Theor Appl Genet (2015) 128:965–984
DOI 10.1007/s00122-015-2483-3

ORIGINAL PAPER

QTL for fatty acid composition and yield in linseed (*Linum usitatissimum* L.)

Santosh Kumar · Frank M. You · Scott Duguid · Helen Booker · Gordon Rowland · Sylvie Cloutier

Received: 14 October 2014 / Accepted: 11 February 2015 / Published online: 8 March 2015

Abstract

Key message **The combined SSR-SNP map and 20 QTL for agronomic and quality traits will assist in marker assisted breeding as well as map-based cloning of key genes in linseed.**

Abstract Flax is an important nutraceutical crop mostly because it is a rich source of omega-3 fatty acids and antioxidant compounds. Canada is the largest producer and exporter of oilseed flax (or linseed), creating a growing need to improve crop productivity and quality. In this study, a genetic map was constructed based on selected 329 single nucleotide polymorphic markers and 362 simple sequence repeat markers using a recombinant inbred line population of 243 individuals from a cross between the Canadian varieties CDC Bethune and Macbeth. The genetic map consisted of 15 linkage groups comprising 691 markers with an average marker density of one marker every 1.9 cM. A total of 20 quantitative trait loci (QTL) were identified corresponding to 14 traits. Three QTL each for oleic acid and stearic acid, two QTL each for linoleic acid and iodine value and one each for palmitic acid, linolenic acid, oil content, seed protein, cell wall, straw weight, thousand seed weight, seeds per boll, yield and days to maturity were identified. The QTL for cell wall, straw weight, seeds per boll, yield and days to maturity all co-located on linkage group 4. Analysis of the candidate gene regions underlying the QTL identified proteins involved in cell wall and fibre synthesis, fatty acid biosynthesis as well as their metabolism and yield component traits. This study provides the foundation for assisting in map-based cloning of the QTL and marker assisted selection of a wide range of quality and agronomic traits in linseed and potentially fibre flax.

Communicated by A. J. Bervillé.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2483-3) contains supplementary material, which is available to authorized users.

S. Kumar · S. Cloutier (✉)
Department of Plant Science, University of Manitoba, 66 Dafoe Road, Winnipeg, MB R3T 2N2, Canada
e-mail: sylvie.j.cloutier@agr.gc.ca

Present Address:
S. Kumar
Brandon Research Centre, 2701 Grande Valley Road, Brandon, MB R7A 5Y3, Canada

F. M. You · S. Duguid
Cereal Research Centre, Agriculture and Agri-Food Canada, 101 Route 100, Morden, MB R6M 1Y5, Canada

H. Booker · G. Rowland
Crop Development Centre, University of Saskatchewan, 51 Campus Drive, Saskatoon, SK S7N 5A8, Canada

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

Introduction

Flax (*Linum usitatissimum* L., $2n = 2x = 30$) is a self-pollinated annual crop belonging to the Linaceae family that has been known to mankind since the Paleolithic era (30,000 years ago) (Zohary 1999). It is believed to have originated in India or the Middle East (Vavilov 1951). Flax is a dual purpose crop grown for its stem fibre (fibre flax) and oil (linseed, oilseed flax or flaxseed). The linseed and fibre types belong to the same species but are morphologically, anatomically, physiologically and agronomically distinct as a result of divergent selection (Diederichsen and Ulrich 2009). Linseed plants are shorter and more

branched, and branches cover a greater proportion of the main stem compared to fibre flax. The seeds are larger and the seed yield is higher because this plant type was primarily selected for its seed yield and quality.

Fibre flax is mainly grown in Europe, Russia and China while Canada is the world's largest producer and exporter of linseed (FAOSTAT 2014). Flaxseed oil is a component of various biodegradable products such as linoleum flooring, paints and varnishes. Flax oil is rich in omega-3 fatty acids which improve cardiac and bone health when consumed by humans and animals (Kim and Ilich 2011; Leyva et al. 2011). Flax stems contain cellulose-rich bast fibres used in textile industry whereas the shorter lignin-rich xylem fibres are manufactured into biodegradable composites for housing and automotive industries (Summerscales et al. 2010). Hybrid composite fibres made of flax and carbon fibres have been characterized with improved water absorption behaviour and overall strength (Dhakal et al. 2013). In the past decade, the flax industry has devoted resources to develop dual purpose flax where high-value products from linseed stems could be used in the pulp, technical fibre and biofuel industries. The government–industry partnership has generated ~$25 million in funding (Canadian flax industry update 2011; www.flaxcouncil.ca/files/web/FlaxIndustryUpdate).

The importance and value of the crop have been the justification for the development of genetic resources such as the first whole genome shotgun (WGS) sequence (Wang et al. 2012), expressed sequence tags (ESTs) (Venglat et al. 2011), a simple sequence repeat (SSR) consensus genetic map (Cloutier et al. 2012), a physical map (Ragupathy et al. 2011) and thousands of single nucleotide polymorphic (SNP) markers (Kumar et al. 2012). The ~370 Mb flax genome is estimated to have undergone whole genome duplication approximately, 5–9 million years ago. However, recent studies based on flax transcriptome analyses suggest an earlier whole genome duplication event, dating 20–40 million years ago (Muhlhausen and Kollmar 2013; Sveinsson et al. 2013). QTL underlying linolenic and linoleic acid contents were identified on two linkage groups using an SSR-based linkage map from a cross between SP2047, a Solin™ line generated using mutation breeding with 2–4 % linolenic acid (LIN) and UGG5-5, a brown seeded breeding line with 63–66 % LIN (Cloutier et al. 2011). The gene encoding fatty acid desaturase 3A co-located with the QTL on linkage group 7 and was hypothesized to be the candidate gene. QTL for seed quality traits (Soto-Cerda et al. 2014b), yield and agronomic components (Soto-Cerda et al. 2014a) were identified by association mapping based on more than 450 SSR markers from a core collection of 407 flax accessions.

Oil content (OIL) in current Canadian linseed varieties, ranges between 45 and 50 % but levels less than 32 % have been observed in fibre flax accessions. Linseed oil is composed of five main fatty acids: palmitic (PAL, C16:0, 6 %), stearic (STE, C18:0, 4.4 %), oleic (OLE, C18:$1^{\Delta 9}$, 24.2 %), linoleic (LIO, C18:$2^{\Delta 9,\ 12}$, 15.3 %) and linolenic (LIN, C18:$3^{\Delta 9,\ 12,\ 15}$, 50.1 %) (Westcott and Muir 2003). STE is the primary substrate that undergoes three desaturations, to first be converted into OLE, then LIO and finally, LIN. Linseed's high proportion of LIN imparts the oil with the drying properties desired for the fabrication of paints, varnishes and linoleum floorings (Cullis 2007). A few varieties with more than 70 % LIN have been developed (Friedt et al. 1995) and, a high-LIN variety with 68 % LIN called NuLin™ 50 has been registered in Canada (www.viterra.ca). The high-LIN content makes the oil highly susceptible to oxidation rendering it unsuitable for many food applications. To meet the fatty acid composition profile of the margarine industry, mutation breeding efforts led to development of varieties with reduced LIN levels (~3 %), known as Linola™ or Solin™ (Green 1986; Rowland 1991). The fatty acid composition of Solin™ oil is similar to other premium polyunsaturated oils such as sunflower. Solin™ oil (C16:0, 6 %), (C18:0, 4 %), (C18:$1^{\Delta 9}$, 16 %), (C18:$2^{\Delta 9,\ 12}$, 72 %) and (C18:$3^{\Delta 9,\ 12,\ 15}$, 2 %) has a high LIO content (Green 1995). Oils from these varieties have higher solidification temperatures that are suitable for the margarine industry (Dribnenki and Green 1995; Dribnenki et al. 2007). Unfortunately, the market niches for these specialty linseed varieties remain underdeveloped.

Linseed yields of 2.5 tonnes per hectare (t/ha) were reported in field plots, (Soto-Cerda et al. 2014b) although the crop yield worldwide is less than 1 t/ha (FAOSTAT 2014). Even in the major producing countries, yield averages only 1.4 t/ha. Yield therefore, remains the most important breeding objective because linseed must compete against other crops. The comparatively small investments in flax breeding, the lack of a good hybrid system and the industry's sensitivity to genetically modified linseed leaves breeders with limited, short and medium term options to tackle the yield gap. Hence, conventional breeding approaches incorporating markers associated with quantitative trait loci (QTL) for yield and yield components represent an attractive option.

In this study, a genetic map was developed using SSR and SNP markers based on a population (BM) of 243 recombinant inbred lines (RILs) generated from a cross between CDC Bethune and Macbeth. The phenotypic data was collected in four consecutive years at two locations and QTL analyses were performed for agronomic, seed and fibre traits.

Materials and methods

Plant materials, experimental design, phenotyping and statistical analysis

A RIL population of 243 individuals from a cross between the Canadian cultivars CDC Bethune (Rowland et al. 2002)

and Macbeth (Duguid et al. 2003b) was grown in the field in a type-2 modified augmented design (MAD) (Lin and Poushinsky 1985). Main plots were arranged in grids of seven rows and seven columns. Each main plot was divided into seven parallel subplots (1.5 m × 2 m with a 20 cm row spacing) with a plot control (CDC Bethune) located in the centre of each main plot. Additional subplots controls, 19 each of the cultivars Hanley (Duguid et al. 2003a) and Macbeth, were assigned to randomly selected main plots and, the RILs were randomly assigned to the remaining subplots. Field experiments were carried out at two locations, namely Morden, Manitoba (M) and Kernen Crop Research Farm located near Saskatoon, Saskatchewan (S) in Canada, during 2009, 2010, 2011 and 2012, for a total of eight environments. Phenotypic data for fatty acid composition (PAL, STE, OLE, LIO and LIN), iodine value (IOD), oil content (OIL), seed yield (YLD), fibre components [cell wall (CW), cellulose (CEL) and lignin (LIG)] were obtained from all eight environments while data for straw weight (SW), height (HGT), thousand seed weight (TSW), seeds per unit area (SPA), seeds per boll (SPB), seed protein content (PRO) and days to maturity (DM) were obtained from six environments (2010–2012). A total of 1 g of seed from each line/location/year was subsampled to measure OIL (%) and fatty acid composition as previously described (Soto-Cerda et al. 2014a). PRO was obtained by near-infrared (NIR) spectroscopy calibrated against the Combustion Analysis Reference Method and expressed on an N × 6.25 dry basis. Fibre traits (CW, CEL, and LIG) were determined by NIR spectroscopy while SW measured the fresh weight of the straw of each 0.5 m row after boll stripping. Phenotypic measurements of the agronomic traits YLD, HGT, TSW, SPA, SPB and DM were collected as previously described (Soto-Cerda et al. 2014b).

The phenotypic data was adjusted for soil heterogeneity using the MAD pipeline (You et al. 2013). Broad-sense heritability (H) for each trait was estimated as in Soto-Cerda et al. (2014a). The error variance was estimated using the data of the three control cultivars (one main plot control and two subplot controls). The SAS GLM procedure and a custom Perl script were used to calculate the broad-sense heritability. Analysis of variance (ANOVA) was performed with the PROC GLM (fixed model) procedure of SAS. An overall mean dataset was generated by averaging the phenotypic traits over all environments and Pearson correlation coefficients between pairs of traits were determined by PROC CORR. All analyses of phenotypic data were carried out using SAS v9.3 (SAS Institute, NC, USA).

Genetic and physical mapping

A genetic map of the BM population consisting of 389 SSR markers from 243 RILs was previously constructed and used in conjunction with two other maps to build a consensus genetic map of flax (Cloutier et al. 2012). To discover SNPs from the same population, genomic DNA from the parental lines and 96 RILs were sequenced as paired end reads of 100 bp using an Illumina HiSeq 2000 as previously described (Kumar et al. 2012). The genome-wide SNP discovery was performed using a modified AGSNP pipeline for genotyping-by-sequencing (GBS) (Kumar et al. 2012; You et al. 2012). The strict criteria applied in the pipeline eliminated false SNPs caused by mis-genoptyping, producing an initial subset of 122,066 SNPs. Subsequently, the SNPs that had >20 % missing data were removed, reducing the number to 3230. The 389 SSRs previously reported for this population (Cloutier et al. 2012) and the 3230 SNPs generated by GBS were assigned coordinates based on the WGS sequence assembly of CDC Bethune (Wang et al. 2012), available on Phytozome (http://phytozome.jgi.doe.gov/). For SNP markers, the precise coordinates of alternative bases/alleles were used whereas for SSRs, the coordinates corresponding to the first base of the repeat were used. SSR and SNP marker names were tagged with linkage group number, contig number, scaffold number and SNP position on the scaffold. A SSR–SNP genetic map using the above 389 SSRs and 3230 SNPs was generated using MapDisto (Lorieux 2012). Of the 3230 SNPs, we retained only 379 corresponding to SNPs that did not map with other SNPs and were approximately spaced every 4 cM apart and that mapped to regions not covered by SSRs. Information on the genomic context of each SNP is provided as a list comprising 100 bp flanking sequence on either side of each SNP (ESM1). Finally, the 389 SSRs and 379 SNPs were used to generate a combined SSR–SNP genetic map for the BM population mentioned above using JoinMap (Stam 1993) with the goal of producing a genetic map achieving comprehensive coverage using a minimal marker set.

QTL analyses

QTL analyses were performed using the single trait multiple interval mapping (MIM) function (Kao et al. 1999) as implemented in the Windows Qgene version 4.3.10. Cofactors selection performed using forward, backward and stepwise manner with the scan interval set at 10 cM yielded consistent results with all three methods and hence, the 'forward cofactor selection' method with default settings was adopted. The LOD thresholds for individual traits were determined by 1000 permutations at $P < 0.05$ level (Churchill and Doerge 1994). QTL were declared when the LOD score was greater than the threshold value in the overall mean dataset and in at least half of the environments. The estimates of additive effect and coefficient of determination (R^2) explaining the percentage of phenotypic

Table 1 Description of the linkage groups (LGs) constituting the combined simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) based genetic map of the recombinant inbred line population derived from CDC Bethune/Macbeth

Linkage group	Number of markers	Marker type		Density (marker/cM)	Length of LGs (cM)
		SSR	SNP		
1	47	23	24	1.7	80
2	52	38	14	2.1	107
3	58	35	23	2.0	115
4	55	34	21	2.0	108
5	28	22	6	1.3	37
6	61	38	23	1.2	71
7	43	20	23	1.5	64
8	69	37	32	1.8	127
9	51	22	29	1.5	75
10	48	17	31	1.7	83
11	25	11	14	3.1	77
12	61	31	30	1.7	103
13	35	12	23	2.3	82
14	43	15	28	2.2	93
15	15	7	8	3.0	44
Total	691	362	329	1.9	1266

variance governed by the putative QTL was obtained by MIM analysis. The genetic map showing the QTL was drawn using Mapchart version 2.2 (Voorrips 2002).

Gene identification

The gene identification was performed using the Gbrowse feature of the CDC Bethune genomic DNA sequence available on the Phytozome server (http://phytozome.jgi.doe.gov/). Either the region between the markers where a QTL was identified or the 1Mbp region up- and downstream of the marker coordinates linked to a QTL was scanned for all annotated genes and their functional classes were determined. In the absence of annotations, the predicted gene and its encoded protein sequence were analyzed by BLAST against the nucleotide sequence and encoded protein databases available at the National Center for Biotechnology Information (NCBI) and Swiss Institute of Bioinformatics (Expasy). The list of gene sequences was then curated based on their putative role in the phenotypic trait(s).

Results

Genetic map

A total of 691 (362 SSRs and 329 SNPs) of the selected 389 SSRs and 379 SNPs grouped into 15 linkage groups (LGs) totalling 1266 cM (Table 1, ESM2). The average LG length was 84 cM and the average marker density was estimated to be one marker per 1.9 cM with LG 11 having the highest marker density and LG6 the lowest.

Phenotypic analysis

The mean, range and broad-sense heritability (*H*) for PAL, STE, OLE, LIO, LIN, IOD, OIL, PRO, CW, SW, TSW, SPB, YLD and DM are summarized in Table 2. Broad-sense heritability was high for all fatty acids (0.98–0.99) and low for a complex trait such as yield (0.53) (Table 2). Transgressive segregants were observed for all traits (Fig. 1). Significant variability for genotype (G), year (Y) and location (L) was observed for all traits except for L of PAL (Table 3). Interactions between G, Y, and L were significant for all traits except in four instances. Some correlations between traits were observed using Pearson correlation coefficients (Table 4). For example, LIN was negatively correlated to all of its precursors in the FA biosynthetic pathway, (Table 4) while TSW was correlated with OIL, PRO and YLD but negatively correlated with LIN, SPB and DM. YLD was also positively correlated with SPB.

QTL analyses

A total of 20 QTL were detected for 14 traits based on the RIL population evaluated in multiple environments (Table 5; Fig. 2). QTL for CW, SW, SPB, YLD and DM co-located with marker Lu2031 on LG4 (Fig. 2). The QTL for CW had a peak LOD of 8 and accounted for 14 % of the phenotypic variation whereas for SW, the LOD peak

Table 2 Mean and range for fibre, seed and yield related traits calculated across environments and locations in a recombinant inbred line population derived from CDC Bethune/Macbeth

Trait	Abbreviation	CDC Bethune/Macbeth recombinant inbred line population			Broad sense heritability (*H*)	CDC Bethune	Macbeth
		Mean ± SE	Min.	Max.			
Palmitic acid (%)	PAL	5.0 ± 0.1	4.5	5.9	0.98	5.0 ± 0.1	4.9 ± 0.1
Stearic acid (%)	STE	3.6 ± 0.2	2.6	4.6	0.98	3.7 ± 0.2	3.8 ± 0.2
Oleic acid (%)	OLE	20.0 ± 1.5	16.4	25.3	0.99	21.4 ± 1.7	17.9 ± 1.3
Linoleic acid (%)	LIO	16.0 ± 0.3	13.2	18.3	0.99	15.1 ± 0.3	16.2 ± 0.4
Linolenic acid (%)	LIN	55.4 ± 1.4	49.9	61.7	0.98	54.7 ± 1.6	57.2 ± 1.1
Iodine value	IOD	189.8 ± 2.7	179.0	198.7	0.98	187.6 ± 3.2	193.2 ± 2.2
Oil content (%)	OIL	45.2 ± 0.6	42.7	48.5	0.95	44.6 ± 0.7	46.3 ± 0.8
Seed protein (%)	PRO	25.0 ± 1.3	23.3	27	0.90	24.6 ± 1.4	24.1 ± 1.3
Cell wall (%)	CW	79.9 ± 1.1	78.0	81.8	0.92	79.3 ± 1.2	80.2 ± 1.2
Straw weight (g)	SW	24.7 ± 4.3	14.4	36.3	0.83	24.7 ± 4.3	21.5 ± 2.5
Thousand seed weight (g)	TSW	5.8 ± 0.3	5.0	7.0	0.82	5.5 ± 0.3	6.0 ± 0.3
Seeds per boll	SPB	7.0 ± 0.5	5.5	8.0	0.81	7.1 ± 0.6	6.9 ± 0.4
Yield (t/ha)	YLD	1.6 ± 2.2	1.3	1.8	0.53	1.6 ± 0.3	1.7 ± 0.2
Days to Maturity	DM	94.7 ± 3.1	90	102	0.89	95.7 ± 3.5	96.0 ± 2.7

was 19 and accounted for 30 % of the phenotypic variation based on the mean dataset (Table 5). The 11 QTL for fatty acid composition and IOD were located on five LGs (1, 3, 5, 6 and 7) with QTL for OLE, LIN and IOD co-located on LG5 and QTL for LIO and OLE co-located on LG3. Three major QTL for OLE (*QOle.BM.crc-LG3-1*, *QOle.BM.crc-LG3-2*, *QOle.BM.crc-LG5*) were identified, contributing 38 % of the phenotypic variability observed for OLE content (Table 5). One QTL each for OIL (*QOil.BM.crc-LG8*) and seed protein (*QPro.BM.crc-LG11*) were identified with LOD scores of 7, accounting for 13 and 11 % of the phenotypic variation, respectively. A minor QTL for TSW (*QTsw.BM.crc-LG15*) was identified in three out of the four environments studied with a LOD peak of 5 and accounted for 9 % of the phenotypic variation.

Genes underlying QTL

The genomic locations of markers that were identified to harbour QTL were scanned using the Gbrowse feature on Phytozome (http://phytozome.jgi.doe.gov/), for the presence of all annotated genes (ESM3) and a subset of genes known to affect the underlying traits is listed in Table 6. The region including SSR marker Lu2031 on LG4 and associated with QTL for CW, SW, SPB, YLD and DM is proximal to a number of candidate genes encoding proteins that could be involved in controlling these traits (Table 6). They include several genes potentially regulating CW synthesis, such as UDP-glycosyltransferase superfamily, laccase, xylosidase and nudix hydrolase family proteins. Genes like plasmodesmata callose-binding protein, pectate lyase family protein are candidates for SW while remorin family protein and COBRA-like protein could play a role in both CW and SW. Genes encoding pollen Ole 1 family and actin cross-linking proteins have putative functions in controlling SPB. Various transcription factors such as GRAS family, WRKY family and genes like Teosinte Branched 1 found at the locus could potentially have an effect on yield. Light-regulated and senescence-related genes are putative candidates for controlling DM (Table 6).

The proteins underlying QTL for fatty acid composition include 3-ketoacyl–acyl carrier protein synthases (KASI and KASIII), histidine box containing HXXXD-type acyl-transferase (H-AT), fatty acid desaturases (SAD and FAD2), fatty acid reductases and fatty acid synthase 1 (FAS1). The QTL for PRO harbours candidate genes such as late embryogenesis abundant protein, seed storage 2S albumin superfamily protein and ribosomal protein S5. The QTL for TSW on LG15 encompasses genes encoding proteins for leucine-rich receptor-like protein kinase family, cytochrome P450 family, as well as WRKY and GRAS family transcription factors. While the above list is not exhaustive, it aims to show that the QTL identified herein underlie candidate genes with functions known to be related to the traits (Table 6).

Discussion

The current study was intended to identify QTL for agronomic, fibre and seed quality traits from a cross between two modern Canadian varieties: CDC Bethune (Rowland

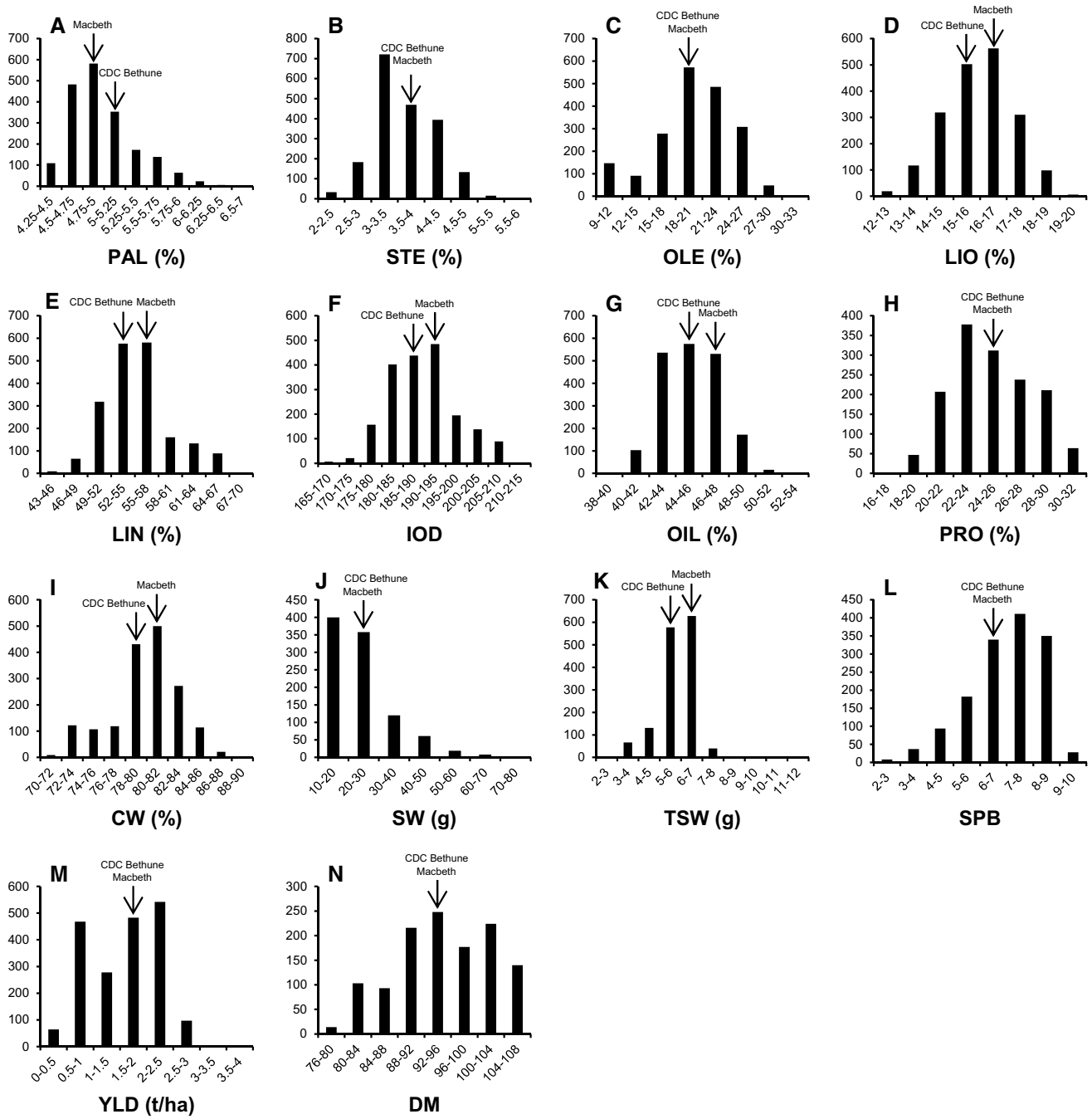


Fig. 1 Phenotypic distribution of palmitic acid (**a**), stearic acid (**b**), oleic acid (**c**), linoleic acid (**d**), linolenic acid (**e**), iodine value (**f**), oil content (**g**) seed protein (**h**), cell wall (**i**), straw weight (**j**), thousand seed weight (**k**), seeds per boll (**l**), yield (**m**) and days to maturity (**n**) in the flax CDC Bethune/Macbeth RIL population evaluated in multiple environments. *Arrows* indicate the interval corresponding to the mean phenotypic values of the parents

et al. 2002) and Macbeth (Duguid et al. 2003). CDC Bethune was developed from the cross between NorMan (Kenaschuk and Hoes 1986) and FP857 while Macbeth originated from the cross between M2701 and AC Linora (Kenaschuk and Rashid 1993). These varieties were chosen because they are adapted to Canadian growing conditions and QTL identified can be easily incorporated into adapted germplasm. Both have similar maturity, height, seed size, oil quality and LIN content with high protein content and good lodging resistance but CDC Bethune consistently out yields Macbeth in co-op trials (Duguid et al. 2003; Rowland et al. 2002).

Table 3 Analysis of variance for estimation of components of variance due to year, location and genotype for the 14 traits from which QTL were identified

Trait[a]	Source of variations													
	G		Y		L		Y × L		Y × G		L × G		Y × L × G	
	MS	%TSS	MS	%TSS	MS	%TSS	MS	%TSS	MS	%TSS	MS	%TSS	MS	%TSS
PAL	0.93***	55.59	15.80***	11.51	0.03	0.01	10.74***	7.82	0.03***	5.19	0.03***	1.98	0.03***	4.74
STE	1.24***	31.70	56.24***	17.66	192.44***	20.14	12.38***	3.89	0.02***	1.90	0.09***	2.40	0.02***	1.61
OLE	31.09***	15.20	1267.60***	7.59	19065.79***	38.04	2506.82***	15.01	0.71***	1.04	1.54***	0.75	0.64***	0.93
LIO	7.85***	45.56	164.47***	11.68	743.58***	17.61	22.43***	1.59	0.26***	4.45	0.54***	3.11	0.18	3.12
LIN	31.58***	17.01	1048.98***	6.92	15560.33***	34.20	2490.12***	16.42	0.95***	1.53	2.29***	1.23	0.80***	1.27
IOD	104.73***	15.30	3967.68***	7.10	64833.84***	38.66	7576.69***	13.55	2.74***	1.20	7.26***	1.06	2.58***	1.12
OIL	10.63***	20.98	647.77***	15.66	2906.27***	23.42	192.75***	4.66	1.03***	6.12	0.98***	1.94	0.86***	5.03
PRO	3.06***	4.06	878.32***	9.52	6573.50***	35.64	1761.64***	19.10	0.69***	1.84	0.85***	1.13	0.65***	1.73
CW	3.47***	3.59	581.79***	7.37	991.29***	4.19	5333.25***	45.06	2.28*	7.07	2.24	2.31	2.03	4.15
SW	102.06***	20.42	3027.66***	4.95	20999.86***	17.15	–[b]	–[b]	14.18	5.62	52.48***	10.29	–[b]	–[b]
TSW	0.78***	14.12	134.24***	19.80	75.55***	11.15	93.25***	27.51	0.23***	8.29	0.20***	3.62	0.21***	7.57
SPB	1.63***	11.44	45.00***	2.58	868.90***	49.87	29.23***	3.36	0.55***	7.71	0.95***	6.67	0.42***	5.76
YLD	11.31***	2.39	6198.21***	16.00	11508.33***	9.90	12412.83***	32.04	12.05***	7.62	12.49***	2.63	6.72***	4.22
DM	34.08***	0.04	15168.08***	18.25	15353.03***	18.47	1885.35***	2.27	7.19***	0.01	10.74***	0.01	8.55***	0.01

G genotype, *Y* year, *L* location, *MS* mean square values, *%TSS* percentage of the total sum of squares

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$

[a] Traits as per Table 2

[b] Could not be estimated because of missing data for Morden

Table 4 Pearson correlation coefficients calculated from the means across environments and locations amongst the traits for which QTL were detected in the recombinant inbred line population CDC Bethune/Macbeth

Trait[a]	PAL	STE	OLE	LIO	LIN	IOD	OIL	PRO	CW	SW	TSW	SPB	YLD
STE	−0.24***												
OLE	0.06	0.15*											
LIO	0.30***	−0.42***	−0.06										
LIN	−0.32***	−0.07	−0.86***	−0.41***									
IOD	−0.31***	−0.25***	−0.92***	−0.16*	0.96***								
OIL	−0.14*	0.06	0.27***	0.01	−0.22***	−0.22***							
PRO	−0.07	−0.35***	0.04	0.20**	−0.06	0.03	0.10						
CW	−0.10	−0.06	−0.13*	0.01	0.14*	0.15*	0.08	0.10					
SW	0.25***	0.01	0.09	0.09	−0.17**	−0.17**	0.19**	−0.23***	−0.25***				
TSW	−0.09	−0.17**	0.17**	0.07	−0.13*	−0.09	0.28***	0.36***	0.12	−0.28***			
SPB	0.04	−0.28***	−0.16*	0.31***	0.03	0.13*	0.07	0.33***	0.25***	−0.42***	−0.13*		
YLD	−0.21***	0.15*	−0.02	−0.06	0.05	0.04	0.12	−0.01	0.03	−0.15*	0.13*	0.32***	
DM	0.03	0.33***	0.18**	−0.19**	−0.13*	−0.21***	−0.12	−0.37***	−0.32***	0.60***	−0.33***	−0.66***	−0.18**

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$

[a] Traits as per Table 2

Table 5 Putative QTL detected for fibre traits, fatty acid composition, yield and days to maturity in the linseed CDC Bethune/Macbeth RIL population

Trait[a]	Linkage group	Marker	QTL[b]	Environment	LOD[c]	Add[d]	R^{2e}
PAL	7	Lu402/Lu7-1820805	*QPal.BM.crc-LG7*	M-2009	7	−0.11	0.13
				S-2009	5	−0.09	0.08
				M-2010	4	−0.09	0.08
				S-2011	4	−0.08	0.07
				M-2012	4	−0.09	0.07
				S-2012	5	−0.09	0.10
				Overall mean	6	−0.10	0.12
STE	1	Lu2183a/Lu1-2670961	*QSte.BM.crc-LG1*	M-2009	4	−0.09	0.07
				S-2009	5	−0.09	0.08
				S-2010	4	−0.07	0.07
				S-2012	4	−0.08	0.06
				Overall mean	3	−0.08	0.06
STE	3	Lu3-8415336/Lu2164	*QSte.BM.crc-LG3*	M-2009	3	0.11	0.06
				M-2010	7	0.21	0.13
				M-2011	10	0.25	0.17
				S-2011	6	0.16	0.11
				M-2012	10	0.23	0.17
				Overall mean	7	0.16	0.12
STE	11	Lu2128/Lu11-19000928	*QSte.BM.crc-LG11*	M-2009	5	−0.13	0.09
				S-2009	6	−0.13	0.10
				S-2010	9	−0.13	0.15
				S-2011	49	−0.13	0.07
				M-2012	59	−0.17	0.09
				S-2012	4	−0.11	0.07
				Overall mean	5	−0.13	0.08
OLE	3	Lu3-3979616/Lu3-5950394	*QOle.BM.crc-LG3-1*	M-2010	8	0.89	0.13
				M-2011	7	0.85	0.13
				M-2012	4	0.60	0.07
				S-2012	4	0.49	0.08
				Overall mean	7	0.66	0.13
OLE	3	Lu658/Lu3150	*QOle.BM.crc-LG3-2*	S-2009	6	−0.44	0.10
				S-2010	4	−0.26	0.07
				M-2012	8	−0.65	0.14
				S-2012	5	−0.42	0.09
				Overall mean	7	−0.51	0.13
OLE	5	Lu5-9728492	*QOle.BM.crc-LG5*	S-2009	6	0.44	0.11
				M-2010	5	0.54	0.09
				S-2010	5	0.29	0.10
				M-2011	6	0.56	0.10
				S-2011	4	0.43	0.08
				M-2012	7	0.60	0.12
				S-2012	5	0.40	0.09
				Overall mean	7	0.47	0.12
LIO	3	Lu3-3979616/Lu3-5950394	*QLio.BM.crc-LG3*	M-2009	5	−0.37	0.09
				S-2009	4	−0.31	0.07
				M-2010	7	−0.49	0.12
				M-2011	4	−0.42	0.08
				Overall mean	5	−0.32	0.08

Table 5 continued

Trait[a]	Linkage group	Marker	QTL[b]	Environment	LOD[c]	Add[d]	R^{2e}
LIO	6	Lu2545	*QLio.BM.crc-LG6*	M-2009	3	−0.25	0.06
				M-2010	3	−0.26	0.06
				S-2011	4	−0.23	0.07
				M-2012	5	−0.32	0.09
				S-2012	3	−0.19	0.05
				Overall mean	4	−0.23	0.07
LIN	5	Lu5-9728492	*QLin.BM.crc-LG5*	M-2009	3	−0.49	0.06
				S-2009	5	−0.47	0.10
				M-2010	4	−0.54	0.07
				S-2010	6	−0.47	0.10
				M-2011	4	−0.62	0.08
				S-2011	4	−0.48	0.07
				M-2012	6	−0.65	0.10
				S-2012	5	−0.48	0.09
				Overall mean	5	−0.52	0.10
IOD	5	Lu5-9728492	*QIod.BM.crc-LG5*	M-2009	3	−0.77	0.06
				S-2009	6	−0.91	0.11
				M-2010	3	−0.94	0.06
				S-2010	9	−0.89	0.15
				M-2011	5	−1.14	0.09
				S-2011	4	−0.86	0.08
				M-2012	6	−1.09	0.10
				S-2012	6	−0.86	0.10
				Overall mean	7	−0.97	0.12
IOD	6	Lu6-2260313/Lu6-2330258	*QIod.BM.crc-LG6*	S-2009	6	−0.90	0.10
				S-2010	5	−0.68	0.08
				M-2011	3	−0.97	0.06
				M-2012	4	−0.93	0.07
				Overall mean	5	−0.86	0.08
OIL	8	Lu8-22516618/Lu3189	*QOil.BM.crc-LG8*	S-2009	3	−0.32	0.06
				M-2010	7	−0.40	0.12
				M-2011	6	−0.35	0.10
				M-2012	6	−0.32	0.11
				Overall mean	7	−0.36	0.13
PRO	11	Lu11-21716266/Lu52	*QPro.BM.crc-LG11*	S-2009	3	0.26	0.06
				M-2011	6	0.20	0.10
				S-2011	4	0.24	0.05
				M-2012	7	0.26	0.12
				Overall mean	6	0.22	0.11
CW	4	Lu2031	*QCw.BM.crc-LG4*	M-2009	3	−0.38	0.06
				S-2009	4	−0.51	0.08
				M-2010	6	−0.48	0.11
				S-2010	10	−0.52	0.17
				Overall mean	8	−0.25	0.14
SW	4	Lu2031	*QSw.BM.crc-LG4*	M-2010	12	5.09	0.21
				S-2010	8	2.13	0.14
				S-2012	12	1.71	0.20
				Overall mean	19	2.47	0.30

Table 5 continued

Trait[a]	Linkage group	Marker	QTL[b]	Environment	LOD[c]	Add[d]	R^{2e}
TSW	15	Lu2010a/Lu2001	*QTsw.BM.crc-LG15*	S-2011	5	0.09	0.09
				M-2012	4	0.15	0.07
				S-2012	6	0.10	0.11
				Overall mean	5	0.10	0.09
SPB	4	Lu2031	*QSpb.BM.crc-LG4*	M-2010	21	−0.65	0.33
				S-2010	4	−0.18	0.07
				M-2011	6	−0.38	0.11
				Overall mean	12	−0.22	0.20
YLD	4	Lu2031	*QYld.BM.crc-LG4*	S-2009	4	0.80	0.08
				M-2010	17	−2.01	0.27
				S-2010	5	−1.23	0.09
				M-2011	8	−0.94	0.15
				S-2011	4	0.59	0.07
				S-2012	4	0.72	0.06
				Overall mean	4	−0.42	0.08
DM	4	Lu2031	*QDm.BM.crc-LG4*	M-2010	19	0.02	0.31
				M-2012	5	0.01	0.09
				S-2012	10	0.02	0.17
				Overall mean	20	0.01	0.31

[a] Traits as per Table 2

[b] QTL nomenclature is as previously used in flax (Cloutier et al. 2011): Q for QTL followed by trait abbreviation, a period, *pmap* showing QTL analysis performed on physical map, a period, originating laboratory, a hyphen and linkage group in which the QTL is located

[c] The estimated LOD score at the QTL peak

[d] Additive effect contributed by the two parents at a defined locus. Positive values indicate a positive contribution to the trait by CDC Bethune and the negative values indicate a positive contribution by Macbeth. The scores are presented as absolute values

[e] R^2 is the percentage of phenotypic variation explained by the QTL in the cross

The final goal of QTL analyses is often to envision map-based cloning towards the identification of the functional unit(s) affecting the trait. Map-based cloning requires a high-density map combined with precise phenotyping. Molecular markers such as amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD) and SSR have been developed in flax (Adugna et al. 2006; Cloutier et al. 2009, 2011; Everaert et al. 2001; Fu et al. 2002; Oh et al. 2000; Roose-Amsaleg et al. 2006; Spielmeyer et al. 1998; Stegniĭ et al. 2000). Three QTL studies have been published based on an RFLP and RAPD map of 94 markers (Oh et al. 2000), an AFLP map of 213 markers (Spielmeyer et al. 1998) and an SSR map of 113 markers (Cloutier et al. 2011). More recently, a consensus map of 770 SSR markers based on the genetic mapping of three bi-parental populations constituted a major improvement over previous maps but remained insufficiently saturated for many applications such as the fine-mapping of QTL and the subsequent map-based cloning of the candidate functional units (Cloutier et al. 2012). The even more recent development of SNPs affords this opportunity (Kumar et al. 2012). In the current study, SNPs and SSRs were selected at regular intervals to create a fairly evenly saturated map with the aim of localizing QTL. The 243 CDC Bethune/Macbeth RIL population mapped with 691 informative markers (362 SSRs and 329 SNPs) into 15 LGs formed the basis for the QTL analyses of stem fibre, seed quality and agronomic traits.

Population and QTL analyses

Previous studies recommended that large populations of approximately 400 were preferable for QTL studies (Vales et al. 2005). However, real data analysis and simulation based studies have concluded that phenotyping precision and high-density molecular markers can improve the accuracy of QTL mapping even with a smaller population size (Yang et al. 2013). The combined use of a decently sized population, the MAD design, the number of locations and years, established phenotyping protocols and a genetic map with well distributed markers should all be factors contributing positively towards the accuracy of the QTL detected.

The QTL analyses were performed using the MIM method which uses multiple marker intervals simultaneously for mapping putative QTL. MIM uses the maximum likelihood for estimating genetic parameters and Cockerham's model for interpreting genetic parameters (Kao et al. 1999). Compared to some other interval mapping algorithms, the MIM model is more precise (Kao et al. 1999; Li et al. 2007), although each algorithm has its unique characteristics (Jourjon et al. 2005). The intent here was not a comparative analysis of algorithms and, as such, the MIM model has been used.

Trait QTL and underlying genes

The QTL for CW, SW, SPB, yield and DM were all linked to SSR marker Lu2031 on LG4 and the underlying genomic region could be classified as gene-rich. A study suggests that 15 % of the genes in a genome may be involved in CW synthesis, remodeling or turnover (Carpita et al. 2001). Expression profiling and reverse genetics approaches identified a number of genes involved with CW synthesis such as COBRA-like protein, laccase and glycosyltransferase family proteins which were also identified as genes underlying cell wall QTL in the current study (Brown et al. 2005). As cell walls are the primary component of the straw, it is not surprising to observe that CW and SW have a QTL in common.

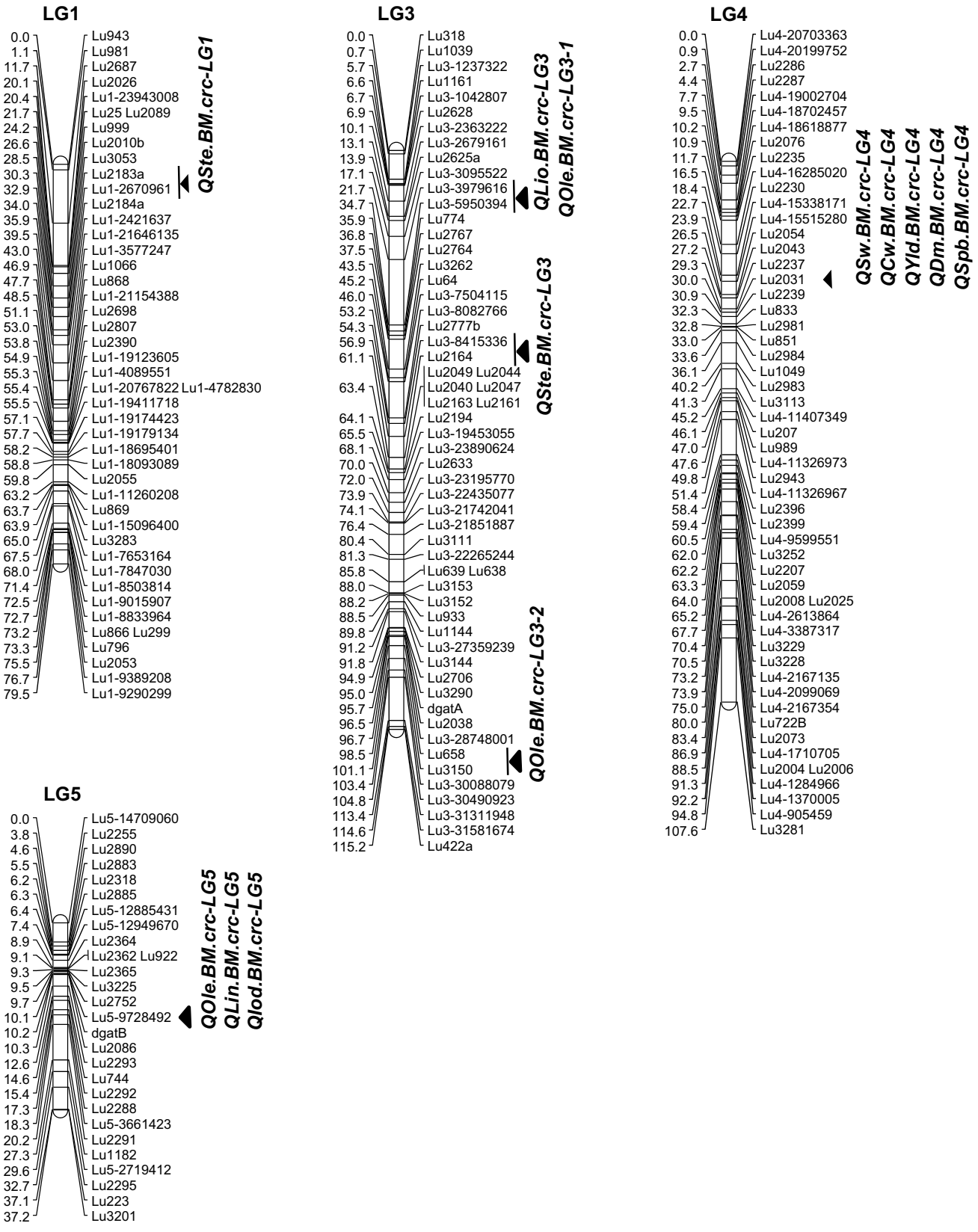
One way of measuring plant fertility is to determine the number of seeds produced per ovary. One factor influencing SPB is pollen success and, as such, genes involved in pollen–stigma interactions, pollen germination, pollen tube growth and egg cell fertilization will have an effect on the number of seeds produced. The gene encoding protein Ole e 1, which is involved in pollen hydration, pollen tube growth, and pollen–stigma interactions, is therefore, a candidate for *QSpb.BM.crc-LG4* (Muschietti et al. 1994). These functions are consistent with the localization of Ole e 1 in the pollen exine, the sub-apical and apical regions of the pollen tube and the germination medium (Prado et al. 2013).

Genes encoding proteins known to affect yield in oilseed crops such as diacylglycerol acyltransferase (DGAT), whose transcription level was correlated in Arabidopsis with enhanced seed weight and oil deposition (Jako et al. 2001), were also present at the YLD QTL. DGAT regulates seed oil concentration and yield in soybean (Eskandari et al. 2013). The WRKY type DNA-binding transcription factors are involved in seed cavity enlargement during seed development which determines the final seed size and yield potential (Kang et al. 2013). The WRKY and GRAS transcription factors are also key regulators of stress signaling in plants (Chen et al. 2012; Ma et al. 2010) that contribute to yield stability in changing environmental conditions.

Fig. 2 The combined SSR–SNP map of the CDC Bethune/Macbeth recombinant inbred line population comprising 691 markers (362 SSRs and 329 SNPs). *Arrows* indicate the position of QTL. QTL nomenclature: Q for QTL followed by trait abbreviation [cell wall (*Cw*); straw weight (*Sw*); palmitic acid (*Pal*); stearic acid (*Ste*); oleic acid (*Ole*); linoleic acid (*Lio*); linolenic acid (*Lin*); iodine value (*Iod*); oil content (*Oil*); seed protein (*Pro*); thousand seed weight (*Tsw*); seeds per boll (*Spb*); yield (*Yld*) and days to maturity (*Dm*)], a period, the population, a second period, the originating laboratory, a hyphen and the linkage group where the QTL located

TSW is a yield component and WRKY and GRAS transcription factors were identified underlying its QTL. Genes encoding a cytochrome P450 family protein and a leucine-rich receptor-like kinase family protein known to affect grain shape and size were also detected near the TSW QTL (Huang et al. 2013). Underlying the QTL for DM on LG4 was a Lir1 protein whose expression is regulated by the circadian clock and activated by light (Reimmann and Dudler 1993). Mutation studies identified a role for Lir1 in the photoperiodic regulation of flowering suggesting a potential role in determining DM.

The current study identified QTL for PAL on LG7 harbouring genes for H-AT which belongs to the BAHD acyltransferases and facilitates acyl group transfer (St-Pierre and Luca 2000). The KASIII enzyme catalyzes the condensation of acetyl–CoA and malonyl–CoA to initiate fatty acid synthesis. The malonyl-thioesters undergo recurring condensation with acetyl–CoA to form PAL catalyzed by KASI (Baud and Lepiniec 2010). The QTL for STE, identified on LG1, LG3 and LG11, accounting for 26 % of the phenotypic variation, differed from a previously reported QTL for STE discovered through association mapping on LG7 that accounted for 19 % of the variation (Soto-Cerda et al. 2014a). The STE QTL on LG1 contained genes for H-AT to facilitate fatty acid chain elongation whereas the SAD gene uses STE as a substrate to form OLE through the creation of a double bond at the Δ9 position, thereby affecting STE (Ohlrogge and Jaworski 1997). The STE QTL on LG3 encompassed genes for KASI and KASII involved in synthesis of PAL and STE, respectively, affecting the production of precursors. QTL for OLE on LG3 encompassed genes for fatty acid reductase which adds CoA to long-chain aldehydes to form long-chain acyl–CoA (Riendeau et al. 1982), FAD2 which converts OLE to LIO (Shanklin and Cahoon 1998), various acyl-transferases and acyl-activating enzymes also known as acyl–CoA synthetases (Conti et al. 1996). The QTL on LG5 harboured a gene for FAS which catalyzes de novo biosynthesis of fatty acids. Unlike animal FAS, a multi-subunit multifunctional protein, plant FAS is a dissociable multi-subunit complex consisting of enzymes such as KAS family, 3-ketoacyl–ACP reductase, 3-hydroxyacyl–ACP dehydrase and enoyl–ACP reductase (Ohlrogge and Jaworski 1997) which catalyzes


LG1
0.0 Lu943
1.1 Lu981
11.7 Lu2687
20.1 Lu2026
20.4 Lu1-23943008
21.7 Lu25 Lu2089
24.2 Lu999
26.6 Lu2010b
28.5 Lu3053
30.3 Lu2183a
32.9 Lu1-2670961
34.0 Lu2184a
35.9 Lu1-2421637
39.5 Lu1-21646135
43.0 Lu1-3577247
46.9 Lu1066
47.7 Lu868
48.5 Lu1-21154388
51.1 Lu2698
53.0 Lu2807
53.8 Lu2390
54.9 Lu1-19123605
55.3 Lu1-4089551
55.4 Lu1-20767822 Lu1-4782830
55.5 Lu1-19411718
57.1 Lu1-19174423
57.7 Lu1-19179134
58.2 Lu1-18695401
58.8 Lu1-18093089
59.8 Lu2055
63.2 Lu1-11260208
63.7 Lu869
63.9 Lu1-15096400
65.0 Lu3283
67.5 Lu1-7653164
68.0 Lu1-7847030
71.4 Lu1-8503814
72.5 Lu1-9015907
72.7 Lu1-8833964
73.2 Lu866 Lu299
73.3 Lu796
75.5 Lu2053
76.7 Lu1-9389208
79.5 Lu1-9290299
QSte.BM.crc-LG1
LG3
0.0 Lu318
0.7 Lu1039
5.7 Lu3-1237322
6.6 Lu1161
6.7 Lu3-1042807
6.9 Lu2628
10.1 Lu3-2363222
13.1 Lu3-2679161
13.9 Lu2625a
17.1 Lu3-3095522
21.7 Lu3-3979616
34.7 Lu3-5950394
35.9 Lu774
36.8 Lu2767
37.5 Lu2764
43.5 Lu3262
45.2 Lu64
46.0 Lu3-7504115
53.2 Lu3-8082766
54.3 Lu2777b
56.9 Lu3-8415336
61.1 Lu2164
Lu2049 Lu2044
63.4 Lu2040 Lu2047
Lu2163 Lu2161
64.1 Lu2194
65.5 Lu3-19453055
68.1 Lu3-23890624
70.0 Lu2633
72.0 Lu3-23195770
73.9 Lu3-22435077
74.1 Lu3-21742041
76.4 Lu3-21851887
80.4 Lu3111
81.3 Lu3-22265244
85.8 Lu639 Lu638
88.0 Lu3153
88.2 Lu3152
88.5 Lu933
89.8 Lu1144
91.2 Lu3-27359239
91.8 Lu3144
94.9 Lu2706
95.0 Lu3290
95.7 dgatA
96.5 Lu2038
96.7 Lu3-28748001
98.5 Lu658
101.1 Lu3150
103.4 Lu3-30088079
104.8 Lu3-30490923
113.4 Lu3-31311948
114.6 Lu3-31581674
115.2 Lu422a
QLio.BM.crc-LG3
QOle.BM.crc-LG3-1
QSte.BM.crc-LG3
QOle.BM.crc-LG3-2
LG4
0.0 Lu4-20703363
0.9 Lu4-20199752
2.7 Lu2286
4.4 Lu2287
7.7 Lu4-19002704
9.5 Lu4-18702457
10.2 Lu4-18618877
10.9 Lu2076
11.7 Lu2235
16.5 Lu4-16285020
18.4 Lu2230
22.7 Lu4-15338171
23.9 Lu4-15515280
26.5 Lu2054
27.2 Lu2043
29.3 Lu2237
30.0 Lu2031
30.9 Lu2239
32.3 Lu833
32.8 Lu2981
33.0 Lu851
33.6 Lu2984
36.1 Lu1049
40.2 Lu2983
41.3 Lu3113
45.2 Lu4-11407349
46.1 Lu207
47.0 Lu989
47.6 Lu4-11326973
49.8 Lu2943
51.4 Lu4-11326967
58.4 Lu2396
59.4 Lu2399
60.5 Lu4-9599551
62.0 Lu3252
62.2 Lu2207
63.3 Lu2059
64.0 Lu2008 Lu2025
65.2 Lu4-2613864
67.7 Lu4-3387317
70.4 Lu3229
70.5 Lu3228
73.2 Lu4-2167135
73.9 Lu4-2099069
75.0 Lu4-2167354
80.0 Lu722B
83.4 Lu2073
86.9 Lu4-1710705
88.5 Lu2004 Lu2006
91.3 Lu4-1284966
92.2 Lu4-1370005
94.8 Lu4-905459
107.6 Lu3281
QSw.BM.crc-LG4
QCw.BM.crc-LG4
QYld.BM.crc-LG4
QDm.BM.crc-LG4
QSpb.BM.crc-LG4
LG5
0.0 Lu5-14709060
3.8 Lu2255
4.6 Lu2890
5.5 Lu2883
6.2 Lu2318
6.3 Lu2885
6.4 Lu5-12885431
7.4 Lu5-12949670
8.9 Lu2364
9.1 Lu2362 Lu922
9.3 Lu2365
9.5 Lu3225
9.7 Lu2752
10.1 Lu5-9728492
10.2 dgatB
10.3 Lu2086
12.6 Lu2293
14.6 Lu744
15.4 Lu2292
17.3 Lu2288
18.3 Lu5-3661423
20.2 Lu2291
27.3 Lu1182
29.6 Lu5-2719412
32.7 Lu2295
37.1 Lu223
37.2 Lu3201
QOle.BM.crc-LG5
QLin.BM.crc-LG5
Qlod.BM.crc-LG5

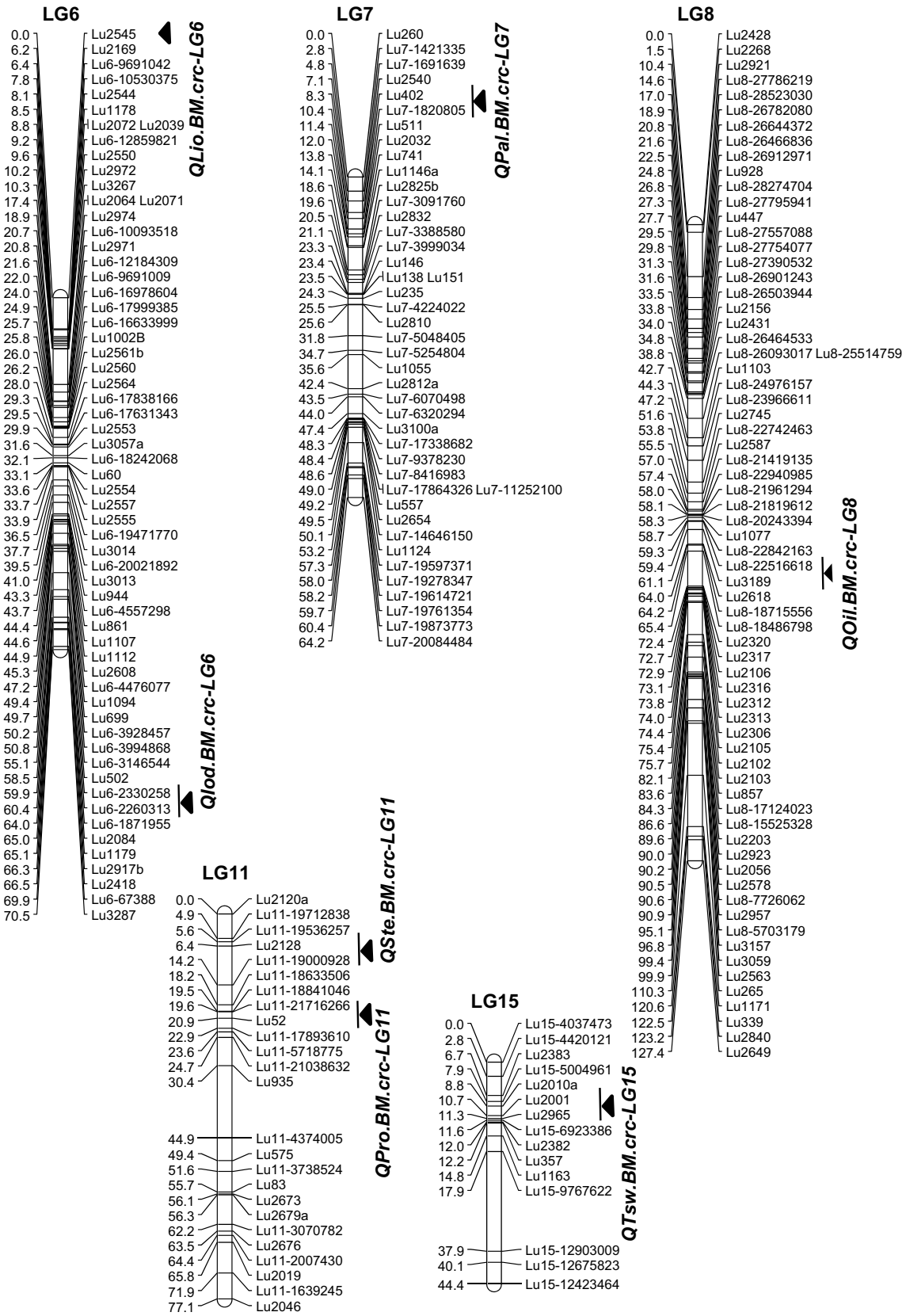


Fig. 2 continued

Table 6 Candidate genes discovered underlying the QTL identified for 14 traits

Trait[a]	LG	Marker interval	Size (cM)	Gene[b]	Gene ID	Encoded protein
PAL	7	Lu402/Lu7-1820805	2	*KAS IIIb-2*	Lus10028925	3-ketoacyl-acyl carrier protein synthase III
					Lus10028818	Flavin containing amine oxidoreductase family protein
					Lus10028813	HXXXD-type acyl-transferase family protein
STE	1	Lu2183a/Lu1-2670961	2.6	*SAD3-1*	Lus10018926	Steroyl-ACP desaturase
					Lus10014695	HXXXD-type acyl-transferase family protein
					Lus10003621	Long-chain fatty alcohol dehydrogenase family protein
					Lus10003557	Myb domain protein
					Lus10010142	Phosphoenolpyruvate carboxylase family protein
STE	3	Lu3-8415336/Lu2164	4.1	*KAS Ib-1*	Lus10040883	3-ketoacyl-acyl carrier protein synthase I
					Lus10029225	Myb-like HTH transcriptional regulator family protein
					Lus10024747	Acetyl-CoA carboxylase 1
				mtKAS-1	Lus10014622	3-ketoacyl-acyl carrier protein synthase II
					Lus10019568	Alcohol dehydrogenase transcription factor Myb/SANT-like family protein
					Lus10019567	Lipase class family protein
STE	11	Lu2128/Lu11-19000928	7.8		Lus10039330	HXXXD-type acyl-transferase family protein
					Lus10039316	Plastidic pyruvate kinase beta subunit
					Lus10039296	NADH-ubiquinone oxidoreductase-related
OLE	3	Lu3-3979616/Lu3-5950394	12.9		Lus10019187	Fatty acid reductase
					Lus10019183	HXXXD-type acyl-transferase family protein
					Lus10003470	Myb-like transcription factor family protein
					Lus10003688	Acyl-CoA N-acyltransferases (NAT) superfamily protein
OLE	3	Lu658/Lu3150	2.6		Lus10017154	Membrane bound O-acyl transferase (MBOAT) family protein
					Lus10017096	Myb domain protein 108
					Lus10017089	Fatty acid desaturase 2
					Lus10017079	O-acetyltransferase family protein
					Lus10037735	Acyl-activating enzyme 7
OLE	5	Lu5-9728492	0.2		Lus10000592	HXXXD-type acyl-transferase family protein
					Lus10006668	Acyl-CoA N-acyltransferases (NAT) superfamily protein
				KAS IIIa-2	Lus10032246	3-ketoacyl-acyl carrier protein synthase III
					Lus10032323	Acyl-CoA N-acyltransferases (NAT) superfamily protein
				KAS II-1	Lus10034886	Fatty acid synthase 1
				KCS10	Lus10028105	3-ketoacyl-CoA synthase 10
				KCS12-3	Lus10029880	3-ketoacyl-CoA synthase 7
					Lus10024036	Myb family transcription factor
LIO	3	Lu3-3979616/Lu3-5950394	12.9		Lus10019187	Fatty acid reductase
					Lus10019183	HXXXD-type acyl-transferase family protein
					Lus10037023	Oxidoreductase, 2OG-Fe(II) oxygenase family protein
					Lus10003470	Myb-like transcription factor family protein
					Lus10003688	Acyl-CoA N-acyltransferases (NAT) superfamily protein
LIO	6	Lu2545	6.2	*KCS5-2*	Lus10012611	3-ketoacyl-CoA synthase 1
					Lus10004809	Myb-like HTH transcriptional regulator family protein
					Lus10004833	Glycerol-3-phosphate acyltransferase 6

Table 6 continued

Trait[a]	LG	Marker interval	Size (cM)	Gene[b]	Gene ID	Encoded protein
LIN	5	Lu5-9728492	0.2		Lus10000592	HXXXD-type acyl-transferase family protein
					Lus10006668	Acyl-CoA N-acyltransferases (NAT) superfamily protein
				KAS IIIa-2	Lus10032246	3-ketoacyl-acyl carrier protein synthase III
					Lus10032323	Acyl-CoA N-acyltransferases (NAT) superfamily protein
				KAS II-1	Lus10034886	Fatty acid synthase 1
				KCS10	Lus10028105	3-ketoacyl-CoA synthase 10
				KCS12-3	Lus10029880	3-ketoacyl-CoA synthase 7
					Lus10024036	Myb family transcription factor
IOD	5	Lu5-9728492	0.2		Lus10000592	HXXXD-type acyl-transferase family protein
					Lus10006668	Acyl-CoA N-acyltransferases (NAT) superfamily protein
				KAS IIIa-2	Lus10032246	3-ketoacyl-acyl carrier protein synthase III
					Lus10032323	Acyl-CoA N-acyltransferases (NAT) superfamily protein
				KAS II-1	Lus10034886	Fatty acid synthase 1
				KCS10	Lus10028105	3-ketoacyl-CoA synthase 10
				KCS12-3	Lus10029880	3-ketoacyl-CoA synthase 7
					Lus10024036	Myb family transcription factor
IOD	6	Lu6-2330258/Lu6-2260313	0.6		Lus10017610	Plastidic glucose-6-phosphate translocator
OIL	8	Lu8-22516618/Lu3189	1.7		Lus10022222	UDP-Glycosyltransferase superfamily protein
					Lus10022219	Sec23/Sec24 transport family protein
					Lus10022211	Ureidoglycolate amidohydrolase
					Lus10022332	Glyceraldehyde-3-phosphate dehydrogenase C2
					Lus10022308	NAD-dependent glycerol-3-phosphate dehydrogenase family protein
					Lus10022302	Kunitz family trypsin and protease inhibitor protein
					Lus10022292	Senescence-related gene
					Lus10011062	Phospholipid/glycerol acyltransferase family protein
PRO	11	Lu11-21716266/Lu52	1.3		Lus10023623	Pectate lyase family protein
					Lus10023631	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
					Lus10019030	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
					Lus10031300	Asparagine synthetase
					Lus10031136	Zincin-like metalloproteases family protein
					Lus10031129	Ribosomal protein S5/Elongation factor G/III/V family protein
CW	4	Lu2031	0.7		Lus10036840	UDP-Glycosyltransferase superfamily protein
					Lus10036740	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
					Lus10036722	Zinc knuckle (CCHC-type) family protein
					Lus10015817	Remorin family protein
					Lus10034348	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
					Lus10034379	COBRA-like protein
					Lus10034439	Laccase
					Lus10041457	β-xylosidase
					Lus10041481	Laccase/Diphenol oxidase family protein
					Lus10041626	Nudix hydrolase homolog

Table 6 continued

Trait[a]	LG	Marker interval	Size (cM)	Gene[b]	Gene ID	Encoded protein
SW	4	Lu2031	0.7		Lus10036914	Plasmodesmata callose-binding protein
					Lus10036773	NAC-like, activated by AP3/PI
					Lus10036721	Pectate lyase family protein
					Lus10015817	Remorin family protein
					Lus10015791	Exostosin family protein
					Lus10034379	COBRA-like protein precursor
YLD	4	Lu2031	0.7		Lus10036891	WRKY DNA-binding protein
					Lus10036824	Fatty acid reductase
					Lus10036819	HXXXD-type acyl-transferase family protein
					Lus10036773	NAC-like, activated by AP3/PI
					Lus10036761	AGAMOUS-like
					Lus10015760	TEOSINTE BRANCHED 1, cycloidea, PCF (TCP)-domain family protein 20
					Lus10041540	RGA like GRAS family transcription factor
					Lus10041457	β-xylosidase
					Lus10041511	Ferulic acid 5-hydroxylase
DM	4	Lu2031	0.7		Lus10041522	CCR-like; Lir1 Light regulated protein
					Lus10041535	Senescence-related gene
					Lus10041483	Phytochrome and flowering time regulatory protein (PFT1)
					Lus10036746	Phytochrome kinase substrate

The genes were identified using the CDC Bethune annotated sequence available at http://phytozome.jgi.doe.gov/

[a] Traits as per Table 2

[b] Genes involved in fatty acid synthesis (You et al. 2014)

synthesis of PAL. In addition, various acyl-transferases and a KASI gene were also identified on LG5, which may play a role in OLE in plants.

Cloutier et al. (2011) identified two QTL for LIO on LG7 and LG16 accounting for 34 and 20 % of the phenotypic variability in the doubled haploid population generated from a cross between SP2047 and UGG5-5. Soto-Cerda et al. (2014a) identified three QTL for LIO on LG3, LG5 and LG12 using association studies, one of which validated a previously identified QTL (Cloutier et al. 2011). In the current study, two QTL were identified for LIO on LG3 and LG6 accounting for 15 % of the total phenotypic variation but they differed from the QTL identified in the two studies mentioned above. This is not surprising because in the first instance, the cross between a low and a high linolenic acid line was more likely to identify major genes such as *fad2* and *fad3* (Cloutier et al. 2011) while in the second, the authors looked at a broad germplasm collection with variable allelic richness and LIN ranging from 2 to 65 % (Soto-Cerda et al. 2014a) contrasting with the present study using a bi-parental population between two conventional linseed varieties both with 55–57 % LIN. The lack of segregation of the major allele QTL was more likely to unravel minor allele QTL in the latter. Indeed, association mapping is considered more powerful from the increased polymorphism point of view, potentially resulting in the identification of additional QTL. The limitation then becomes the availability of these alleles in the breeding germplasm. In QTL studies based on linkage mapping from bi-parental populations, there are only two alleles and the strength of the QTL identification depends in part on the size of the mapping populations and the number of polymorphic markers available. The current study is based on a bi-parental population of 243 RILs developed from two modern Canadian varieties and a genetic map consisting of 691 markers (SSRs and SNPs). Taken together, these attributes are more likely to facilitate the application of these QTL in Canadian germplasm.

In addition to the acyl-transferases identified for other fatty acids, candidate genes underlying these LIO QTL included genes for fatty acid reductase and KAS1 mainly involved in synthesis of PAL. Genes encoding Myb transcription factors on both LIO QTL suggest their involvement in regulating *fad2* genes that contain Myb interacting sites on their gene promoters (Peng and Weselake 2011; Thambugala and Cloutier 2014). In previous reports, QTL for LIN and LIO co-located (Cloutier et al. 2011; Soto-Cerda et al. 2014a), an observation that differs from our results where QTL for LIN and IOD were identified at the same location as a QTL for OLE on LG5, contributing

~10 and 12 % of phenotypic variation for LIN and IOD, respectively.

In dicot species like flax, seed reserve components usually consist of seed storage proteins and/or storage lipids such as waxes or triacylglycerols (TAGs) which are formed by esterification of fatty acids (PAL, STE, OLE, LIO and LIN) in a glycerol backbone. In addition to various fatty acid biosynthesis genes, a number of transcription factors, secretory and transporter proteins also regulate seed oil content (Baud and Lepiniec 2010). The QTL for seed oil encompassed genes such as UDP–glycosyltransferase which are highly expressed during the early stages of seed development in high oil lines of soybean (Wei et al. 2008). The genes for Kunitz family trypsin inhibitor protein as well as the senescence-related proteins accumulated more transcripts in high oil compared to low oil lines (Wei et al. 2008). A proteomic study in flax characterized 2S albumin (conlinin) and 11S globulin (legumin, glutenin type A, cupin) proteins (Barvkar et al. 2012). The seed protein QTL on LG11 had multiple copies of genes encoding seed storage proteins such as 2S albumin family protein, late embryogenesis abundant (LEA) proteins and glycoproteins (Table 6). In addition, we also detected 7S globulin (48 kDa glycoprotein precursor) protein. The 2S albumins are small (1.7–2.2S), sulphur-rich plant storage proteins synthesized and stored in developing seeds and act as a source of nutrients during subsequent germination and seedling growth (Boutilier et al. 1999). Flax contains two seed specific 2S albumins type proteins: conlinin1 and conlinin2 (Truksa et al. 2003). LEA proteins protect other proteins from aggregation during seed desiccation at maturity or osmotic stresses associated with low temperature (Goyal et al. 2005). Other proteins identified underlying the PRO QTL included the hydroxyproline-rich glycol-proteins which play a vital role in plant growth and development including stress tolerance (Showalter et al. 2010).

The identification of a number of genes that potentially affect the studied traits is a good start but additional gene expression data and fine-mapping of QTL will be required to pinpoint genes/markers affecting traits and to distinguish causal polymorphism(s). In addition, genes/markers validated by other QTL studies or association studies will provide greater confidence for their use in breeding.

Conclusions

The current study described the first high-density SSR–SNP map in linseed and identified QTL for important agronomic, fibre and seed specific traits. The evenly distributed map, use of RILs from a conventional cross and phenotypic data from eight environments for most of the traits assisted in identification of major and minor QTL for fourteen traits that are important to the commercial success of flax. The availability of the sequenced genome of CDC Bethune, one of the parents of the population, allowed identification of several putative effector genes and transcription factors underlying the traits. Candidate genes based on known functionality related to the traits were discussed. This information combined with the resequencing data of the Canadian core collection of 407 accessions and its phenotypic assessment promise to bolster the efficiency of map-based cloning particularly for highly heritable and major QTL. Despite these resources, the definitive identification and validation of the functional unit(s) underlying QTL remain challenging considering that most QTL are not associated with structural changes in protein coding sequences. Nevertheless, the identification of marker Lu2031 tightly linked (0.7 cM) to a number of QTL regulating seed and straw yield as well as DM is an immediate outcome of the QTL mapping presented herein and it likely will be a valuable marker for breeding dual purpose flax.

Author contribution statement SC, SD, and GR designed the experiment. SD and HB produced the phenotypic data and SC, the genotypic data. FMY performed the SNP discovery analysis. SK performed data analysis and co-wrote the manuscript with SC.

Acknowledgments The authors are grateful to Gord Penner and Evelyn Loewen, the Cloutier lab members and the breeding teams at the Morden Research Station and the Crop Development Centre for technical assistance. We are grateful to Andrzej Walichnowski for his proofreading of the manuscript. 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.

Conflict of interest The authors declare that they have no conflict of interest.

References

Adugna W, Labuschagne MT, Viljoen CD (2006) The use of morphological and AFLP markers in diversity analysis of linseed. Biodivers Conserv 15:3193–3205

Barvkar VT, Pardeshi VC, Kale SM, Kadoo NY, Giri AP, Gupta VS (2012) Proteome profiling of flax (*Linum usitatissimum*) seed: characterization of functional metabolic pathways operating during seed development. J Proteome Res 11:6264–6276

Baud S, Lepiniec L (2010) Physiological and developmental regulation of seed oil production. Prog Lipid Res 49:235–249

Boutilier K, Hattori J, Baum BR, Miki BL (1999) Evolution of 2S albumin seed storage protein genes in the Brassicaceae. Biochem Syst Ecol 27:223–234

Brown DM, Zeef LA, Ellis J, Goodacre R, Turner SR (2005) Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. Plant Cell 17:2281–2295

Carpita N, Tierney M, Campbell M (2001) Molecular biology of the plant cell wall: searching for the genes that define structure, architecture and dynamics. Plant Mol Biol 47:1–5

Chen L, Song Y, Li S, Zhang L, Zou C, Yu D (2012) The role of WRKY transcription factors in plant abiotic stresses. Biochim Biophys Acta 1819:120–128

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

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, Niu Z, Duguid S (2011) SSR-based linkage map of flax (*Linum usitatissimum* L.) and mapping of QTLs underlying fatty acid composition traits. Mol Breed 28:437–451

Cloutier S, Miranda E, Ragupathy R, Radovanovic N, Reimer E, Walichnowski A, Ward K, Rowland G, Duguid S (2012) Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). Theor Appl Genet 125:1783–1795

Conti A, Rogers J, Verdejo P, Harding CR, Rawlings AV (1996) Seasonal influences on stratum corneum ceramide 1 fatty acids and the influence of topical essential fatty acids. Int J Cosmet Sci 18:1–12

Cullis C (2007) Flax: genome mapping and molecular breeding in plants. Springer, Berlin

Dhakal HN, Zhang ZY, Guthrie R, Macmullen J, Bennett N (2013) Development of flax/carbon fibre hybrid composites for enhanced properties. Carbohydr Polym 96:1–8

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

Dribnenki JCP, Green AG (1995) Linola™ '947' low linolenic acid flax. Can J Plant Sci 75:201–202

Dribnenki JCP, McEachern SF, Chen Y, Green AG, Rashid KY (2007) 2149 solin (low linolenic flax). Can J Plant Sci 87:297–299

Duguid SD, Kenaschuk EO, Rashid KY (2003) Macbeth flax. Can J Plant Sci 83:803–805

Eskandari M, Cober E, Rajcan I (2013) Using the candidate gene approach for detecting genes underlying seed oil concentration and yield in soybean. Theor Appl Genet 126:1839–1850

Everaert I, Riek JD, Loose MD, Waes JV, Bockstaele EV (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

FAOSTAT (2014) Production of crops: flax fibre and linseed: area harvested and production (tonnes). http://faostat3faoorg/. Accessed June 2014

Friedt W, Bickert C, Schaub H (1995) In vitro breeding of high linolenic, doubled haploid lines of linseed (*Linum usitatissimum* L.) via androgenesis. Plant Breed 114:322–326

Fu Y-B, Peterson G, Diederichsen A, Richards KW (2002) RAPD analysis of genetic relationships of seven flax species in the genus *Linum* L. Genet Resour Crop Evol 49:253–259

Goyal K, Walton LJ, Tunnacliffe A (2005) LEA proteins prevent protein aggregation due to water stress. Biochem J 388:151–157

Green AG (1986) Genetic control of polyunsaturated fatty acid biosynthesis in flax (*Linum usitatissimum*) seed oil. Theor Appl Genet 72:654–661

Green A (1995) New linseed breed low in ALA. Australian New Crops Newsletter

Huang R, Jiang L, Zheng J, Wang T, Wang H, Huang Y, Hong Z (2013) Genetic bases of rice grain shape: so many genes, so little known. Trends Plant Sci 18:218–226

Jako C, Kumar A, Wei Y, Zou J, Barton DL, Giblin EM, Covello PS, Taylor DC (2001) Seed-specific over-expression of an Arabidopsis cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant Physiol 126:861–874

Jourjon M-F, Jasson S, Marcel J, Ngom B, Mangin B (2005) MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinfromatics 21:128–130

Kang X, Li W, Zhou Y, Ni M (2013) A WRKY transcription factor recruits the SYG1-like protein SHB1 to activate gene expression and seed cavity enlargement. PLoS Genet 9:e1003347

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216

Kenaschuk EO, Hoes JA (1986) NorLin flax. Can J Plant Sci 66:171–173

Kenaschuk EO, Rashid KY (1993) AC Linora flax. Can J Plant Sci 73:839–841

Kim Y, Ilich JZ (2011) Implications of dietary [alpha]-linolenic acid in bone health. Nutrition 27:1101–1107

Kumar S, You FM, Cloutier S (2012) Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. BMC Genom 13:684

Leyva DR, Zahradka P, Ramjiawan B, Guzman R, Aliani M, Pierce GN (2011) The effect of dietary flaxseed on improving symptoms of cardiovascular disease in patients with peripheral artery disease rationale and design of the FLAX-PAD randomized controlled trial. Contemp Clin Trials 32:724–730

Li H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. Genetics 175:361–374

Lin CS, Poushinsky G (1985) A modified augmented design (type 2) for rectangular plots. Can J Plant Sci 65:743–749

Lorieux M (2012) MapDisto: fast and efficient computation of genetic linkage maps. Mol Breed 30:1231–1235

Ma H-S, Liang D, Shuai P, Xia X-L, Yin W-L (2010) The salt- and drought-inducible poplar GRAS protein SCL7 confers salt and drought tolerance in *Arabidopsis thaliana*. J Exp Bot 61:4011–4019

Muhlhausen S, Kollmar M (2013) Whole genome duplication events in plant evolution reconstructed and predicted using myosin motor proteins. BMC Evol Biol 13:202

Muschietti J, Dircks L, Vancanneyt G, McCormick S (1994) LAT52 protein is essential for tomato pollen development: pollen expressing antisense LAT52 RNA hydrates and germinates abnormally and cannot achieve fertilization. Plant J 6:321–338

Oh TJ, Gorman M, Cullis CA (2000) RFLP and RAPD mapping in flax (*Linum usitatisimum*). Theor Appl Genet 101:590–593

Ohlrogge JB, Jaworski JG (1997) Regulation of fatty acid synthesis. Annu Rev Plant Physiol Plant Mol Biol 48:109–136

Peng F, Weselake R (2011) Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis. BMC Genom 12:286

Prado N, Alché JdD, Casado-Vela J, Mas S, Villalba M, Rodríguez R, Batanero E (2013) Nanovesicles are secreted during pollen germination and pollen tube growth: a possible role in fertilization. Mol Plant. doi:10.1093/mp/sst153

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

Reimmann C, Dudler R (1993) Circadian rhythmicity in the expression of a novel light-regulated rice gene. Plant Mol Biol 22:165–170

Riendeau D, Rodriguez A, Meighen E (1982) Resolution of the fatty acid reductase from *Photobacterium phosphoreum* into acyl protein synthetase and acyl-CoA reductase activities. Evidence for an enzyme complex. J Biol Chem 257:6908–6915

Roose-Amsaleg C, Cariou-Pham E, Vautrin D, Tavernier R, Solignac M (2006) Polymorphic microsatellite loci in *Linum usitatissimum*. Mol Ecol Notes 6:796–799

Rowland GG (1991) An EMS-induced low-linolenic-acid mutant in McGregor flax (*Linum usitatissimum* L.). Can J Plant Sci 71:393–396

Rowland GG, Hormis YA, Rashid KY (2002) CDC Bethune flax. Can J Plant Sci 82:101–102

Shanklin J, Cahoon EB (1998) Desaturation and related modifications of fatty acids 1. Annu Rev Plant Physiol Plant Mol Biol 49:611–641

Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR (2010) A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. Plant Physiol 153:485–513

Soto-Cerda BJ, Duguid S, Booker H, Rowland G, Diederichsen A, Cloutier S (2014a) Association mapping of seed quality traits using the Canadian flax (*Linum usitatissimum* L.) core collection. Theor Appl Genet 127:881–896

Soto-Cerda BJ, Duguid S, Booker H, Rowland G, Diederichsen A, Cloutier S (2014b) Genomic regions underlying agronomic traits in linseed (*Linum usitatissimum* L.) as revealed by association mapping. J Integr Plant Biol 56:75–87

Spielmeyer W, Green AG, Bittisnich D, Mendham N, Lagudah ES (1998) Identification of quantitative trait loci contributing to Fusarium wilt resistance on an AFLP linkage map of flax (*Linum usitatissimum*). Theor Appl Genet 97:633–641

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: join map. Plant J 3:739–744

Stegniĭ VN, Chudinova IuV, Salina EA (2000) RAPD analysis of the flax (*Linum usitatissimum* L.) varieties and hybrids of various productivity. Genetics 36:1370–1373

St-Pierre B, Luca VD (2000) Chapter nine evolution of acyltransferase genes: origin and diversification fo the BAHD superfamily of acyltransferases involved in secondary metabolism. In: John T, Romeo RILV, Vincenzo De L (eds) Recent advances in phytochemistry. Elsevier, Amsterdam, pp 285–315

Summerscales J, Dissanayake NPJ, Virk AS, Hall W (2010) A review of bast fibres and their composites. Part1—fibres as reinforcements. Compos Part A Appl Sci Manuf 41:1329–1335

Sveinsson S, McDill J, Wong GK, Li J, Li X, Deyholos MK, Cronk QC (2013) Phylogenetic pinpointing of a paleopolyploidy event within the flax genus (*Linum*) using transcriptomics. Ann Bot. doi:10.1093/aob/mct306

Thambugala D, Cloutier S (2014) Fatty acid composition and desaturase gene expression in flax (*Linum usitatissimum* L.). J Appl Genet 55:423–432

Truksa M, MacKenzie Samuel L, Qiu X (2003) Molecular analysis of flax 2S storage protein conlinin and seed specific activity of its promoter. Plant Physiol Biochem 41:141–147

Vales MI, Schon CC, Capettini F, Chen XM, Corey AE, Mather DE, Mundt CC, Richardson KL, Sandoval-Islas JS, Utz HF, Hayes PM (2005) Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust. Theor Appl Genet 111:1260–1270

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. Chron Bot 13:1–366

Venglat P, Xiang D, Qiu S, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M, Bekkaoui F, Sharpe A, Wang E, Rowland G, Selvaraj G, Datla R (2011) Gene expression analysis of flax seed development. BMC Plant Biol 11:74

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Geraldes A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe A, Wong GK, Wang J, Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. Plant J 72:461–473

Wei W-H, Chen B, Yan X-H, Wang L-J, Zhang H-F, Cheng J-P, Zhou X-A, Sha A-H, Shen H (2008) Identification of differentially expressed genes in soybean seeds differing in oil content. Plant Sci 175:663–673

Westcott NA, Muir AD (2003) Flax seed lignan in disease prevention and health promotion. Phytochem Rev 2003:401–417

Yang J, Li C, Gong X, Gupta S, Lance R, Zhang G, Loughman R, Zhu J (2013) Large population with low marker density verse small population with high marker density for qtl mapping: a case study for mapping qtl controlling barley net blotch resistance. In: Zhang G, Li C, Liu X (eds) Advance in barley sciences. Springer, The Netherlands, pp 301–315

You FM, Deal KR, Wang J, Britton MT, Fass JN, Lin D, Dandekar A, Leslie CA, Aradhya M, Luo MC, Dvorak J (2012) Genome-wide SNP discovery in walnut with an AGSNP pipeline updated for SNP discovery in allogamous organisms. BMC Genom 13:354

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

You FM, Li P, Kumar S, Ragupathy R, Li Z, Fu YB, Cloutier S (2014) Genome-wide identification and characterization of the gene families controlling fatty acid biosynthesis in flax (*linum usitatissimum* L). J Proteomics Bioinform 7:310–326

Zohary D (1999) Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East. Genet Resour Crop Evol 46:133–142