

# Quantitative trait loci involved in regulating seed oil composition in *Arabidopsis thaliana* and their evolutionary implications

Anushree Sanyal · C. Randal Linder

Received: 4 September 2011 / Accepted: 18 October 2011 / Published online: 10 November 2011  
© Springer-Verlag 2011

**Abstract** Fatty acid composition is an important determinant of seed oil quality. Overall, 72 QTL for 12 fatty acid traits that control seed oil composition were identified in four recombinant inbred line (RIL) populations (*Ler-0* × *Sha*, *Ler-0* × *Col-4*, *Ler-2* × *Cvi*, *Ler-0* × *No-0*) of *Arabidopsis thaliana*. The identified QTL explained 3.2–79.8% of the phenotypic variance; 33 of the 59 QTL identified in the *Ler-0* × *Sha* and the *Ler-0* × *Col* RIL populations co-located with several a priori candidate genes for seed oil composition. QTL for fatty acids 18:1, 18:2, 22:1, and fatty acids synthesized in plastids was identified in both *Ler-0* × *Sha* and *Ler-0* × *Col-4* RIL populations, and QTL for 16:0 was identified in the *Ler-0* × *Sha* and *Ler-0* × *No-0* RIL populations providing strong support for the importance of these QTL in determining seed oil composition. We identified melting point QTL in three RIL populations, and fatty acid QTL collocated with two of them, suggesting that the loci could be under selection for altering the melting point of seed oils to enhance adaptation and could be useful for breeding purposes. Nuclear-cytoplasmic interactions and epistasis were rare. Analysis of the genetic correlations between these loci and other fatty acids

indicated that these correlations would tend to strongly enhance selection for desirable fatty acids.

## Introduction

Seed oils, mainly composed of triacylglycerols (TAGs), are used for both nutritional and industrial purposes. The quality of the seed oil is determined by the constituent fatty acids (FAs) found in the TAGs. The different FA types and their relative proportions in TAGs vary widely both inter- and intra-specifically (Eckey 1954; Hilditch and Williams 1964; Gurr 1980; Harwood 1980; Welch et al. 2000; Marsalkiene et al. 2009). Further, seed oil composition appears to have been under strong selection on a latitudinal cline at both the macro- and micro-evolutionary scales (Linder 2000). The composition of seed oils influences their melting points and differs according to temperature regimens. At lower temperature or higher latitude, the melting point is lower because of greater proportions of unsaturated FAs (Linder 2000). Thus, identifying genomic regions affected by local adaptation and characterizing the traits and genes regulating seed oil composition under different environmental conditions will be useful for breeding oil-seed crops and understanding seed oil evolution.

In recent years, our understanding of seed TAGs biosynthetic pathways has progressed rapidly (Browse and Somerville 1991; Ohlrogge and Browse 1995; Harwood 1996; Broun et al. 1999; Lung and Weselake 2006; Cahoon et al. 1997; Barker et al. 2007). However, the understanding of the genetics underlying the regulation of FA and TAG synthesis is still limited (Ohlrogge and Jaworski 1997; Sharma et al. 2008; Baud and Lepiniec 2009). The proportions of some FAs (i.e., the saturated FAs in seed oils) are nearly constant under a wide variety of

---

Communicated by C. Quiros.

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-011-1742-1) contains supplementary material, which is available to authorized users.

---

A. Sanyal · C. Randal Linder  
Section of Integrative Biology, School of Biological Sciences,  
University of Texas at Austin, Austin, TX 78712, USA

A. Sanyal (✉)  
Department of Agronomy, University of Wisconsin-Madison,  
1575 Linden Drive, Madison, WI 53706, USA  
e-mail: asanyal2@wisc.edu

environmental conditions, and thus under strict genetic control (Hilditch 1956; Calvin 1965; Harris et al. 1978; Djemel et al. 2005).

Previous studies have shown considerable variation in the seed oil composition of *Arabidopsis* (Millar and Kunst 1999; O'Neill et al. 2003) and suggested that populations derived from selected crosses between parental lines differing in oil composition will be useful for investigating these traits. Hobbs et al. (2004) identified QTLs responsible for regulating the seed oil composition of *A. thaliana* in the *Ler-2* × *Cvi* recombinant inbred line (RIL) population. However, since QTL are often population-specific (Lynch and Walsh 1998), determining the full genetic architecture of complex traits using a single population is likely to describe only a fraction of the “global genetic architecture” for said traits (Symonds et al. 2005). To more fully elucidate genetic variations responsible for controlling quantitative trait variation, many experimental segregating populations, derived from a diverse array of accessions, are needed (Tonsor et al. 2005).

The aim of this study is to map QTL for FA traits in four RIL populations (*Ler-0* × *Sha*, *Ler-0* × *Col-4*, *Ler-2* × *Cvi*, *Ler-0* × *No-0*) for capturing the range of genetic variation in seed oil composition. QTL for FA traits, which influence the seed oil composition and melting point (MP), can be very useful in understanding adaptation to temperature and for selecting suitable genotypes for breeding better quality oil in the *Brassica* and other crop species. A further objective is to identify promising candidate genes (CGs) in seed oil biosynthesis giving additional insight into the aforementioned purposes of breeding and seed oil evolution.

## Materials and methods

### Mapping populations

Four recombinant inbred line (RIL) mapping populations (Table 1) were used: *Ler-0* × *Sha* (El-Lithy et al. 2004),

*Ler-0* × *Col-4* (Lister and Dean 1993), *Ler-2* × *Cvi* (Alonso-Blanco et al. 1998), and *Ler-0* × *No-0* (Magliano et al. 2005). All were obtained from the Arabidopsis Biological Resource Center (<http://www.arabidopsis.org/abrc/>). All four populations were previously genotyped in the F8 generation or later by their creators. Genotyping details for each population can be found in their original descriptions and at the *Arabidopsis* Information Service (<http://www.arabidopsis.org>) and the Nottingham *Arabidopsis* Stock Center (<http://nasc.nott.ac.uk>).

### Determination of fatty acid composition of triacylglycerols

Triacylglycerols (TAGs) were extracted twice from 15–20 seeds of each RIL line (unreplicated) and all of the parental lines (two replicates) for all four RIL populations. The unreplicated design was chosen to optimize the power of QTL detection given the number of phenotypical observations, which are very labor intensive (Capelle et al. 2010). All the extracted TAGs from the RIL and parental lines were converted to fatty acid methyl esters (FAMES) (Metcalf and Wang 1981). The FAMES were resolved on a Hewlett-Packard 5890A gas chromatograph using a 30 m 70% Cyanopropyl polysilphenylene-siloxane capillary column (BPx70, SGE, Inc.) and detected by flame ionization. For all runs, an initial oven temperature of 180°C was maintained for 5.5 min and then ramped to 240°C at a rate of 7.5°C/min. Ramping was followed by a final time of 0.50 min at 240°C. Injection volume was 2 µl with a 1:100 split ratio. The results for each run were compiled and analyzed by HP Chemstation software (version A.04.02) with the proportion of each fatty acid estimated by the area under the curve for each peak as a proportion of the total area under the peaks. We measured nine of the eleven fatty acids found in *A. thaliana* seeds (16:0, 18:0, 18:1, 18:2, 18:3, 20:0, 20:1, 20:2 and 22:1). The other two fatty acids (20:3 and 22:0) occurred in minor amounts (each <1% of the total) that could not always be accurately

**Table 1** Four RIL populations used for QTL mapping

Mapping population	Number <sup>a</sup>	Markers <sup>b</sup>	QTL <sup>c</sup>	Linkage group lengths <sup>d</sup>				
				1	2	3	4	5
<i>Ler-0</i> × <i>Col-4</i>	91	773	23	114	68	71	82	100
<i>Ler-0</i> × <i>Sha</i>	96	76	36	91.8	68	75	71.8	79.8
<i>Ler-2</i> × <i>CVI</i>	80	293	7	127	82	83	83	115
<i>Ler-0</i> × <i>No</i>	64	46	6	70	88	93	67	83

The size of the populations, number of genetic markers, and number of major QTL with LOD  $\geq 3$  found and linkage group lengths are presented

<sup>a</sup> Number of recombinant inbred lines used

<sup>b</sup> Number of markers used

<sup>c</sup> Number of major QTL identified

<sup>d</sup> Linkage group lengths in centimorgans

measured. The chromatography peaks were identified using three size standards: RM-6, NHI-F and 189-19 (Supelco).

## Statistical analyses

### QTL analysis

The relative proportions of individual FAs, total saturated FAs (16:0, 18:0 and 20:0), melting point of TAGs and FAs synthesized in plastids ( $\leq 18:1$ ) were used as traits for QTL mapping (Windows QTL Cartographer v2.5; Wang et al. 2007, R/qtl; Broman et al. 2003). Prior to the QTL analysis, phenotypic data for seed oil composition for the four RIL populations were analyzed with Systat 12 and the variation within each trait was assessed using a one-way analysis of variance to compare the means of the four populations. Interval mapping (IM) based on the LOD-score test (Lander and Botstein 1989) and composite interval mapping (CIM) with model 6 activated in the Zmapqtl module were performed. Default parameters (5 co-factors and a window size of 10) were used for QTL detection. Statistically significant QTL ( $p = 0.05$ ) were determined using permutation-derived empirical thresholds based upon 1,000 permutations for each trait mapped (Churchill and Doerge 1994). The estimated additive effect and the percentage of variance explained by each major effect QTL were obtained from the software using the CIM model.

Analysis of the seed oil composition included the ten individual FAs, the total saturated FAs, fatty acids synthesized in plastids and melting point of triacylglycerols were measured. QTL analysis of fatty acids synthesized in plastids was done to examine maternal or plastid-specific effects. QTL analysis of the total saturated fats and estimated melting point of the triacylglycerols were done as evidence suggests these traits may be under selection (Linder 2000). The melting point of each RIL's triacylglycerols was estimated using the relative proportion of each constituent FA to weight the melting point of the FA. Therefore, the estimated melting point of the triacylglycerols of each RIL was the average of the weighted melting points of the constituent fatty acids (Malkin 1954). This estimated measure of the TAGs melting point may not precisely match the empirically determined melting point (Stumpf 1980), but it does not influence the relative ordering and values of the melting points. Thus, our estimated melting points of the TAGs in a RIL are useful even if slightly deviant from the absolute melting point.

### Epistasis

The four RIL populations were also analyzed for epistatic interactions between QTLs by building a stepwise QTL

model using R/qtl (Broman et al. 2003). *T* tests and ANOVAs were performed using Systat 12 (Systat 12 2007) to determine the effects of the genotypes for each trait in the four RIL populations. One-way ANOVAs were performed where, for each FA, the relative proportion of a FA was the dependent variable and the genotypes were the independent factors.

### Cytoplasmic inheritance in the *Ler-2* × *Cvi* RIL population

The *Ler-2* × *Cvi* RIL population was reciprocally crossed, so we tested for cytoplasmic interactions for all major QTL identified in the *Ler-2* × *Cvi* RIL population, using the model  $Y = (Q_1 + \dots + Q_n + \text{cytoplasm} + Q_1 * \text{cytoplasm} \dots + Q_n * \text{cytoplasm})$  with *Y* as the trait,  $Q_1$  to  $Q_n$  as the identified QTL and cytoplasm as the covariate. The tests were performed in R/qtl (Broman et al. 2003). We also tested for nuclear-cytoplasmic interactions for all QTL identified in the *Ler-2* × *Cvi* RIL population by performing two-way ANOVAs in Systat 12 where, for each test, an individual FA was the dependent variable, and the cytoplasm and markers were the independent factors.

### Genetic correlations among traits

We estimated genetic correlations among traits for each RIL population using the Pearson product-moment pairwise correlation between the 12 traits within each of the four RIL populations (Systat 12) to test whether there were significant correlations between traits (Table S1–S4). *P* values were adjusted for multiple-hypothesis testing in R (version 2.9.0 alpha) using the *q* value false discovery rate (FDR) plug-in (Storey and Tibshirani 2003; Storey et al. 2004) and the Benjamini and Hochberg method (BH-FDR) (Benjamini and Hochberg 1995) in the built in function, *p*-adjust. Default parameters were used for the *q* value module except that the bootstrap method was used wherever possible.

### Melting point indices of fatty acids

To assess how seed oil melting point (MP) might be affected by genetic correlations between fatty acid QTL, we developed an index for comparing the relative effects of significant genetic correlations among FAs on the melting points of TAGs (Sanyal and Linder, manuscript submitted). For a pair of significantly correlated FAs, the difference between the individual FAs melting points ( $MP_{FA1}$ ,  $MP_{FA2}$ ) and the average melting point of the TAGs ( $MP_{TAG}$ ) was calculated. The product of these differences was then multiplied by the FAs' genetic correlation ( $corr_{FA1,FA2}$ ) and the two FAs' average proportions ( $p_{FA1}$ ,  $p_{FA2}$ )

in a set of RILs. The difference between the FAs' melting points and the melting point of the TAG accounts for the degree to which change in one FA would affect the TAG melting point due to a change in the correlated FA; the correlation coefficient determines the degree to which selection on one FA would affect the other FA, and the relative proportions of the two FAs involved in the interaction accounts for the relative weight of each FA in determining the melting point of the TAG prior to selection. The said parameters are mathematically annotated by the equation:

$$(\text{MP}_{\text{FA1}} - \text{MP}_{\text{TAG}}) * (\text{MP}_{\text{FA2}} - \text{MP}_{\text{TAG}}) * \text{corr}_{\text{FA1,FA2}} \\ * P_{\text{FA1}} * P_{\text{FA2}}$$

For each RIL set, the indices for the pairs of FAs were standardized by dividing each index by the maximum index value of the set of indices calculated. Therefore, the largest scaled index has a value of 1 with all other indices being between  $-1$  and  $1$ .

#### Candidate gene and QTL co-location analysis

To determine whether CGs co-localized with QTL more often than expected, the QTL confidence interval (CI) was randomized across the linkage maps 1,000 times, and the distribution of the proportion of QTL co-locating with CGs was determined for each RIL population. The number of QTL placed matched the number of QTL found for each RIL population, and the number of possible locations matched the number of markers used for each RIL set. Afterwards, the proportion of times the CGs collocated with the QTL at rates equal to or greater than the rates we observed to obtain a  $p$  value was calculated. If this proportion of the randomized QTL co-locating with the CGs was less than the original co-location of QTL and CGs 95% of the time, the co-location was deemed significant (Pelgas et al. 2011).

#### Map alignments

There is considerable variation in the total length of the linkage maps in the four populations, which ranged from 386.4 to 490 cM (Table 1). Hence, the linkage positions of markers in each population were aligned with their corresponding physical positions on the Col physical map (ARABIDOPSIS GENOME INITIATIVE 2000). QTL physical positions were determined from the known flanking marker positions. Further, the physical positions of the QTL were estimated by using a least-squares regression approach based on the coarse linear relationship between the physical and linkage map position of each marker. The QTL in the *Ler-0* × Col-4 population was mapped directly on the physical map. The QTL mapped in the different

populations was considered to be the same locus when the physical positions of markers within 2-LOD support intervals overlapped (Symonds et al. 2005). In addition, the co-location of the QTL with CGs further confirmed the overlap of QTL in the *Ler-0* × Sha and *Ler-0* × Col-4 populations.

## Results

All 12 FA traits in each of the four RIL populations showed continuous phenotypic variation (Table 3; Figs. S1–S4) and transgressive segregation exhibiting both higher and lower values than the high- and low-value parent, respectively, suggesting action of multiple genes for the traits and that selection could act on genetically diverse populations to rapidly change seed oil compositions to suit different conditions. ANOVA showed significant variation for relative proportions of 12 FA traits with the exception of 20:2 among RILs in the four RIL populations at  $p \leq 0.005$  indicating that a strong genetic component underlies the FA traits.

The *Ler-0* × Sha population had the largest number of QTL, 36 for all 12 FA traits scattered across all five chromosomes. The *Ler-0* × Col-4 population had 23 QTL mapped for all traits with the exception of 20:0 and 20:1 on all five chromosomes. The *Ler-2* × Cvi population had seven QTL mapped on all chromosomes with the exception of chromosome 3 and *Ler-0* × No-0 population had six QTL mapped on all five chromosomes. All four populations had one or more QTL co-locating with other QTL (Tables 4, 5, 6, 7; Figs. 2, 3, 4, 5).

#### Mapping of QTL for fatty acid composition

A total of 72 QTL was mapped with individual effects that explained from 3.2 to 79.8% of within-population variation. The QTL for 16:0, 18:1, 18:2, 22:1, and FAs present in plastids were observed in multiple populations (Tables 4, 5, 6, 7). Models including gene-by-gene interactions showed the presence of epistasis for only 20:1 and 20:2 in the *Ler-0* × Sha population and 18:1 in the *Ler-0* × Col-4 RIL populations.

In the *Ler-0* × Sha population, three QTL hot spots for FA traits were identified where seven to nine QTL co-located on chromosomes 1, 3, and 4 (Table 4; Fig. 2). Interestingly, all nine QTL on chromosome 4 were in the same location. In addition, overlapping QTL to a lesser extent was also observed on chromosomes 2 and 5, respectively. All QTL collocated with at least one QTL. Several of these QTL have the highest LOD peak at the same location suggesting that the same QTL is influencing several traits (Table 4; Fig. 2). Importantly, the four MP

QTL co-locate with one or more FA QTL whose correlations could affect the melting point (Table S1). The largest QTL for 20:1, which explained 79.8% of the phenotypic variance, was identified on chromosome 4, respectively. One to four QTL for each of the 12 FA traits identified across all chromosomes explained 29.1–94.6% of the phenotypic variance. The four QTL for 20:1 explained 94.6% of the phenotypic variance. QTL effects were unidirectional for 18:0, 18:2, MP and saturated FA traits where the ‘*Ler*’ allele was the high allele for all of them excluding 18:2 where it was a low allele. QTL effects were in opposite directions for the other eight FA traits where the ‘*Ler*’ allele was the high allele at some loci and low allele at other loci for the same trait suggesting considerable genetic variation (Table 4). In the *Ler*-0 × *Col*-4 population, the 23 QTL identified for 10 FA traits explained 7–41.3% of the phenotypic variance (Table 5; Fig. 3). One to four QTL identified for each of the FA traits explained 7.9–58.6% of the phenotypic variance. The largest QTL for 18:1 on chromosome 3 explained 41.3% of the phenotypic variance. In addition, one or more QTL co-located on chromosomes 1, 2 and 3, respectively, and their LOD peaks were often in the same position suggesting that the same QTL could be controlling multiple traits. The QTL effects were unidirectional for seven of the FA traits where the ‘*Ler*’ allele was the high allele for 16:0, 18:2, 20:2 and saturated FA traits, and the low allele for 18:1, 18:3 and 22:1, respectively. The QTL effects were in opposite directions at different loci for 18:0, MP, and FAs present in plastids suggesting considerable genetic variation (Table 5). The MP QTL co-locates with 22:1 on chromosome 1, which may not affect the MP, as the correlation is not significant (Table S2). Seven QTL for five FA traits were identified in the *Ler*-2 × *Cvi* population, which explained 13.7–47.8% of the phenotypic variance. The three QTL for 16:0 explained 47.8% of the phenotypic variance for that trait. The QTL mapped on chromosomes 1 and 5 collocated with each other (Table 6). The six QTL identified in the *Ler*-0 × *No*-0 population explained 13.9–23.4% of the phenotypic variance. Two QTL for 16:0 explained 37.3% of the phenotypic variance. Furthermore, the QTL for 16:0 collocated with a QTL for saturated FAs on chromosome 3 (Table 7).

#### Co-location of QTL and candidate genes

In the *Ler*-0 × *Sha* population, 22 of the 36 QTL co-located with one or more CGs, and 9 of the 15 a priori identified CGs co-located with at least one QTL. Remarkably, all the QTL on chromosomes 3 and 4 co-located with relevant CGs, consistent with respect to functions. Two QTL each on chromosomes 2 and 5 also co-located with CGs. None of the seven co-locating QTL on chromosome 1 co-located with a CG. The QTL for 16:0, 18:1, 18:2, 18:3, 20:0, 20:1, 20:2, and

22:1 on chromosomes 3 and 4 collocated with FAE1.1 (AT3G10280), FAE1.2 (AT3G52160), FAE1.3 (AT4G34250), FAE1.4 (AT4G34520) and FAD2 (AT2G29980) genes (Table 4). Seven of these FAs are either the reactants or products of all the four FAE genes with the exception of 16:0 FA on chromosome 3 (Table 2; Fig. 1). The FAE genes elongate 18:1 and 18:2 to 20:1 and 20:2, 20:1 to 22:1, and also 20:0 to 22:0, respectively. The FAD2 also converts 18:1 to 18:2, which are precursors for 20:1 and 20:2, and thus can affect the proportions of those four FAs. The co-location of SAD1 (AT5G16240) and SAD2 (AT5G16230) with a QTL for 20:0 on chromosome 5 is relevant, as 18:0 is a substrate for both the genes and could affect the proportions of 20:0, which is formed from 18:0, respectively. The co-location of DAGAT (AT2G19450) with a QTL for 20:0 on chromosome 2 is not clear. The preference of DAGAT for a certain FA over others could affect the proportions of 20:0, respectively (Table 2). Furthermore, we observed that on chromosomes 2, 3 and 4, the CGs collocated within 2 cM of the QTL peak, respectively. On chromosome 2, DAGAT collocated within 0.2 cM of the QTL peak for 20:0, while on chromosome 3, FAD2 and FAE1.1 collocated within 1–2 cM of the QTL peak for 18:1, 18:2 and 20:2, and FAE1.2 collocated at the QTL peak for 18:3, respectively. Similarly, FAE1.3 and FAE1.4 collocated within 0.4 cM of the QTL peak for 20:2, respectively (Table 4). In the *Ler*-0 × *Col*-4 population, 11 of the 23 QTL collocated with CGs on all five chromosomes. The co-location of 16:0 with KASII on chromosome 1 is relevant as it regulates the proportions of 16:0 and 18:0. Similarly, the co-location of FAD3 with QTL for 18:2 and 20:2 on chromosome 2 correlates with the desaturase activity of FAD3 (18:2–18:3 and 18:2 is also elongated to 20:2; thus, the proportions for 20:2 could also be affected). As seen in the *Ler*-0 × *Sha* population, in the *Ler*-0 × *Col*-4 population, the co-location of FAE1.1, FAE1.3, FAE1.4 and FAD2 with QTL for 18:1, 18:2 and 22:1 on chromosomes 3 and 4 and co-location of SAD1 and SAD2 with a QTL for 18:0 on chromosome 5 was observed. In all the cases, the co-locations make sense with respect to functional relevance (Tables 2, 5). The co-location of a QTL for 22:1 with FATB (AT1G08510) cannot be explained. None of QTL in the *Ler*-2 × *Cvi* and *Ler*-0 × *No*-0 populations collocated with CGs. However, wherever co-location of CGs was not observed, other genes, which have putative function related to lipid synthesis, were seen (Tables 6, 7).

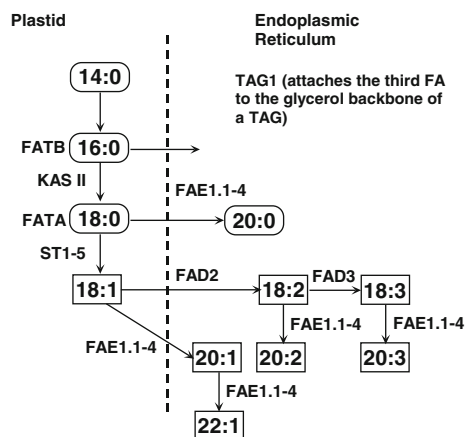
#### Genetic correlation between FA traits

Associations amongst FAs are subject to autocorrelation because they are expressed as percentages of total identified FAs. The correlation was also greater between FAs when one was included in a composite trait, thus we



**Table 2** Candidate fatty acid synthesis genes in *A. thaliana* and their functions

Candidate genes	Function
KASII	Elongates 14:0–16:0 and 18:0
FATB	Terminates elongation of 16:0
FATA	Terminates elongation at 18:0
Stearoyl-ACP desaturase (5 copies: SAD1–5)	Desaturates 18:0–18:1
FAD2	Desaturates 18:1–18:2
FAD3	Desaturates 18:2–18:3
FAE1 (4 copies: FAE1.1–1.4)	Elongates both saturated and unsaturated C18 fatty acids to C20 and C22
DAGAT (or TAG1)	Attaches final fatty acid to diacylglycerol to form a TAG

**Fig. 1** Relevant FA and TAG synthesis pathways in *A. thaliana*. FAs are boxed with saturated fatty acids having rounded corners. The distinction between the plastids and the ER is to designate where the reactions occur. 16:0, 18:0 and 18:1 are all exported to the ER and are incorporated into TAGs. Some 18:2 and 18:3 are synthesized in the plastids but these are not exported to the ER

observed that  $r$  was 0.935 between 16:0 and saturated FAs, which are comprised of 16:0, 18:0 and 18C saturated FAs, respectively. In the biosynthetic pathway, the 14C, 16C and 18C saturated FAs are converted to 18C, 20C, and 22C unsaturated FAs, and so the two types of FAs would be expected to be negatively correlated. As expected, the results indicate significant positive correlations between saturated FAs (16:0, 18:0 and 20:0, Fig. 1; Tables S1–S4) and negative correlations between saturated and unsaturated FAs (e.g., between 18:0 and 18:2, Tables S1–S4). As expected, negative correlations between unsaturated FAs were also observed when one of the FAs is the substrate and the other FA is the product of an enzyme reaction (for e.g., 18:2 and 18:3, Fig. 1; Table S1) in a metabolic pathway. In some cases, the correlations are hard to interpret because the up-regulation or down-regulation of the gene could determine the correlations. Further, in the *Ler-0* × *Sha* population, the four MP QTL on chromosomes 1, 3, 4, and 5 co-locate with two to eight FA QTL for all nine individual

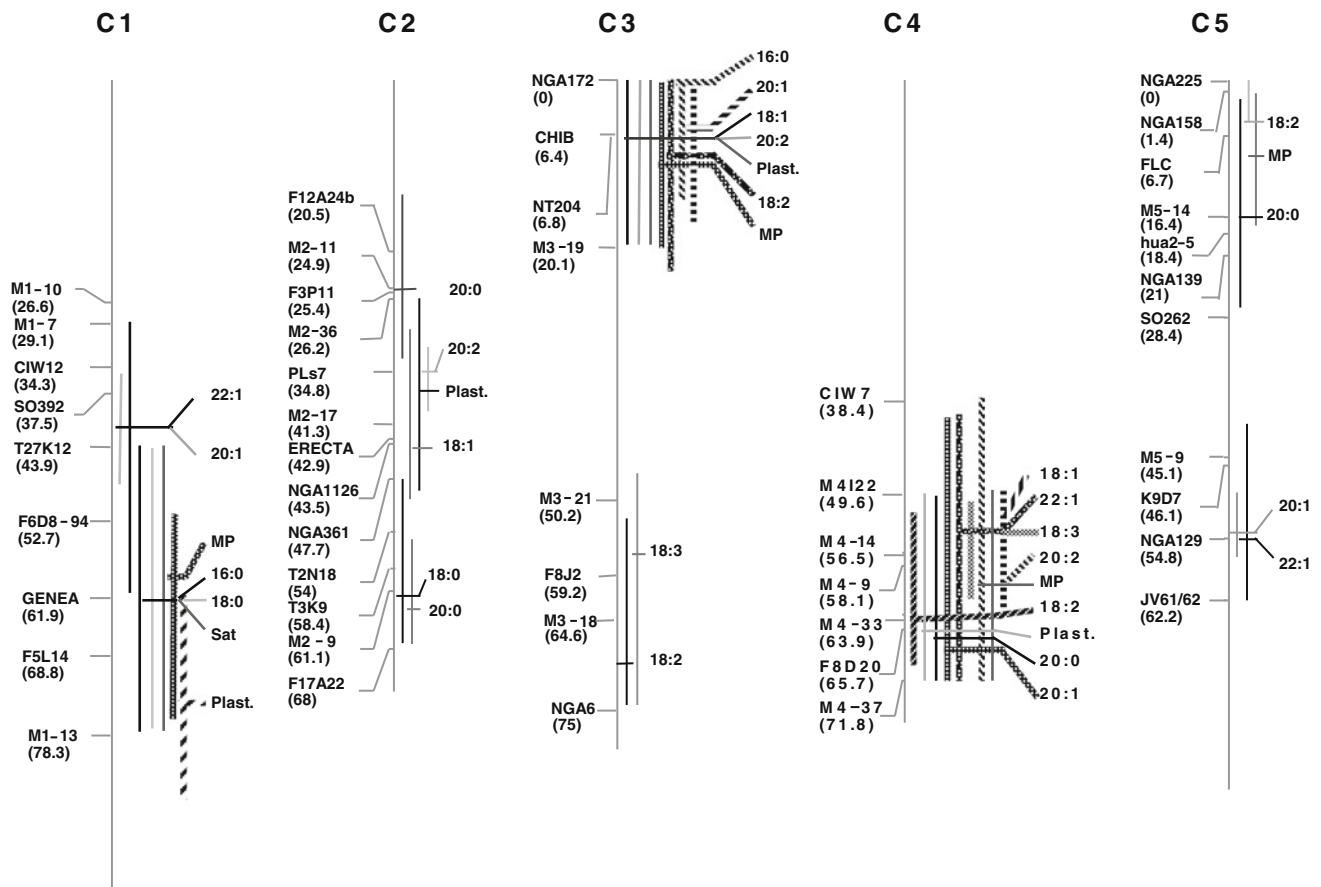
FAs. Significant correlations between MP and all individual FAs with the exception of 22:1 were observed indicating that these correlations with the individual FAs will affect the MP in the *Ler-0* × *Sha* population only. In the *Ler-0* × *Col-4* population, only 22:1 co-locates with MP QTL, but the correlation was not significant and will not affect the MP.

#### QTL overlap between populations

Three QTL for 18:1, 18:2 and FAs synthesized in plastids on chromosome 3 and one QTL for 22:1 on chromosome 4 (Tables 4, 5; Figs. 2, 3) were identified in both *Ler-0* × *Sha* and *Ler-0* × *Col* populations. For all four QTL, the effects of the ‘*Ler*’ allele in the *Ler-0* × *Sha* and *Ler-0* × *Col-4* populations were in the opposite directions suggesting considerable variation at these loci. The *Sha* and *Col-4* alleles are quite different where the differences in the relative proportions of the FAs in the two ecotypes range from 1.6 to 5.4%, respectively (Table 3). All three QTL on chromosome 3 in both the populations collocated with two relevant CGs, FAE1.1 and FAD2, and one QTL on chromosome 4 collocated with two relevant CGs, FAE1.3 and FAE1.4, as aforementioned (Tables 4, 5). In addition, a QTL for 16:0 was also detected on chromosome 3 in the *Ler-0* × *Sha* and *Ler-0* × *No-0* RIL populations. In this case, the *Ler* allele was the “high” allele in both populations and did not co-locate with any CG.

#### Cytoplasmic effects and nuclear-cytoplasmic interactions

We found a significant cytoplasmic effect and a nuclear-cytoplasmic interaction only in the *Ler-2* × *Cvi* population for the QTL: LCv4.16b, for trait 16:0 on chromosome 4 (Table 6). A cytoplasmic effect was detected for the *Ler* cytoplasm ( $p = 0.027$ ), which is interesting because 16:0 is synthesized in the plastid. The nuclear-cytoplasmic interaction ( $p = 0.007$ ) might have been caused because 16:0 is synthesized in the plastid but the enzymes for its synthesis



**Fig. 2** Major QTL for the 12 seed oil composition traits in the *Ler-0* × *Col* RIL population. C1 to C5 indicate chromosome numbers. On the left hand, just the markers which overlap with the 95% confidence interval (CI) of one or more QTL peaks are listed. The *first line* represents the name of the marker and the *second line* is the position of the marker on the chromosomes in cM. On the *right*

*side* of the chromosome, the *vertical lines* indicate the 95% CI of the major QTL. The *horizontal lines* indicate the position of the maximum height of the QTL peak and they have the trait associated with them on the *right*. *Horizontal and vertical bar patterns* match each other except when multiple peaks share same peak position

are nuclear encoded. However, it is not clear why other plastid-synthesized FAs (18:0, 18:1) do not show nuclear-cytoplasmic interactions.

### Epistasis

Three significant epistatic interactions in the *Ler-0* × *Sha* and *Ler-0* × *Col-4* populations were detected suggesting that epistasis is not generally common in TAG synthesis. In *Ler-0* × *Sha*, epistatic interactions between two QTL for 20:1 and two QTL for 20:2 on chromosomes 3 and 4 (Table 4) were observed. The ‘*Ler*’ allele was the high allele for both the QTL for 20:1 suggesting that the same regulatory element could be controlling both. For 20:2, the ‘*Ler*’ allele was the low allele for the QTL on chromosome 3 and high allele for the QTL on chromosome 4, suggesting that there could be some regulation of the total amount of the FA. In the *Ler-0* × *Col-4* population, epistatic interactions were observed between two QTL for 18:1 on

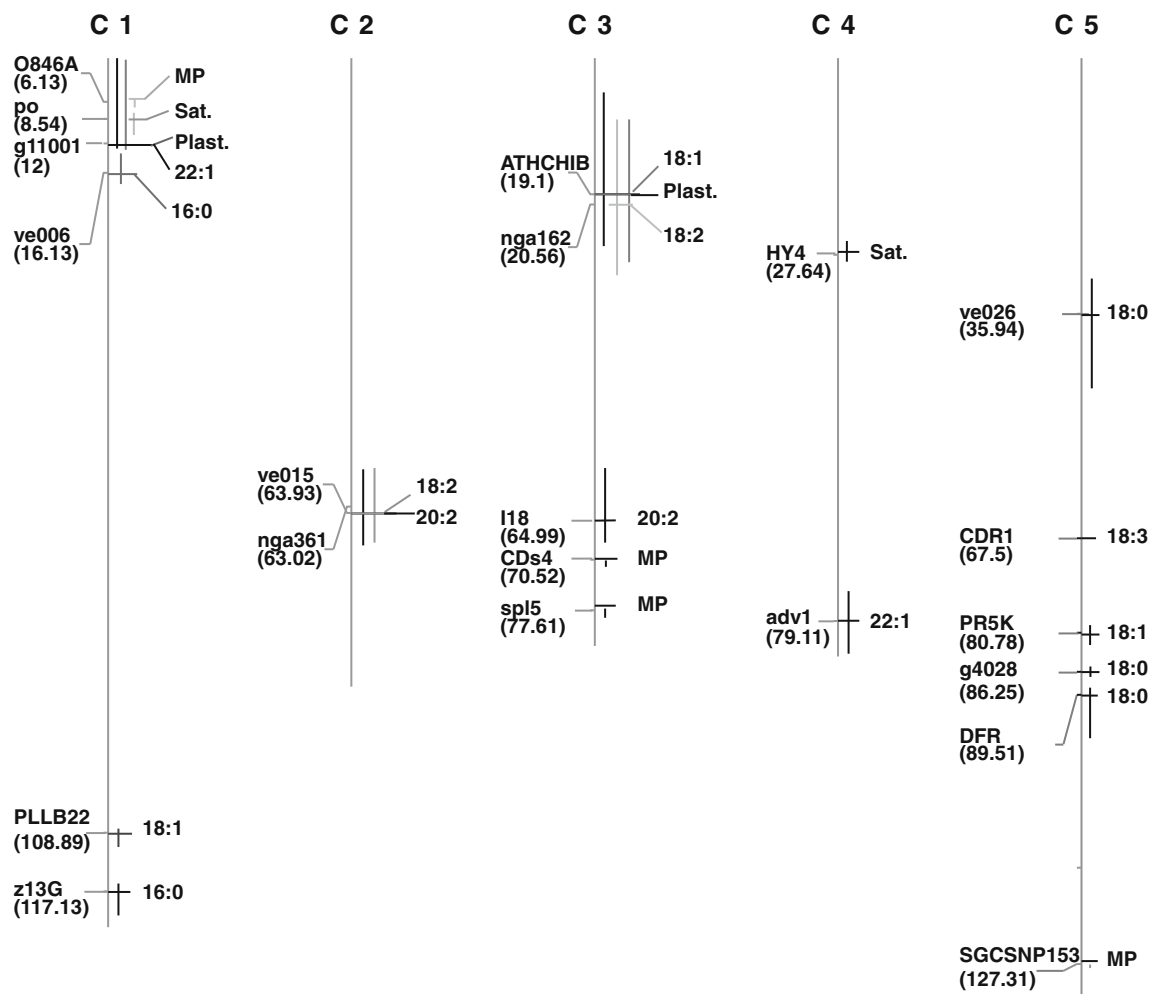
chromosomes 1 and 5 (Table 5). The ‘*Ler*’ allele was the low allele in both the QTL.

### Discussion

The mapping of 72 QTL for FA composition in seed TAGs in the four RIL populations identified and characterized several new genomic regions that contribute to regulating TAG FA proportions in *A. thaliana*. The most remarkable was the co-location of nine FA QTL at the bottom of chromosome 4 between 38 and 71.8 cM and the co-location of seven individual FA QTL at the top of chromosome 3 between 0 and 22.8 cM in the *Ler-0* × *Sha* RIL population.

### Comparison with previous work

Prior work (Hobbs et al. 2004) identified seven QTL in the *Ler-2* × *Cvi* RIL population. Interestingly, the seven QTL



**Fig. 3** Major QTL for the 12 seed oil composition traits in the *Ler-0* × *Col* RIL population. To interpret this figure see the legend for Fig. 2

detected in our study for the same RIL population were different from those of Hobbs et al. (2004). Two out of seven QTL—one for 18:0 and another for 18:3 on chromosomes 1 and 2, respectively—correspond to the same locations in both the studies but with small effect in this study. Some of the differences may simply be due to the strength of the QTL peaks in the two studies. However, since Hobbs et al. did not report QTL of small effect, it is unknown whether any of our QTL with major effect co-locates with the QTL of small effect detected in their studies.

The differences in QTL presence and absence in the two studies and the differences in whether QTL were of major or minor effect could have several causes. Genotype × environment interactions ( $G \times E$ ) could have created different phenotypes due to both differences in the growing environment and the genotypes used in the two studies (Ungerer and Rieseberg 2003). Only nine lines were common to both studies. The differences in QTL detected might also be, partly, due to cytoplasmic effects

because all the 80 lines used in our study had the *Ler* cytoplasm, whereas for Hobbs et al. (2004), only 25 of the 50 lines had the *Ler* cytoplasm.

#### Shared QTL between populations

The map positions of three QTL for 18:1, 18:2, and FAs present in plastids on chromosome 3 and one QTL for 22:1 on chromosome 4 showed strong overlap in the two genomic regions in the *Ler-0* × *Sha* and *Ler-0* × *Col-4* populations. In addition, the co-location of the same relevant CGs with the QTL in both the populations (FAE1.1 and FAD2 on chromosome 3 and FAE1.3 and FAE1.4 on chromosome 4, Tables 4, 5) is significant. Two QTL for 16:0 on chromosome 3 in the *Ler-0* × *Sha* and *Ler-0* × *No-0* populations shows an overlap but do not co-locate with CGs (Tables 4, 7). However, several other FA lipid synthesis genes and transcription factors also co-locate with these QTL peaks. Hence, the QTL may have been mapped due to variation at the same or different loci. It is possible that at each locus



**Table 3** Fatty acid composition of seed oils of the four RIL populations

	16:0	18:0	18:1	18:2	18:3	20:0	20:1	20:2	22:1	Sat	FAP	MP
<i>Ler-0</i>	0.068	0.037	0.193	0.276	0.167	0.017	0.207	0.014	0.019	0.123	0.298	12.19
Sha	0.057	0.019	0.168	0.374	0.182	0.017	0.147	0.019	0.016	0.094	0.245	7.76
Mean	0.082	0.025	0.196	0.324	0.170	0.016	0.175	0.017	0.017	0.101	0.281	10.986
Min value	0.053	0.016	0.142	0.260	0.116	0.011	0.135	0.009	0.011	0.034	0.226	13.761
Max value	0.099	0.033	0.279	0.399	0.207	0.026	0.219	0.026	0.026	0.142	0.347	9.137
SD	0.007	0.003	0.033	0.031	0.018	0.002	0.022	0.005	0.003	0.012	0.032	1.010
SE	0.001	0.000	0.003	0.003	0.002	0.000	0.002	0.000	0.000	0.001	0.003	0.103
Variance	0	0	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	1.021
<i>Ler-0</i>	0.068	0.037	0.193	0.276	0.167	0.017	0.207	0.014	0.019	0.123	0.298	12.19
Col-4	0.089	0.044	0.155	0.318	0.206	0	0.147	0.021	0.019	0.133	0.288	11.44
Mean	0.082	0.035	0.182	0.291	0.193	0.006	0.177	0.016	0.017	0.123	0.299	12.127
Min value	0.069	0.028	0.085	0.250	0.155	0.000	0.104	0.009	0.007	0.102	0.205	9.905
Max value	0.099	0.052	0.236	0.329	0.238	0.020	0.203	0.030	0.035	0.148	0.371	14.611
SD	0.007	0.005	0.023	0.018	0.017	0.006	0.018	0.003	0.004	0.011	0.024	0.945
SE	0.001	0.000	0.002	0.002	0.002	0.001	0.002	0.000	0.000	0.001	0.003	0.099
Variance	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.892
<i>Ler-2</i>	0.064	0.043	0.218	0.251	0.166	0.023	0.208	0.012	0.016	0.129	0.324	12.73
Cvi	0.071	0.032	0.139	0.318	0.159	0.023	0.204	0.028	0.025	0.127	0.243	11.35
Mean	0.069	0.034	0.191	0.295	0.154	0.020	0.202	0.016	0.020	0.123	0.294	13.430
Min value	0.051	0.000	0.124	0.229	0.093	0.000	0.130	0.000	0.000	0.101	0.237	9.404
Max value	0.108	0.045	0.284	0.355	0.216	0.032	0.235	0.028	0.036	0.148	0.401	14.823
SD	0.011	0.006	0.030	0.032	0.026	0.007	0.021	0.006	0.007	0.010	0.032	0.820
SE	0.001	0.001	0.003	0.004	0.003	0.001	0.002	0.001	0.001	0.001	0.004	0.092
Variance	0.000	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.673
<i>Ler-0</i>	0.068	0.037	0.193	0.276	0.167	0.017	0.207	0.014	0.019	0.123	0.298	12.19
No-0	0.069	0.038	0.186	0.267	0.184	0.032	0.187	0.017	0.019	0.139	0.293	11.57
Mean	0.070	0.035	0.194	0.271	0.177	0.027	0.193	0.015	0.018	0.132	0.298	13.746
Min value	0.059	0.025	0.143	0.247	0.136	0.000	0.144	0.009	0.012	0.095	0.228	10.841
Max value	0.094	0.042	0.260	0.318	0.229	0.037	0.233	0.020	0.029	0.147	0.375	14.743
SD	0.006	0.003	0.022	0.013	0.016	0.008	0.014	0.002	0.003	0.011	0.023	0.695
SE	0.001	0.000	0.003	0.002	0.002	0.001	0.002	0.000	0.000	0.001	0.003	0.087
Variance	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.483

there might be as many as three or as few as two alleles segregating among the parental accessions. One explanation could be that the *Ler-0* allele significantly differs in effect from the Sha, Col-4, and No-0 alleles.

#### Population-specific QTL

The mapping of different QTL in different populations (Tables 4, 5, 6, 7; Figs. 2, 3, 4, 5) in this study is likely due to an accession-specific allele as otherwise the QTL would have been identified in all four populations since all of them share the *Ler* accession as the parent. These four populations mimic a set of test crosses. Therefore, the Sha, Col-4, Cvi, and No-0 alleles must be very different from the *Ler* allele and hence, any QTL identified in one population only indicates that the allele was from the

“non-*Ler*” accession of the population from which the parent was mapped. Thus, this design proved exquisitely useful for identifying novel alleles.

Differences in the QTL identified for the FA traits in the four RIL populations could be due to segregation of meaningful genetic variation among natural lineages as there was no environmental variation between populations. It was also seen that the majority of genetic variation in TAGs in *A. thaliana* is additive in nature.

#### Candidate genes and FA regulation

The complete genome sequence of *A. thaliana* and knowledge of the seed oil biosynthetic pathway in *A. thaliana* helped identify CGs near the QTL. A total of 22 of the 36 QTL collocated with CGs in the *Ler-0* × Sha RIL

**Table 4** Quantitative trait loci detected for the 12 seed oil composition traits in the *Ler-0* × Sha RIL population

QTL	LG <sup>a</sup>	Trait	LOD score	Peak max <sup>b</sup> (cM)	Closest marker	CI <sup>c</sup> (cM)	Additive <sup>d</sup> effect	Variance <sup>e</sup> (%)	Candidate genes	Allele effect	Lipid genes <sup>f</sup>
LS1.16	1	16:0	8.28	62	Genea	43.7–77.9	0.53	0.255		Ler↑	49
LS1.18	1	18:0	7.24	62	Genea	44–77.5	0.37	0.13		Ler↑	49
LS1.Sat	1	Saturated	10.18	62	Genea	43.7–77.8	0.57	0.315		Ler↑	49
LS1.MP	1	MP <sup>g</sup>	5.55	59	Genea	51.9–76	0.31	0.09		Ler↑	48
LS1.201	1	20:1	3.93	41.6	T27K12	35.1–48.3	−0.21	0.041		Ler↓	28
LS1.221	1	22:1	7.67	41.6	T27K12	28.9–61.3	−0.47	0.212		Ler↓	68
LS1.P	1	FA ≤ 18:1	4.13	74.2	M1-13	61.6–85.9	0.24	0.053		Ler↑	39
LS2.18	2	18:0	13.65	61.2	M2-9	47.7–67.3	0.59	0.336		Ler↑	31
LS2.20b	2	20:0	4.75	63.1	M2-9	54.9–67.4	0.42	0.167		Ler↓	29
LS2.202	2	20:2	3.77	34.8	PLs7	31.9–39.6	0.18	0.029		Ler↑	4
LS2.P	2	FA ≤ 18:1	6.33	36.9	PLs7	26.1–49.1	−0.28	0.075	FAD3	Ler↓	20
LS2.181	2	18:1	4.07	43.5	nga1126	29.8–50.1	−0.23	0.051		Ler↓	25
LS2.20a	2	20:0	3.74	24.8	M2-11	13.7–33.3	0.42	0.121	DAGAT	Ler↑	19
LS3.181	3	18:1	23.8	6.5	CHIB	0–19.7	0.79	0.578	FAE1.1, FAD2	Ler↑	59
LS3.201	3	20:1	6.39	6	CHIB	0–17.1	0.28	0.075	FAE1.1, FAD2	Ler↑	55
LS3.202	3	20:2	29.5	6.5	CHIB	0–19.7	−0.81	0.615	FAE1.1, FAD2	Ler↓	59
LS3.P	3	FA ≤ 18:1	23.9	6.5	CHIB	0–19.7	0.76	0.528	FAE1.1, FAD2	Ler↑	59
LS.182a	3	18:2	17.7	8.9	nt204	0–22.8	−0.75	0.498	FAE1.1, FAD2	Ler↓	59
LS3.MP	3	MP <sup>g</sup>	9.79	10.9	nt204	2.4–19.8	0.51	0.223	FAE1.1, FAD2	Ler↑	53
LS3.16	3	16:0	3.65	0.1	nga172	0–14.3	−0.33	0.103	FAE1.1, FAD2	Ler↓	55
LS3.182b	3	18:2	4.61	69.1	M3-18	52.4–74.7	−0.20	0.04	FAE1.2	Ler↓	27
LS3.183	3	18:3	4.79	56.2	F8J2	47–74.7	0.42	0.168	FAE1.2	Ler↑	27
LS4.181	4	18:1	4.57	52	M4122	49.1–63.4	−0.26	0.065	FAE1.3, FAE1.4	Ler↓	34
LS4.182	4	18:2	5	64	M4-33	51.7–69.9	−0.22	0.041	FAE1.3, FAE1.4	Ler↓	34
LS4.183	4	18:3	4.01	53.7	M4-14	50.4–62	−0.37	0.123	FAE1.3, FAE1.4	Ler↓	34
LS4.221	4	22:1	10.27	53.7	M4-14	40–71.8	0.58	0.30	FAE1.3, FAE1.4	Ler↑	34
LS4.202	4	20:2	11.55	60.3	M4-9	38–71.8	0.34	0.112	FAE1.3, FAE1.4	Ler↑	34
LS4.MP	4	MP <sup>g</sup>	13.02	60.3	M4-9	49–71.8	0.55	0.286	FAE1.3, FAE1.4	Ler↑	34
LS4.20	4	20:0	4.67	65.9	F8D20	49.7–71.8	0.36	0.122	FAE1.3, FAE1.4	Ler↑	34
LS4.201	4	20:1	34.42	67.7	F8D20	40.4–71.8	0.91	0.798	FAE1.3, FAE1.4	Ler↑	34
LS4.P	4	FA ≤ 18:1	6.39	65.8	F8D20	49.4–71.8	−0.29	0.084	FAE1.3, FAE1.4	Ler↓	34
LS5.182	5	18:2	4.45	5	Flc	0–5	−0.19	0.034		Ler↓	14
LS5.MP	5	MP <sup>g</sup>	6.97	9.1	Flc	1.6–17.4	0.34	0.113	SAD1, SAD2	Ler↑	29
LS5.201	5	20:1	3.14	54.1	nga129	49.3–57	−0.19	0.032		Sha↓	19
LS5.221	5	22:1	7.87	54.9	nga129	41.1–62.2	−0.47	0.198	SAD1, SAD2	Ler↓	39
LS5.20	5	20:0	5.93	16.4	M5-14	2.3–27.2	0.39	0.144		Ler↑	33

<sup>a</sup> LG = Linkage group<sup>b</sup> Position of the highest point of the QTL peak<sup>c</sup> 95% confidence interval of the QTL peak<sup>d</sup> Estimate of the change of the average phenotype produced by substituting a single allele of one type with that of another type<sup>e</sup> The proportion of the phenotypic variance explained by the QTL<sup>f</sup> The number of putative lipid synthesis genes collocating with a QTL peak<sup>g</sup> Estimated melting point of the triacylglycerols

population, and 11 of the 23 QTL in the *Ler-0* × Col-4 RIL population collocated with CGs. Most of the co-locations with QTL for individual FAs were functionally relevant

with the exception of the co-locations of FATB with a QTL for 22:1 on chromosome 1 in the *Ler-0* × Col-4 RIL population, and the co-location of FAE1.1 and

**Table 5** Quantitative trait loci detected for 12 seed oil composition traits in the *Ler-0* × *Col-4* RIL population

QTL	LG <sup>a</sup>	Trait	LOD score	Peak max <sup>b</sup> (cM)	Closest marker	CI <sup>c</sup> (cM)	Additive effect (%) <sup>d</sup>	Variance <sup>e</sup> (%)	Candidate genes	Allele effect	Lipid genes <sup>f</sup>
LC.221	1	22:1	3.24	12.2	g11001	0–12.7	−0.42	0.12	FATB	Ler↓	28
LC1.P	1	FA ≤ 18:1	3.43	12.2	g11001	0.2–12.9	0.29	0.07	FATB	Ler↑	27
LC1.Sat	1	Saturated	3.32	8.62	po	7.7–10.8	0.41	0.13	FATB	Ler↑	16
LC1.16a	1	16:0	3.6	16.3	ve006	13.4–17.7	0.37	0.128	KASII	Ler↑	28
LC1.16b	1	16:0	3.81	117.2	mi425	116–120.5	0.40	0.141		Ler↑	22
LC1.181	1	18:1	3.20	109	PLL22	108.3–110.9	−0.35	0.074		Ler↓	2
LC1.MP	1	MP <sup>e</sup>	4.07	6.2	O846A	5.8–6.9	0.61	0.114		Ler↑	1
LC2.182	2	18:2	5.63	64	ve015	57.8–68.5	0.36	0.117	FAD3	Ler↑	17
LC2.202	2	20:2	4.10	64	ve015	57.6–68.1	0.34	0.101	FAD3	Ler↑	17
LC3.181	3	18:1	13.32	19.1	ATHCHIB	4.8–26.4	−0.66	0.413	FAE1.1, FAD2	Ler↓	41
LC3.P	3	FA ≤ 18:1	11.83	19.1	ATHCHIB	8.6–28.7	−0.59	0.321	FAE1.1, FAD2	Ler↓	36
LC3.182	3	18:2	13.47	20.6	nga162	8.6–30.5	0.59	0.338	FAE1.1, FAD2	Ler↑	37
LC3.202	3	20:2	12.55	64	I18	57.6–68.1	0.34	0.101		Ler↑	10
LC3.MPa	3	MP <sup>e</sup>	3.21	70.6	CDs4	70.6–71.5	−0.41	0.095		Ler↓	3
LC3.MPb	3	MP <sup>e</sup>	3.1	77.7	sp15	77.6–78.9	−0.41	0.092		Ler↓	1
LC4.221	4	22:1	4.53	79.1	adv1	74.9–83.7	−0.43	0.12	FAE1.3, FAE1.4	Ler↓	17
LC4.Sat	4	Saturated	3.9	27.7	HY4	25.7–28.6	0.46	0.134		Ler↑	2
LC5.18a	5	18:0	7.58	36.1	ve026	31–46.4	0.58	0.277	SAD1, SAD2	Ler↑	19
LC5.18b	5	18:0	3.27	86.3	g4028	85.5–87	−0.30	0.108		Ler↓	6
LC5.18c	5	18:0	3.43	89.6	DFR	88.5–95.6	−0.30	0.112		Ler↓	7
LC5.181	5	18:1	4	81	PR5K	79.7–82.5	−0.41	0.099		Ler↓	0
LC5.183	5	18:3	3.15	67.6	CDR1	67.4–67.9	−0.49	0.079		Col↑	1
LC5.MP	5	MP <sup>e</sup>	3.6	127.4	SGCSNP153	127.4–127.9	0.38	0.102		Col↑	1

See the legend for Table 4

**Table 6** Quantitative trait loci detected for the 12 seed oil composition traits in the *Ler-2* × *Cvi* RIL population

QTL	LG <sup>a</sup>	Trait	LOD score	Peak max <sup>b</sup> (cM)	Closest marker	CI <sup>c</sup> (cM)	Additive <sup>d</sup> effect	Variance <sup>e</sup> (%)	Allele effect	Lipid genes <sup>f</sup>
LCv1.16	1	16:0	3.10	50	GB.112L	43.7–54.4	0.37	0.126	Ler↑	4
LCv1.201	1	20:1	3.17	51.1	GB.112L	44.6–54.6	−0.40	0.137	Ler↓	4
LCv2.MP	2	MP <sup>f</sup>	4.79	16	CH.266C	14.8–18.4	0.47	0.185	Ler↑	1
LCv4.16a	4	16:0	3.61	19.8	CC.288L	9.3–23.1	0.48	0.156	Cvi↑	22
LCv4.16b	4	16:0	5.07	58.2	B9-1.8	51.1–68.9	−0.48	0.196	Ler↓	20
LCv5.20	5	20:0	3.5	73.8	DFR	69.2–79.2	−0.41	0.153	Cvi↓	6
LCv5.P	5	FA ≤ 18:1	3.93	72.8	DFR	65.4–77.8	0.65	0.181	Cvi↑	6

See the legend for Table 4

FAD2 with a QTL for 16:0 in the *Ler-0* × *Sha* RIL population. We observed six functionally relevant CGs (FAE1.1, FAE1.3, FAE1.4, FAD2, SAD1, and SAD2), which co-located with the same QTL in the two populations (Tables 4, 5). We also observed that in *Ler-0* × *Sha* RIL population, the CGs FAE1.1, FAE1.2, FAE1.3, FAE1.4, FAD2 and DAGAT collocate within 1–2 cM of some of the associated peaks. In the *Ler-0* × *Col-4* RIL

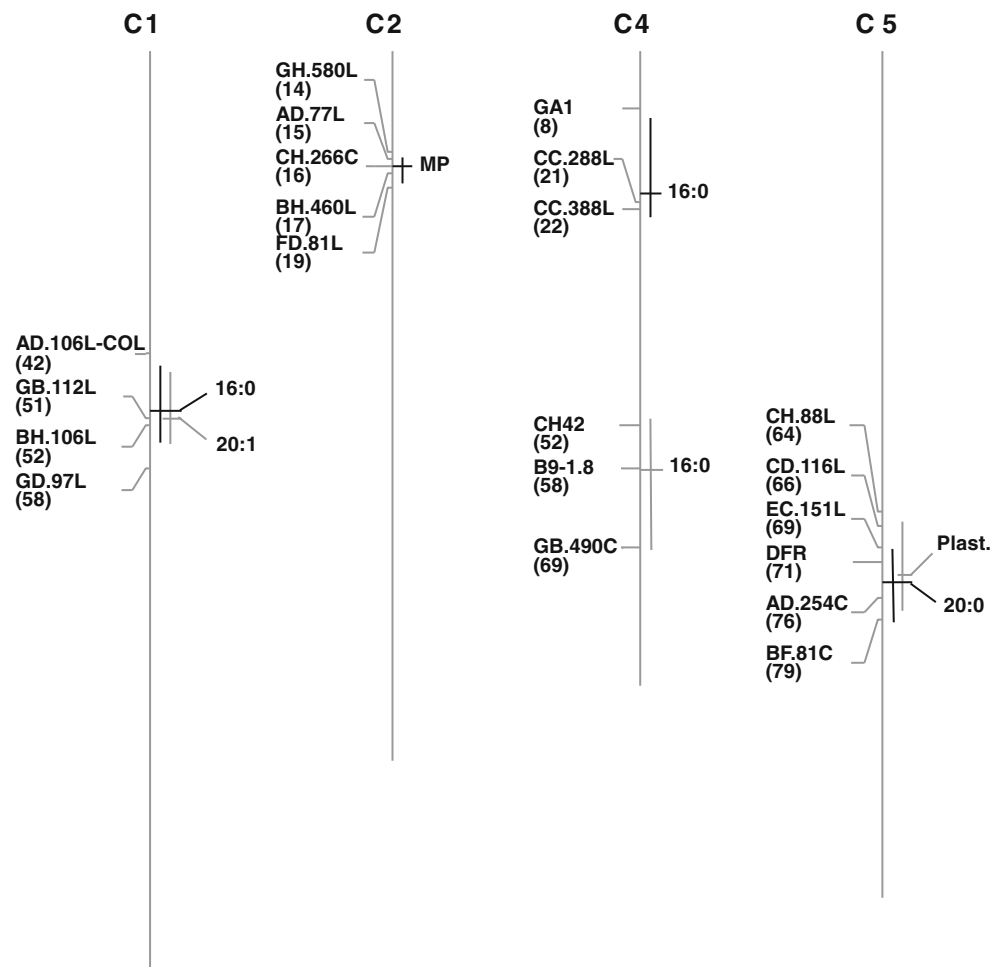
population, KASII collocates at the QTL peak for 16:0 on chromosome 1 respectively. Previous studies have shown that QTLs might be accurate to within 2 cM or less (Price 2006) which makes these very promising candidate genes to study. In addition, all other QTL in all the four populations, which did not co-locate with CGs collocated with one or more lipid synthesis genes and transcription factors with the exception of QTL LC5.181 in the *Ler-0* × *Col-4*

**Table 7** Quantitative trait loci detected for the 12 seed oil composition traits in the *Ler-0* × No-0 RIL population

QTL	LG <sup>a</sup>	Trait	LOD score	Peak max <sup>b</sup> (cM)	Closest marker	CI <sup>c</sup> (cM)	Additive <sup>d</sup> effect	Variance <sup>e</sup> (%)	Allele effect	Lipid genes <sup>f</sup>											
LN1.16	1	16:0	4.28	68	nga111	55.7–70	0.51	0.234	Ler↑	19											
LN2.183	2	18:3	4.57	20.1	MSAT2.38	3–25.8	0.48	0.218	Ler↑	11											
LN3.16	3	16:0	3.17	0.2	nga172	0–6.8	−0.45	0.139	Ler↓	12											
LN3.Sat	3	Saturated	3.28	1.3	nga172	0–7.6	−0.48	0.182	Ler↓	12											
LN4.183	4	18:3	6.36	39	CIW6	33.8–56.9	−0.63	0.218	Ler↓	49	LN5.202	5	20:2	4.9	0	nga225	0–8.2	−0.53	0.227	Ler↓	8
LN5.202	5	20:2	4.9	0	nga225	0–8.2	−0.53	0.227	Ler↓	8											

See the legend for Table 4

**Fig. 4** Major QTL for the 12 seed oil composition traits in the *Ler-2* × Cvi RIL population. To interpret this figure see the legend for Fig. 2

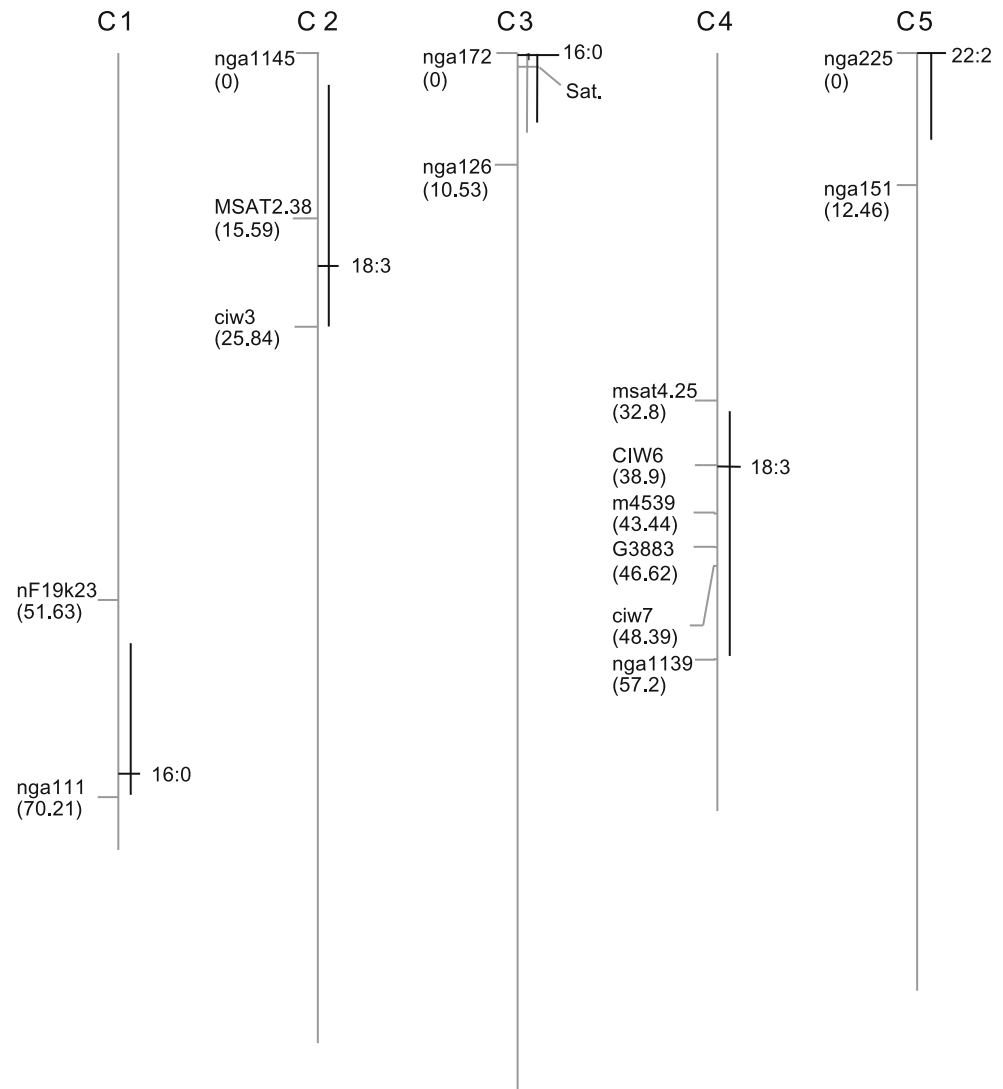


RIL population. Recent studies show that mutations in the regulatory regions and protein-coding changes in transcription factors contribute more often to quantitative trait variation (King and Wilson 1975; Mackay 2001; Korstanje and Paigen 2002; Hsia and McGinnis 2003). Determination of the source of genetic variation in the seed oil composition will require the identification and molecular characterization of the identified QTL and the underlying CGs.

#### Implications for evolutionary studies

Natural variation in seed oil composition in *A. thaliana* is extensive (O'Neill et al. 2003). While all *A. thaliana* populations studied to date have the same set of FAs in their seed oils, the relative proportions of those FAs vary widely. For example, 360 accessions surveyed for very long chain fatty acids (VLCFAs) and polyunsaturated fatty acids (PUFAs) in their seeds (O'Neill et al. 2003) had

**Fig. 5** Major QTL for the 12 seed oil composition traits in the *Ler-0* × No-0 RIL population. To interpret this figure see the legend for Fig. 2



relative proportions of VLCFAs ranging from 13 to 21.2% of total fatty acids and relative proportions of PUFAs from 53.3 to 66.1%. Prior work with 84 accessions in *A. thaliana* has also shown that there is extensive variation in the proportions of 18:1 (5–17.8%), 18:2 (23–37.8%), 18:3 (15–27%), 20:1 (16.4–27.5%) and 16:0 (5.5–10.2%) acids while the variation in the proportions of 18:0, 20:0, 20:2 and 22:1 were <5% (Sanyal and Linder, manuscript submitted). It has been observed that a greater proportion of unsaturated FAs are present at higher latitudes and cooler temperatures, so that seeds with more unsaturated oils germinate earlier and gain competitive advantage (Linder 2000). Here we observe that the composite FA measures of the total proportion of saturated FAs (which includes most of the high melting point FAs), the MPs of the seed oils and the individual FA 22:1 in 84 European accessions of *A. thaliana* have been shown to follow a strong latitudinal cline (Sanyal and Linder, manuscript submitted) similar to what has been demonstrated in other species and

taxonomic groups (Linder 2000). *Arabidopsis thaliana*'s latitudinal cline in seed oil MPs and total saturated fats was mostly due to combinations of FAs and not one or two dominant FAs. Discovery of QTL affecting the MP of seed TAGs in three of the four RIL populations has important evolutionary implications. There is evidence that the MP of seed TAGs has been under selection (Linder 2000), so identification of QTL that affect TAG MPs is a first step towards the discovery of these loci in the wild. The MP QTL in the *Ler-0* × Sha RIL population is particularly important because they co-locate with several FA QTL and their significant correlations could be used to explore how selection might operate on those FA loci and affect MP (Table 8). The genetic correlations between FAs whose QTL co-locate with MP QTL appear to generally enhance selection for altered TAG melting point (Table S1, Table 8). Further, since we did not detect any significant epistatic interactions among non-overlapping QTL, cytoplasmic effects or nuclear-cytoplasmic interactions; there



**Table 8** Net effect of significant genetic correlations between FAs on selection for altered TAG melting points

FA	<i>Ler-0</i> × <i>Sha</i>		<i>Ler-0</i> × <i>Col-4</i>	
	Effect of correlations on selection		Effect of correlations on selection	
	Weighted <sup>a</sup>	Unweighted	FA	Unweighted
16:0	−0.968	0.484	22:1	−0.058
18:0	0.704	0.704		
18:1	0.438	0.219		
18:2	5.639	1.880		
18:3	−0.053	−0.053		
20:0	0.624	0.312		
20:1	2.787	1.394		
20:2	0.227	0.114		
22:1	0.079	0.040		

FAs listed had QTL that co-located with at least one melting point QTL

<sup>a</sup> Weighted values were multiplied by the number of times the FA QTL co-located with the melting point QTL

should not be strong constraints or complicating factors on selection at particular loci from other QTL that we detected. Finally, there could be some constraints on the ability of selection to act for altered TAG melting due to the presence of multiple overlapping FA QTL. Only one of the four MP QTL in the *Ler-0* × *Col-4* collocated with a QTL for 22:1, a minor component of *A. thaliana* seed oil and the correlation was not significant (Tables S1–S2) and the QTL for MP in the *Ler-2* × *Cvi* did not co-locate with any FA QTL. *Ler-0* × *Sha* is therefore a particularly rich source for future work on studying the evolution of seed oil composition.

#### Implications of breeding seed oil crops

The identification of these FA syntheses QTL in the four RIL populations will enable marker-assisted selection to move beneficial QTL alleles into elite agricultural genotypes in seed oil breeding programs especially in the *Brassica* oilseed crops to produce oil for the food industry and also for industrial products. The usefulness of the detailed information available for the *Arabidopsis* genome has helped in making progress in identifying possible CGs involved in the glucosinate pathway and flowering time in crop brassicas, FA desaturases (Hasan et al. 2008; Lagercrantz et al. 1996; Mekhedov et al. 2000; Qiu et al. 2006), and therefore, will also be valuable in elucidating the seed oil biosynthetic pathway.

#### Conclusions

Our results have shown that natural variation for seed oil between five ecotypes (*Ler*, *Col-4*, *Sha*, *Cvi* and *No-0*) can be used to investigate the genetic control of seed oil composition as an alternative approach to screening mutants. The presence of seventy-two QTL for seed oil

composition provides new targets for the control of these traits. Characterization of five QTL observed in multiple populations and the QTL clusters identified especially in the *Ler-0* × *Sha* and *Ler-0* × *Col-4* RIL populations would be useful in elucidating the seed oil biosynthetic pathway. The identification of QTL for seed oil MP suggests that we can select for genotypes which will be adapted to different temperature conditions. Further, seed oil MP QTL co-locate with individual QTL suggesting that these QTL could be very useful for local adaptation and hence for breeding purposes. Several of these QTL map to genomic regions overlapping with several candidate genes on the basis of strong functional arguments and provide opportunity for further studies.

**Acknowledgments** We thank Dr. Thomas E. Juenger for his advice on experiments and comments on this manuscript. We thank Maarten Koornneef and Vaughan Symonds for sharing the mapping information of the *Ler-0* × *Sha* and *Ler-0* × *No-0* RIL populations. We thank Dr. Samuel Trachsel for his valuable comments on the manuscript. We also thank Anandita Agarwal, Petko Ivanov, Sandra Sipe and Sheila Shapouri for their help with data collection.

#### References

- Alonso-Blanco C, Peeters AJ, Koornneef M, Lister C, Dean C et al (1998) Development of an AFLP based linkage map of *Ler*, *Col* and *Cvi Arabidopsis thaliana* ecotypes and construction of a *Ler/Cvi* recombinant inbred line population. *Plant J* 14:259–271
- Barker GC, Larson TR, Graham IA, Lynn JR, King GJ (2007) Novel insights into seed fatty acid synthesis and modification pathways from genetic diversity and quantitative trait Loci analysis of the *Brassica C* genome. *Plant Physiol* 144:1827–1842
- Baud S, Lepiniec L (2009) Regulation of de novo fatty acid synthesis in maturing oilseeds of *Arabidopsis*. *Plant Physiol Biochem* 47:448–455
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol)* 57:289–300

- Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19:889–890
- Broun P, Gettner S, Somerville C (1999) Genetic engineering of plant lipids. *Annu Rev Nutr* 19:197–216
- Browse J, Somerville C (1991) Glycerolipid synthesis: biochemistry and regulation. *Annu Rev Plant Physiol Plant Mol Biol* 42:467–506
- Cahoon EB, Lindqvist Y, Schneider G, Shanklin J (1997) Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position. *Proc Natl Acad Sci USA* 94:4872–4877
- Canvin DT (1965) The effect of temperature on the oil content and fatty acid composition of the oils from several oil seed crops. *Can J Bot* 43:63–69
- Capelle V, Remoue C, Moreau L, Reyss A, Mahe A et al (2010) QTLs and candidate genes for desiccation and abscisic acid content in maize kernels. *BMC Plant Biology* 10:2
- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138:963–971
- Djemel N, Guedon D, Lechevalier A, Salon C, Miquel M et al (2005) Development and composition of the seeds of nine genotypes of the *Medicago truncatula* species complex. *Plant Physiol Biochem* 43:557–566
- Eckey EW (1954) Vegetable fats and oils. Reinhold, New York
- El-Lithy ME, Clercx EJ, Ruys GJ, Koornneef M, Vreugdenhil D (2004) Quantitative trait locus analysis of growth-related traits in a new *Arabidopsis* recombinant inbred population. *Plant Physiol* 135:444–458
- Gurr MI (1980) The biosynthesis of triacylglycerols. In: Stumpf PK, Conn EE (eds) *Lipids: structure and function*, Vol. 4. The biochemistry of plants: a comprehensive treatise. Academic Press, New York, pp 205–248
- Harris HC, McWilliam JR, Mason WK (1978) Influence of temperature on oil content and composition of sunflower seed. *Aust J Agr Res* 29:1203–1212
- Harwood JL (1980) Plant acyl lipids: structure, distribution, and analysis. In: Stumpf PK, Conn EE (eds) *Lipids: structure and function* Vol 4 of *The biochemistry of plants: a comprehensive treatise*. Academic Press, New York, pp 2–55
- Harwood JL (1996) Recent advances in the biosynthesis of plant fatty acids. *Biochim Biophys Acta* 1301:7–56
- Hasan M, Friedt W, Pons-Kühnemann J, Freitag NM, Link K, Snowdon RJ (2008) Association of gene-linked SSR markers to seed glucosinolate content in oilseed rape (*Brassica napus* ssp. *napus*). *Theor Appl Genet* 116:1035–1049
- Hilditch TP (1956) The chemical constitution of natural fats. Chapman and Hall, London
- Hilditch TP, Williams PN (1964) The chemical constitution of natural fats. Wiley, New York
- Hobbs DH, Flintham JE, Hills MJ (2004) Genetic control of storage oil synthesis in seeds of *Arabidopsis*. *Plant Physiol* 136:3341–3349
- Hsia CC, McGinnis W (2003) Evolution of transcription factor function. *Curr Opin Genet Dev* 13:199–206
- King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. *Science* 188:107–116
- Korstanje R, Paigen B (2002) From QTL to gene: the harvest begins. *Nat Genet* 31:235–236
- Lagercrantz U, Putterill J, Coupland G, Lydiat D (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine-scale genome collinearity and congruence of genes controlling flowering time. *Plant J* 9:13–20
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199
- Linder CR (2000) Adaptive evolution of seed oils in plants: accounting for the biogeographic distribution of saturated and unsaturated fatty acids in seed oils. *Am Nat* 156:442–458
- Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J* 4:745–750
- Lung S, Weselake RJ (2006) Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. *Lipids* 41:1073–1088
- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, Massachusetts
- Mackay TF (2001) The genetic architecture of quantitative traits. *Annu Rev Genet* 35:303–339
- Magliano TM, Botto JF, Godoy AV, Symonds VV, Lloyd AM et al (2005) New *Arabidopsis* recombinant inbred lines (*Landsberg erecta* × *Nossen*) reveal natural variation in phytochrome-mediated responses. *Plant Physiol* 138:1126–1135
- Malkin T (1954) The polymorphism of glycerides. In: Holman RT, Lundberg WO, Malkin T (eds) *Progress in the chemistry of fats and other lipids*. Pergamon, New York, pp 1–50
- Marsalkiene N, Sliessaravicius A, Karpaviciene B, Dastikaite A (2009) Oil content and fatty acid composition of seeds of some Lithuanian wild crucifer species. *Agron Res* 7:654–661
- Mekhedov S, Oskar Martínez de Ilárduya OM, Ohlrogge J (2000) Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis. *Plant Physiol* 122:389–401
- Metcalfe LD, Wang CN (1981) Rapid preparation of fatty acid methyl esters using organic base-catalyzed transesterification. *J Chromatogr Sci* 19:530–535
- Millar AA, Kunst L (1999) The natural genetic variation of the fatty-acyl composition of seed oils in different ecotypes of *Arabidopsis thaliana*. *Phytochemistry* 52:1029–1033
- O'Neill CM, Gill S, Hobbs D, Morgan C, Bancroft I (2003) Natural variation for seed oil composition in *Arabidopsis thaliana*. *Phytochemistry* 64:1077–1090
- Ohlrogge J, Browse J (1995) Lipid biosynthesis. *Plant Cell* 7:957–970
- Ohlrogge JB, Jaworski JG (1997) Regulation of fatty acid synthesis. *Annu Rev Plant Physiol Plant Mol Biol* 48:109–136
- Pelgas B, Bousquet J, Meirmans PG, Ritland K, Nathalie I (2011) QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. *BMC Genomics* 12:145
- Price AH (2006) Believe it or not, QTLs are accurate! *Trends Plant Sci* 11:213–216
- Qiu D, Morgan C, Shi J, Long Y, Liu J, Li R, Zhuang X, Wang Y, Tan X, Dietrich E, Weihmann T, Everett C, Vanstraelen S, Beckett P, Fraser F, Trick M, Barnes S, Wilmer J, Schmidt R, Li J, Li D, Meng J, Bancroft I (2006) A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. *Theor Appl Genet* 114:67–80
- Sharma N, Anderson M, Kumar A, Zhang Y, Giblin EM et al (2008) Transgenic increases in seed oil content are associated with the differential expression of novel Brassica-specific transcripts. *BMC Genomics* 9:619
- Storey JD, Tibshirani R (2003) Statistical methods for identifying differentially expressed genes in DNA microarrays. *Methods Mol Biol* 224:149–157
- Storey J, Taylor J, Siegmund D (2004) Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J R Stat Soc Ser B (Stat Methodol)* 66:187–205
- Stumpf PF (1980) Biosynthesis of saturated and unsaturated fatty acids. In: Stumpf PK, Conn EE (eds) *Lipids: structure and function*, edited by P. K. The biochemistry of plants: a comprehensive treatise, vol 4. Academic Press, New York, pp 177–204

- Symonds VV, Godoy AV, Alconada T, Botto JF, Juenger TE et al (2005) Mapping quantitative trait loci in multiple populations of *Arabidopsis thaliana* identifies natural allelic variation for trichome density. *Genetics* 169:1649–1658
- Systat software, Inc (2007) Systat 12 Statistics, Systat Inc. San Jose, CA
- Tonsor SJ, Alonso-Blanco C, Koornneef M (2005) Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. *Plant Cell Environ* 28:2–20
- Ungerer MC, Rieseberg LH (2003) Genetic architecture of a selection response in *Arabidopsis thaliana*. *Evolution* 57:2531–2539
- Wang S, Basten CJ, Zeng ZB (2007) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC (<http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>)
- Welch RM, House WA, Beebe S, Cheng Z (2000) Genetic selection for enhanced bioavailable levels of iron in bean (*Phaseolus vulgaris* L.) seeds. *J Agric Food Chem* 48:3576–3580