Theor Appl Genet (2013) 126:483–495
DOI 10.1007/s00122-012-1995-3

ORIGINAL PAPER

Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents

Mehrzad Eskandari · Elroy R. Cober · Istvan Rajcan

Received: 29 March 2012 / Accepted: 6 October 2012 / Published online: 29 November 2012

Abstract Soybean seed is a major source of oil for human consumption worldwide and the main renewable feedstock for biodiesel production in North America. Increasing seed oil concentration in soybean [*Glycine max* (L.) Merrill] with no or minimal impact on protein concentration could be accelerated by exploiting quantitative trait loci (QTL) or gene-specific markers. Oil concentration in soybean is a polygenic trait regulated by many genes with mostly small effects and which is negatively associated with protein concentration. The objectives of this study were to discover and validate oil QTL in two recombinant inbred line (RIL) populations derived from crosses between three moderately high-oil soybean cultivars, OAC Wallace, OAC Glencoe, and RCAT Angora. The RIL populations were grown across several environments over 2 years in Ontario, Canada. In a population of 203 $F_{3:6}$ RILs from a cross of OAC Wallace and OAC Glencoe, a total of 11 genomic regions on nine different chromosomes were identified as associated with oil concentration using multiple QTL mapping and single-factor ANOVA. The percentage of the phenotypic variation accounted for by each QTL ranged from 4 to 11 %. Of the five QTL that were tested in a population of 211 $F_{3:5}$ RILs from the cross RCAT Angora × OAC Wallace, a "trait-based" bidirectional selective genotyping analysis validated four QTL (80 %). In addition, a total of seven two-way epistatic interactions were identified for oil concentration in this study. The QTL and epistatic interactions identified in this study could be used in marker-assisted introgression aimed at pyramiding high-oil alleles in soybean cultivars to increase oil concentration for biodiesel as well as edible oil applications.

Communicated by B. Diers.

M. Eskandari · I. Rajcan (✉)
Department of Plant Agriculture, Crop Science Building, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada
e-mail: irajcan@uoguelph.ca

E. R. Cober
Agriculture and Agrifood Canada, Eastern Cereal and Oilseed Crop Research Centre, 960 Carling Ave, Ottawa, ON K1A 0C6, Canada

Introduction

Soybean [*Glycine max* (L.) Merrill] is the largest oil seed crop in the world, which accounted for 56 % of global edible oil production in 2011 (Soystats 2011). Soybean seed oil, which accounts for 18–20 % of the seeds weight, is used for human consumption as well as renewable raw materials for a wide variety of industrial products, including biodiesel (Lee et al. 2007; Clemente and Cahoon 2009). Oil concentration in soybean seeds is a complex quantitative trait governed by a number of genes mostly with small effects and under influence from the environment (Burton 1987; Lee et al. 2007). A well-documented negative relationship between seed oil and protein makes it difficult for breeders to develop high-oil soybean genotypes while retaining a high level of protein concentration (Wilcox and Shibles 2001; Hyten et al. 2004).

Molecular markers have been used in the past two decades to discover quantitative trait loci (QTL), or chromosomal regions associated with seed oil concentration in soybean to be exploited in marker-assisted selection (MAS) programs, which could facilitate the development of high-oil genotypes (Diers et al. 1992; Lee et al. 1996; Orf et al. 1999; Csanádi et al. 2001; Specht et al. 2001; Hyten et al.

2004; Panthee et al. 2005; Qi et al. 2011). Since the first documented attempt to detect oil QTL in soybean (Diers et al. 1992), more than 130 QTL have been reported across the 20 linkage groups (Qi et al. 2011; Soybase 2011). Among these oil QTL, however, only a few have been detected in multiple genetic backgrounds or environments, and none have been widely used in marker-assisted selection (MAS) for high oil in soybean breeding programs. This could be due to several factors affecting the usefulness of QTL, including large confidence intervals, QTL × environment and QTL × genetic background interactions, which all impede the use of QTL in breeding programs (Bernardo 2010; Qi et al. 2011).

Current soybean breeding programs use segregating populations derived from elite parents to improve polygenic quantitative traits such as yield (Palomeque et al. 2009). In contrast, most previous studies aimed at detecting oil QTL used mapping populations that were derived from parental lines with large differences for oil concentration or from plant introductions and exotic germplasm (Hyten et al. 2004). Oil QTL detected in populations that are derived from elite parental lines with high or moderately high oil could increase the chance of detecting more practically suitable oil QTL, which could be used in either marker-assisted selection or introgression programs to develop new varieties with higher levels of oil in the seed.

In the present study, two recombinant inbred line (RIL) populations derived from crosses involving three moderately high-oil soybean cultivars with high seed yield and protein concentration were used to address the following objectives: (1) to identify QTL in the OAC Wallace × OAC Glencoe RIL population (main population) across multiple environments, (2) to validate detected QTL in the main population in a different environment and genetic background using a population derived from the cross RCAT Angora × OAC Wallace.

Materials and methods

Plant materials

A population of 203 $F_{3:6}$ RILs derived from a cross between two moderately high-oil concentration soybean genotypes, OAC Wallace and OAC Glencoe, made at the University of Guelph was used as the main population for mapping QTL in this study. Both parental lines were developed at the University of Guelph; OAC Wallace is a 2,750 crop heat unit (CHU) cultivar (OOPSCC 2012) from a cross between OAC Bayfield and OAC 95-06, and OAC Glencoe is a 3075 CHU cultivar (OOPSCC 2012) from a cross between Talon and OAC Dorado. A RILs population comprised of 211 $F_{3:5}$ lines developed from a cross between RCAT Angora and OAC Wallace, also created at the University of Guelph, was used as a validation population to confirm detected oil QTL in the main population in this study. RCAT Angora is classified as a 3,150-CHU cultivar (OOPSCC 2012) and was developed at the University of Guelph, Ridgetown Campus, from a cross of B152 and T8112.

Experimental design

For the main mapping population, the study was conducted over 2 years (2009 and 2010) at three locations: Woodstock, Ridgetown, and Ottawa [Eastern Cereal and Oilseed Research Centre (ECORC), Agriculture and Agri-Food Canada (AAFC)]. The St. Pauls location in Ontario, Canada, was used for the validation population in 2010. The RILs populations were planted using randomized complete block designs (RCBD) with two replications adjusted for spatial variation with nearest neighbor analysis (NNA) in each of seven environments. At the Ottawa location, each plot consisted of four rows 5 m long, with a 40 cm between row spacing. Plots were trimmed to 4 m in length before harvest and all rows were harvested at this location. At the Ridgetown location, each plot consisted of five rows 4 m long, with 43 cm between row spacing. Plots were trimmed to 3.8 m in length after emergence and only the inside three rows were harvested. At the Woodstock location, each plot consisted of four rows 6.2 m long, with 35.5 cm between row spacing. Plots were trimmed to 5 m in length after emergence and all rows were harvested. For the above three locations, 500 soybean seeds were planted in each plot, which were resulted in plant densities of 50, 54, and 59 seeds/m^2 at Ottawa, Ridgetown, and Woodstock, respectively. At the St. Pauls location, each plot consisted of two rows 6.2 m long, with 35.5 cm between row spacing. Plots were trimmed to 5 m in length after emergence and both rows were harvested. At St. Pauls, 250 soybean seeds were planted in each plot, which was resulted in a plant density of 59 seeds/m^2.

Phenotypic and genotypic data collection

Seed oil concentration was measured on a 5-g seed sample using a Minispec nuclear magnetic resonance (NMR) analyzer (Minispec Mq10, Bruker Inc.) for all trials, except for the test at Ottawa in 2010 in which oil concentration was measured using a near infrared transmission (NIT) machine. The measurements were calculated on a moisture-free basis.

Leaf tissue samples for RILs of both populations were collected from the Woodstock location in 2009 and 2010, respectively. Fifteen leaf tissue disks for each RIL were sampled from seven newly emerged leaflets taken from seven different plants per plot using a single-hole punch

designed to fit a 2 mL screw cap tube. Tissue samples were stored at −80 °C after freeze-drying for 72 h using a Savant ModulyoD Thermoquest (Savant Instruments, Holbrook, NY, USA). Genomic DNA was extracted either from 10 seeds for the parental lines or from 15 leaf disks for each RIL using the Sigma GenElute™ DNA Extraction Kit (SIGMA®, Saint Louis, MO, USA) according to manufacturer's directions. For the polymerase chain reactions (PCRs), genomic DNA was diluted 1:100 and stored at 4 °C. The PCR amplifications were performed in 15-μL aliquots of each reaction mix containing 1.5 μL 10× PCR buffer (Invitrogen Life Sciences Burlington, ON, USA), 1.5 μL 50 mM $MgCl_2$, 1 μL 3 mM deoxyribonucleotide triphosphates (dNTPs) (Invitrogen Life Sciences, Burlington, ON, USA), 2 μL 2.25 mM forward and reverse marker primers (Lab Services, University of Guelph, Guelph, Canada), 0.4 μL 2.5 U/μL *Taq* DNA polymerase (Invitrogen Life Sciences, Burlington, ON, USA), 3 μL genomic DNA, and 3.6 μL sterile water. The PCR reactions for the SSR markers were performed in a 96-well Stratagene Robocycler® (Stratagene Inc., La Jolle, CA, USA) machine with the following thermal sequence: 2 min at 95 °C, followed by 35 cycles of 45 s denaturation at 92 °C, 45 s annealing at 47 °C, 45 s extension at 68 °C, and a final extension for 5 min at 72 °C to complete the reactions. The PCR reactions for gene-based markers were the same as the SSR markers, except for the annealing temperatures that were different for each pair of primers (Table 1). To separate the PCR products, 4.5 % (w/v) agarose gels electrophoreses were used. To begin, 3 μL of loading buffer (6×) was added to the PCR products, and 12 μL samples were loaded on the gels using a BRL Sunrise™ 96 Horizontal Gel Electrophoresis System (Life Technologies, Gaithersburg, MD, USA). A 0.5× TBE solution served as a running buffer and the gels were run at a constant 120 V for 3 h using an EC105 Electrophoresis Power Supply (ThermoEC, Holrook, NY, USA). To score DNA bands, they were stained with ethidium bromide and visualized under UV light.

SSR and gene-specific markers

SSR markers in the soybean molecular lab at the University of Guelph, which currently include 555 primer pairs selected from the integrated soybean genetic map (Song et al. 2004), were initially used to search for polymorphisms between OAC Wallace and OAC Glencoe which were the parental lines for the main population. Diacylglycerol (DGAT) genes have been selected for this study due to their involvement in the Kennedy pathway leading to triacylglycerol (TAG), i.e., oil synthesis. Three pairs of gene-based primers (Table 1) were also used in QTL analyses; (1) GmDGAT1B marker which was designed for the isoform of DGAT1 gene on chromosome 17 (Glyma17g06120), (2) GmDGAT2B marker, designed for the isoform of DGAT2 gene on chromosome 16 (Glyma16a21960), and (3) GmDGAT2C marker, designed for another isoform of DGAT2 gene on chromosome 16 (Glyma16a21970). The names of gene-specific markers corresponded to the gene names (http://www.uky.edu/Ag/Agronomy/PLBC/Research/enzymes/DGAT.htm). Selected polymorphic SSR markers along with these gene-specific markers were used in genotyping the entire main population. For the marker validation population, RCAT Angora × OAC Wallace, all markers which were significantly associated with oil concentration in the main population were screened against RCAT Angora and OAC Wallace to identify polymorphic markers between these genotypes. These selected polymorphic markers were tested against 47 RILs with the highest oil concentration selected from the upper tail of the population and also tested against 47 RILs with the lowest oil concentration selected from the lower tail of the population.

Linkage mapping

A linkage map was obtained using the QTL IciMapping software (Li et al. 2007). Markers were assigned to linkage groups based on a minimum likelihood of odds (LOD) ≥3 and recombination frequencies ≤0.45 centiMorgan (cM). Map distances were estimated using Kosambi's mapping function. Segregation distortion was calculated with the same program to determine the departure from the expected 1:1 allelic frequency ratio. Because of varying levels of heterozygous loci among the RILs and the low frequency of the heterozygous lines for a giver marker, the heterozygous lines were eliminated from the analysis, therefore,

Table 1 Primer information and PCR product sizes for three gene-specific markers used in this study

Marker name	Gene accession	Primer sequence		Amplicon length (bp)	T_a °C[a]
		Forward (5′ → 3′)	Reverse (5′ → 3′)		
GmDGAT2B	Glyma16g21960	GGAGCCAAAAGTTCTAATC	GAAAATCCCTCAAAAGTAAA	1,098	51
GmDGAT2C	Glyma16g21970	CAATGACAAGAAAAGAACTAT	AGGAAGAAAAACAAAGTAAA	955	56
GmDGAT1B	Glyma17g06120	AATCTGAGTGGAATCTTTTACAT	GCAGTTCTTGTTTGTGTTAGTC	495	55

[a] The annealing temperature

all reported data is based on two contrasting homozygous genotypes only. Markers extremely distorted toward one of the parents were also excluded from the study.

QTL and statistical analyses

Simple interval mapping (IM) and composite interval mapping (CIM) were performed using MapQTL®6 software (Van Ooijen 2009). To do CIM, the multiple QTL mapping (MQM) algorithm of MapQTL®6 software (Van Ooijen 2009) was used. In MQM analyses, the significant markers resulting from simple IM analyses were used as cofactors. The empirical LOD threshold values were calculated by performing a permutation test with a set of 2,000 iterations at a Type-I error rate of 0.05.

Variance components analyses were performed for combined environments using the REML algorithm in PROC VARCOMP procedure. The estimated variance components were used to calculate broad-sense heritability (H^2) across environments using the following formula:

$$H^2 = [(V_g)/(V_g + V_{gy}/y + V_{gl}/g + V_{gyl}/yl + V_e/ryl)] \quad (1)$$

where V_g, V_{gy}, V_{gl}, V_{gyl}, and V_e refer to genotypic variance, genotype × year variance, genotype × location variance, genotype × year × location variance, and the residual variance, respectively (Falconer and Mackay 1996). Coefficients y, l, and r refer to the number of years, locations, and replications per location per year, respectively. The variance components attributable to variation among genotypes (V_g) and residual variation (V_E) used to estimate the broad-sense heritability at each environment using the following formula:

$$H^2 = V_g + /(V_g + V_E). \quad (2)$$

SAS version 9.2 (SAS Institute Inc., Cary, NC, USA) was used for estimating LSMEANS for seed oil concentration, performing single-marker ANOVA and stepwise regressions, estimating two-way epistatic effects, and building the best-fit model for individual and combined environments. The Type-I error rate of 0.05 was set for all analyses unless specified otherwise. NNA were conducted to obtain adjusted LSMEANS for the oil concentration trait (Bowley 1999). Single-marker effects were calculated separately for individual and combined environments using PROC GLM, with the LSMEANS estimates of seed oil concentration as the dependent variable and the marker genotypic scores as the independent variable. For each marker, an R^2 statistic was calculated as an estimate of the proportion of the phenotypic variance accounted for by the marker. To determine a suitable subset of significant single markers associated with the dependent variable, backward stepwise regressions using the PROG REG procedure for individual and the combined environments were conducted. Backward elimination process was used in stepwise regression analyses since the probability of Type II error or missing markers associated with the dependent variable is less than the forward solution (Field 2005).

Two-way epistatic effects between each pair of markers and the magnitude of variation accounted for by the interactions (R^2) were calculated by the EPISTACY 2.0 macro, which was run in SAS v. 9.2 (Holland 1998). To reduce the experimental-wise error in epistatic interaction analyses, the Type-I error rate ($\alpha = 0.05$) was divided by $g(g-1)/2$, where g is the number of chromosomes in soybean (Holland, 1998), and it was set at $\alpha = 0.0003$ for all pair-wise comparisons. The type-I error rate was set at 0.01 for single-factor ANOVA analyses. Best-fit models were built for individual and combined environments using significant single markers from single-factor ANOVA along with paired markers with significant epistatic interaction effects from EPISTACY analyses to identify the total amount of variation accounted for by the models. After establishing a model which included significant individual markers followed by a backward stepwise regression for each data set, a PROC GLM procedure was used to add interaction terms to the model, one at a time, and they were kept in the model if they remained significant. The best-fit models for individual and combined environments were identified at the highest R^2 value regardless of the number of markers included.

Selective genotyping

A 'trait-based' selective genotyping method (Navabi et al. 2009) was used to confirm those markers associated with oil concentration, detected in the main population, which were also polymorphic between the parental genotypes of the validation population. This analysis was based on a normal approximation of a binomial distribution of allele frequencies, which was applied to the data obtained from bidirectional selective genotyping. Markers were confirmed as associated with the trait if $|D_q| \geq Z_{(\alpha/2)}S_q$, where $|D_q|$ is the absolute value of the difference between allele frequencies of a given marker. $Z_{(\alpha/2)}$ is the ordinate of the area under the curve from $-\infty$ to $z_{(\alpha/2)}$ obtained from the standardized normal distribution which equals $1 - \alpha/2$, and S_q is the standard error of the difference between marker allele frequencies. In this study, $|D_q|$ was estimated as the difference in allele frequencies between two upper and lower tails, and S_q was calculated using the following formula:

$$S_q = \sqrt{\frac{p_u q_u}{n_u} + \frac{p_l q_l}{n_l}} \quad (3)$$

Table 2 Mean, standard deviation, range, heritability, and parental means for soybean seed oil concentration (g/kg) in two RIL populations of OAC Wallace × OAC Glencoe and RCAT Angora × OAC Wallace in different environments (measurements were taken on a zero-moisture basis)

Environment		Mean (g/kg)	Standard deviation (g/kg)	Heritability[b]	Range (g/kg)	OAC Wallace	OAC Glencoe	RCAT Angora
Location	Years							
Ottawa	2009	222.3	6.0	0.87 (0.02)	200.1–234.6	215.70	209.74	NA
Ridgetown	2009	226.0	5.5	0.79 (0.03)	212.3–240.2	231.45	230.79	NA
Woodstock	2009	212.1	5.5	0.81 (0.03)	197.2–224.9	219.33	209.68	NA
Ottawa	2010	214.7	4.4	0.81 (0.02)	203.4–227.8	215.49	218.14	NA
Ridgetown	2010	227.2	5.5	0.81 (0.03)	214.3–240.1	234.32	226.18	NA
Woodstock	2010	228.9	5.2	0.86 (0.02)	210.4–242.7	228.15	229.18	NA
St. Pauls[a]	2010	205.1	13.1	0.76 (0.03)	174.4–233.2	197.20	NA	199.1
Combined environments		219.1	4.4	0.81 (0.03)	205.9–234.1	224.2	221.1	NA

NA not available

[a] The validation population was derived from RCAR Angora × OAC Wallace

[b] Standard errors are in parentheses

where p_u and q_u are alternate allele frequencies for RILs selected from the upper tail, and p_l and q_l are alternate allele frequencies for RILs selected from the lower tail, and n_u and n_l are, respectively, the number of RILs in the upper and lower trails of population.

Results

A genetic linkage map consisting of 80 markers on 18 linkage groups with an approximate total length of 505 cM and an average distance of 6.2 cM among markers, as well as 13 unlinked markers was generated in this study. Apart from inverted orientations of two pairs of closely linked markers on two linkage groups, the mapping orders and relative linkage distances of the SSR markers within the linkage groups did not show any major discrepancies with the integrated public soybean genetic linkage map by Song et al. (2004). Adjacent SSR markers Sat_020 and Satt196 on Chromosome 9 (LG K), and Sat_244 and Satt643 on Chromosome 7 (LG M) although assigned to their corresponding linkage groups, were inverted compared to the reference linkage map of Song et al. (2004). It should be noted that the orientation of SSR markers Sat_244 and Satt643 on Chromosome 7 (LG M) showed the same marker order when compared with the "high-density" soybean genetic linkage map by Hwang et al. (2009). The low density of polymorphic markers for some of the linkage groups in this study resulted in gaps of up to 50 cM between some adjacent markers; hence they were split into separate linkage groups.

The average seed oil concentration for OAC Wallace (224.2 g/kg) and OAC Glencoe (221.1 g/kg) over six different environments did not show any statistical difference from one another. However, OAC Wallace had significantly greater oil concentration than OAC Glencoe at Woodstock and Ridgetown in 2009 and 2010, respectively, while no statistical differences were detected for the rest of the environments (Table 2). Significant transgressive segregation ($P \leq 0.05$) was present in both RILs populations in all environments. The plot basis broad-sense heritability estimates for oil concentration in the main population ranged from 0.79 to 0.87 in different environments and was 0.81 across the environments (Table 2). In the validation population, the heritability estimate for oil concentration was 0.76 (Table 2).

A total of 11 putative QTL located on eight different chromosomes: Chromosome 9 (LG K), 12 (LG H), 13 (LG F), 14 (LG B2), 16 (LG J), 7 (LG M), 1 (LG D1a), and 17 (LG D2), were identified as associated with oil concentration at either individual environments or in the combined environments based on single-factor ANOVAs followed by stepwise regression analyses (Table 3). The stepwise regression was used to ensure that the QTL were independent of each other and not explaining the same variation. The individual phenotypic variances accounted for by each QTL ranged from 4 to 11 %. Seven of the high-oil QTL alleles were contributed by OAC Glencoe and four QTL alleles by OAC Wallace.

The results of MQM analyses revealed seven genomic regions on six different chromosomes associated with the seed oil concentration in either individual environments or in the combined environments (Table 4; Fig. 1). Using the MQM analyses, we detected six putative QTL identified also by single-marker ANOVAs, and also detected an additional QTL on Chromosome 19 (LG L). The QTL flanked by markers Satt182 and Satt523, on chromosome 19, had been detected at Ottawa in 2010 and accounted for 6.8 % of the total phenotypic variation for oil. This QTL received its oil favorable allele from OAC Wallace.

Table 3 Putative QTL associated with soybean seed oil concentration identified by single-factor ANOVA in a RIL population of OAC Wallace × OAC Glencoe at different Ontario locations in 2009 and 2010

Loci	Chr.	Pos[a]	2009						2010		
			Ottawa			Woodstock			Ottawa		
			R^2 [b]	*P* value	Add. effect[c]	R^2	*P* value	Add. effect	R^2	*P* value	Add. effect
Sat_284	17	30.8	–	–	–	–	–	–	–	–	–
Sat_120	13	76	0.05	0.0073	−1.8	0.08	0.0004	−1.5	0.09	0.0002	−1.7
Satt335[e]	13	77.7	–	–	–	–	–	–	0.05	0.0049	−1.2
Satt490	13	98	–	–	–	–	–	–	–	–	–
Satt317	12	89.5	0.10	0.0000	3.0	–	–	–	–	–	–
Satt463	7	50.1	–	–	–	–	–	–	–	–	–
GmDGAT2B	16	–	–	–	–	0.05	0.0082	−1.1	0.04	0.0137	−5.0
Satt129[e]	1	110	–	–	–	–	–	–	–	–	–
Sat_020	9	104	0.06	0.0037	2.9	–	–	–	–	–	–
Satt001[e]	9	50.6	0.05	0.0070	1.6	–	–	–	–	–	–
Satt066[e]	14	78.8	0.04	0.0100	−1.4	–	–	–	–	–	–
Satt273	9	56.6	–	–	–	–	–	–	–	–	–
Model R^2 [f]				0.18			0.10			0.11	

Loci	Chr.	Pos[a]	2010						Combined environments[d]		
			Ridgetown			Woodstock					
			R^2	*P* value	Add. effect	R^2	*P* value	Add. effect	R^2	*P* value	Add. effect
Sat_284	17	30.8	0.08	0.0002	−1.5	–	–	–	0.05	0.0006	−1.9
Sat_120	13	76	0.06	0.0021	−1.7	–	–	–	0.08	0.0005	−1.7
Satt335[e]	13	77.7	–	–	–	–	–	–	–	–	–
Satt490	13	98	–	–	–	0.05	0.0047	1.0	–	–	–
Satt317	12	89.5	0.11	0.0000	3.0	0.05	0.0047	2.3	0.10	0.0002	3.0
Satt463	7	50.1	–	–	–	–	–	–	0.05	0.0008	−2.3
GmDGAT2B	16	–	0.04	0.0133	−3.0	–	–	–	–	–	–
Satt129[e]	1	110	–	–	–	–	–	–	0.05	0.001	1.8
Sat_020	9	104	–	–	–	–	–	–	–	–	–
Satt001[e]	9	50.6	–	–	–	–	–	–	–	–	–
Satt066[e]	14	78.8	–	–	–	–	–	–	–	–	–
Satt273	9	56.6	0.06	0.0023	−1.4	–	–	–	–	–	–
Model R^2 [f]				0.23			0.08			0.27	

[a] As per Song et al. (2004)

[b] The proportion of the total variance accounted for by the loci

[c] Additive effect (g/kg) at each locus was estimated as half the difference of the phenotypic LSMEAN values of each homozygous genotype. The estimates of additive effect are based on the OAC Wallace allele. A negative value for the estimate indicates that the higher mean was obtained for the alternate, OAC Glencoe allele

[d] Data from Ottawa in 2010 was not included in the combined environments due to using a different machine (NIR) to measure oil concentration at that location

[e] The markers also polymorphic between RCAT Angora and OAC Wallace

[f] The proportion of the total variance accounted for by present loci in the model

The QTL flanked by markers Satt317 and Satt302 on Chromosome 12 (LG H) was detected at four different environments and also in the combined environments and its R^2s ranged from 4.5 to 9.3 %. The QTL linked to the gene-specific marker GmDGAT2B on Chromosome 16 (LG J) identified using single-factor ANOVA was also detected by MQM analysis at Woodstock in 2009 and explained 5.8 % of total phenotypic variation for seed oil concentration.

A total of seven two-way epistatic interactions were identified as significantly ($P \leq 0.0003$) associated with oil

Table 4 Genomic regions associated with soybean seed oil concentration identified by multiple QTL mapping (MQM) in a RIL population of OAC Wallace × OAC Glencoe at different Ontario locations in 2009 and 2010

Interval					2009						2010		
					Ottawa			Woodstock			Ottawa		
$Locus_i$	Pos	$Locus_j$	Pos	Chr.	R^2 [a]	LOD[b]	Add. effect[c]	R^2	LOD	Add. effect	R^2	LOD	Add. effect
Satt317	4.34	Satt302	0.0	12	0.07	3.4	2.8	–	–	–	0.05	2.8	1.7
Satt182[f]	0.00	Satt523	11.5	19	–	–	–	–	–	–	0.07	3.6	1.9
Satt510	63.5	Satt335[f]	73.5	13	–	–	–	–	–	–	0.06	3.1	−1.3
Satt490	100.0	Satt335[f]	73.5	13	–	–	–	–	–	–	–	–	–
GmDGAT2B	20.8	Satt132	0.0	16	–	–	–	0.06	2.6	−1.5	–	–	–
Satt001[e]	34.7	Satt273	42.3	9	0.05	2.7	2.1	–	–	–	–	–	–
Satt323	24.8	Satt463	19.7	7	–	–	–	–	–	–	–	–	–
LOD Threshold						2.6			2.5			2.6	

Interval					2010						Combined environments[d]		
					Ridgetown			Woodstock					
$Locus_i$	Pos	$Locus_j$	Pos	Chr.	R^2	LOD	Add. effect	R^2	LOD	Add. effect	R^2	LOD	Add. effect
Satt317	4.34	Satt302	0.0	12	0.1	4.8	2.8	0.05	2.7	1.8	0.09	5.7	2.7
Satt182[f]	0.00	Satt523	11.5	19	–	–	–	–	–	–	–	–	–
Satt510	63.5	Satt335[f]	73.5	13	–	–	–	–	–	–	–	–	–
Satt490	100.0	Satt335[f]	73.5	13	–	–	–	–	–	–	0.05	3.0	1.9
GmDGAT2B	20.8	Satt132	0.0	16	–	–	–	–	–	–	–	–	–
Satt001[e]	34.7	Satt273	42.3	9	–	–	–	–	–	–	–	–	–
Satt323	24.8	Satt463	19.7	7	–	–	–	–	–		0.06	3.9	−1.5
LOD Threshold						2.7			2.5			2.6	

[a] The proportion of the total phenotypic variance accounted for by QTL

[b] The value on the QTL tagged by the first loci

[c] The estimates of additive effect (g/kg) are based on the OAC Wallace allele. A negative value for the estimate indicates that the higher mean was obtained for the alternate, OAC Glencoe allele

[d] Data from Ottawa in 2010 did not take into account the combined analyses due to using a different machine to measure oil concentration

[e] The markers also polymorphic between RCAT Angora and OAC Wallace

[f] The proportion of the total variance accounted for by present loci in the model

concentration across six different environments (Table 5). The proportion of the total phenotypic variance accounted for by each two-way interaction ranged from 9 to 12 %. Five of these seven markers contributing to the significant two-way interactions affecting the oil concentration were also individually associated with oil concentration in the present study. The Satt367 marker, which showed an association with oil concentration by interacting with three other markers, was previously reported as affecting seed oil concentration in soybean (Reinprecht et al. 2006).

The results of best-fit models for separate and combined environments, which were derived from stepwise regressions combined with the significant two-way epistatic interactions, are presented in Table 6. The proportion of phenotypic variation accounted for by each model ranged from 12 to 45 %. The epistatic interactions that were not significant at $P < 0.01$ in their corresponding best-fit model were discarded from the model. Among the eight markers individually associated with oil concentration detected by best-fit models, five markers also showed significant contributions to the trait through interaction with other markers.

Four putative oil QTL on Chromosome 9 (LG K), 13 (LG F), 14 (LG B2), and 19 (LG L) from the population of OAC Wallace × OAC Glencoe have been confirmed in a population of 211 RILs derived from RCAT Angora × OAC Wallace using a bidirectional selective genotyping method (Table 7). In total, there were five SSR markers associated with oil concentration in the OAC Wallace × OAC Glencoe population which were polymorphic between RCAT Angora and OAC Wallace (Tables 3, 4). While QTL linked to markers Satt001 and Satt182 received their beneficial alleles from OAC Wallace, QTL tagged by markers Satt066 and Satt335 received their favorable oil alleles from RCAT Angora.

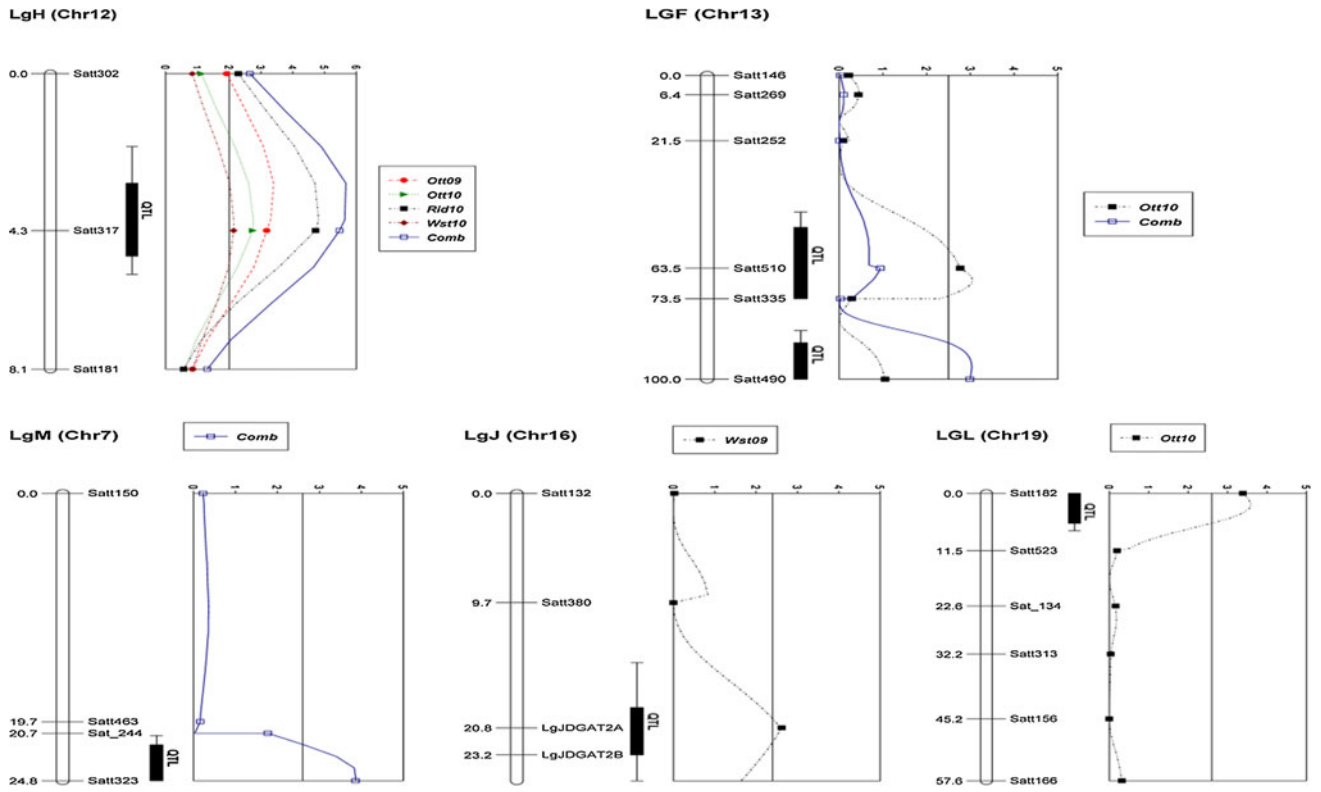

Fig. 1 LOD scores and map distances for oil QTL on chromosomes 7, 12, 13, 16, and 19 in a RIL population derived from the cross of OAC Wallace and OAC Glencoe. The data is based on trials grown in the following locations and years: *Ott09* Ottawa 2009, *Ott10* Ottawa 2010, *Rid10* Ridgetown 2010, *WST09* Woodstock 2009, *WST10* Woodstock 2010, *Comb* combined across environments, except for Ottawa 2010

Table 5 Markers with significant epistatic effects on soybean seed oil concentration and the amount of phenotypic variation accounted for by each interaction in a RIL population of OAC Wallace × OAC Glencoe at three locations in 2009 and 2010

Environment	$Loci_i$	Chr.	$Loci_j$	Chr.	R^2_{ij} [a]	*P* value
Ottawa 2009	Sat_284[c]	17	Satt367	20	0.10	0.0000
	Satt129[c]	1	Satt367	20	0.12	0.0000
Ridgetown 2009	Sat_109	10	Sat_120[c]	13	0.12	0.0003
Woodstock 2009	Satt001[c]	9	Satt560	14	0.12	0.0002
Ottawa 2010	Satt199	18	Satt273[c]	9	0.09	0.0001
Ridgetown 2010	Sat_062	6	Satt411	15	0.12	0.0000
Woodstock 2010	Satt367	20	Satt474	14	0.10	0.0000
Combined[b]	Satt129[c]	1	Satt367	20	0.12	0.0000
	Satt367	20	Satt474	14	0.12	0.0000

[a] The proportion of the total phenotypic variance accounted for by the loci

[b] Data from Ottawa in 2010 did not take into account the combined environments due to using a different machine to measure oil concentrations

[c] Markers solely associated with seed oil concentration detected by either single-factor ANOVA or MQM analyses at any environment

Discussion

It is an old and well-accepted adage in plant breeding, especially for quantitative traits, that crossing "good by good" has a high probability of producing superior new genotypes (Bernardo 2010). This is true also for modern soybean breeding programs, which mostly utilize elite germplasm to improve complex traits, including seed oil concentration. Previous attempts to discover QTL associated with seed oil concentration in soybean mostly focused on populations developed from low and high-oil concentration parents or studying at QTL in exotic germplasm for introgression purposes (Hyten et al. 2004). To our knowledge, our study is the first one aimed at discovering and validating oil QTL using segregating populations derived from hybridizing moderately high-oil soybean modern cultivars with different genetic backgrounds. In the current study, for both populations, while the parental cultivars were not significantly different for seed oil concentration across different environments, the RILs populations exhibited transgressive segregation.

A comparison of the identified oil QTL in the current study with previously reported QTL in many different studies revealed some QTL, which mapped to similar

Table 6 Best-fit models consisting of significant markers (ANOVA) and epistatic interactions (Holland, 1998) for soybean seed oil concentration in a RIL population of OAC Wallace × OAC Glencoe at three locations in 2009 and 2010

Locus	Chr.	2009									2010		
		Ottawa			Ridgetown			Woodstock			Ottawa		
		R^2 [a]	*P* value	Add. effect[b]	R^2	*P* value	Add. effect	R^2	*P* value	Add. effect	R^2	*P* value	Add. effect
Sat_120	13	0.05	0.0073	−1.8	–	–	–	0.08	0.0004	−1.5	0.09	0.0002	−1.7
Satt001	9	0.05	0.0070	1.6	–	–	–	–	–	–	–	–	–
Satt317	12	0.10	<0.0001	1.3	–	–	–	–	–	–	–	–	–
Sat_284xSatt367	15 × 20	0.10	<0.0001	–	–	–	–	–	–	–	–	–	–
Satt129	1	–	–	–	–	–	–	–	–	–	–	–	–
Sat_109 × Sat_120	10 × 13	–	–	–	0.12	0.0003	–	–	–	–	–	–	–
Satt129 × Satt367	1 × 20	–	–	–	–	–	–	–	–	–	–	–	–
GmDGAT2B	16	–	–	–	–	–	–	0.08	0.0082	−1.1	0.04	0.0137	−5.0
Satt001 × Satt560	9 × 14	–	–	–	–	–	–	0.12	0.0002	–	–	–	–
Satt199 × Satt273	18 × 9	–	–	–	–	–	–	–	–	–	0.09	0.0001	–
Satt273	9	–	–	–	–	–	–	–	–	–	–	–	–
Sat_284	17	–	–	–	–	–	–	–	–	–	–	–	–
Sat_062 × Satt411	6 × 15	–	–	–	–	–	–	–	–	–	–	–	–
Satt490	13	–	–	–	–	–	–	–	–	–	–	–	–
Satt367 × Satt474	20 × 14	–	–	–	–	–	–	–	–	–	–	–	–
R^2 [c]		0.45[c]	<0.0001	–	0.12	0.0003	–	0.31	<0.0001	–	0.24	<0.0001	–

Locus	Chr.	2010						Combined environments[d]		
		Ridgetown			Woodstock					
		R^2	*P* value	Add. effect	R^2	*P* value	Add. effect	R^2	*P* value	Add. effect
Sat_120	13	0.06	0.0021	−1.7	–	–	–	0.08	0.0005	−1.7
Satt001	9	–	–	–	–	–	–	–	–	–
Satt317	12	0.11	<0.0001	3.0	0.05	0.0047	2.3	0.10	0.0002	3.0
Sat_284xSatt367	15 × 20	–	–	–	–	–	–	–	–	–
Satt129	1	–	–	–	–	–	–	0.05	0.0010	1.8
Sat_109xSat_120	10 × 13	–	–	–	–	–	–	–	–	–
Satt129xSatt367	1 × 20	–	–	–	–		–	0.12	0.0000	
GmDGAT2B	16	0.04	0.0133	−3.0	–	–	–	–	–	–
Satt001xSatt560	9 × 14	–	–	–	–	–	–	–	–	–
Satt199xSatt273	18 × 9	–	–	–	–	–	–	–	–	–
Satt273	9	0.06	0.0023	−1.4	–	–	–	–	–	–
Sat_284	17	0.08	0.0002	−1.5	–	–	–	0.05	0.0006	−1.9
Sat_062xSatt411	6 × 15	0.12	<.0001	–	–	–	–	–	–	–
Satt490	13	–	–	–	0.05	0.0047	1.0	–	–	–
Satt367xSatt474	20 × 14	–	–	–	0.10	<0.0001	–	–	–	–
R^2 [c]		0.44	<0.0001	–	0.29	<0.0001	–	0.38	<0.0001	–

[a] The proportion of the total phenotypic variance accounted for by the loci

[b] Additive effect (g/kg) at each locus estimated as half the difference of the phenotypic LSMEAN values of each homozygous genotype. The estimates of additive effect are based on the OAC Wallace allele. A negative value for the estimate indicates that the higher mean was obtained for the alternate, OAC Glencoe allele

[c] The proportion of the total Phenotypic variance accounted for by present loci in the best-fit model

[d] Data from Ottawa in 2010 did not take into account the combined environments due to using a different machine to measure oil concentrations

Table 7 Putative QTL associated with soybean seed oil concentration confirmed using a 'trait-based' bidirectional selective genotyping analysis (Navabi et al. 2009) by genotyping 44 % of two high-oil and low-oil subsets of the lines in a RIL population of RCAT Angora × OAC Wallace at St. Pauls in 2010

Locus	Chr.	Allele frequency				D_q^c	S_q^d	*P* value
		High oil		Low oil				
		W[a]	A[b]	W[a]	A[b]			
Satt001	9	0.68	0.32	0.39	0.61	0.29	0.10	<0.01
Satt066	14	0.31	0.69	0.71	0.29	0.40	0.09	<0.01
Satt182	19	0.67	0.33	0.36	0.64	0.31	0.10	<0.01
Satt335	13	0.37	0.63	0.69	0.31	0.32	0.10	<0.01

[a] The frequency of the allele which was shared by OAC Wallace

[b] The frequency of the allele which was shared by RCAT Angora

[c] The absolute value of the difference in marker allele frequencies

[d] The standard error of the difference between marker allele frequencies

genomic regions. For instance, one QTL on Chromosome 12 (LG H), tagged by SSR marker Satt317, had been also reported to be associated with oil concentration in a study by Panthee et al. (2005). These authors used CIM in a population of 101 RILs from a cross between 'N87-984-16' and 'TN93-99' soybean genotypes (Panthee et al. 2005). While the marker that tagged the oil QTL between these two studies is the same, we cannot say with certainty that our study has confirmed the Panthee et al. (2005) QTL.

In another study, Shan et al. (2008) also reported an oil QTL in this genomic region, linked to the marker Satt293, by studying a population of 154 RILs derived from a cross of 'Charleston' × 'Dongnong'. The SSR marker Satt293 is located only about 0.4 cM away from the Satt317 (Song et al. 2004). This genomic region was also identified to be carrying oil QTL based on an integrated map of oil-associated QTL in soybeans (Qi et al. 2011). This integrated map consisting of 20 "consensus" genomic regions associated with oil concentration was constructed by integrating 130 QTL from different studies using a "meta-analysis" method that narrowed down the confidence interval of the QTL to increase the precision and validity. On this linkage group four additional oil QTL have been reported (Lee et al. 1996; Brummer et al. 1997; Qui et al. 1999; Hyten et al. 2004). However, they are at least 34 cM away from the QTL detected in the current study, indicating that they probably represent different QTL.

The putative oil QTL on Chromosome 13 (LG F), which was tagged by markers Satt510 in the interval of Satt510-Satt335 and detected in Ottawa 2010 using MQM, may represent the same QTL as that associated with individual SSR markers Sat_120 and Satt335, which was identified in different specific environments using single-factor ANOVA. The marker Sat_120 is about 4.6 cM and the marker Satt335 is about 6.3 cM away from the marker Satt510 (Song et al. 2004). This putative QTL received its oil beneficial allele from OAC Glencoe. In the present study, also a putative oil QTL was detected in the interval of Satt490–Satt335 in combined environments, which was tagged by Satt490 and inherited its positive oil allele from OAC Wallace. Due to the distance of more than 26 cM between the markers Satt490 and Satt510 and the fact that favorable oil allele at these loci came from different parents, we believe that QTL tagged by these markers are different in this genomic region.

In the current study, an oil QTL marked by Satt335 was detected using bidirectional selective genotyping in a population of RCAT Angora × OAC Wallace. Specht et al. (2001) discovered a putative oil QTL linked to the marker Satt510 in a RIL population derived from a cross of 'Minsoy' × 'Noir 1'. There are two more QTL which have been reported to be associated with oil on this linkage group (Shan et al. 2008; Qi et al. 2011), however, they are located more than 55 cM away from the QTL identified in the present study, and so they are likely different oil QTL on this linkage group.

Hyten et al. (2004) reported a QTL for oil in close proximity to our QTL on Chromosome 19 (LG L) in the interval of Satt182–Satt523. They discovered this putative oil QTL linked to the marker Satt523 by studying a population of 131 F_6-derived RILs from a cross of 'Essex' × 'Williams'. Reinprecht et al. (2006) also detected a QTL associated with the marker Satt182 in a population of 169 RILs derived from a cross between a low linolenic acid line, 'RG10', and a lipoxyygenase null line, 'OX948'. There are also some other reported QTL in close proximity to this genomic region that had been detected in different genetic backgrounds and environments (Diers et al. 1992; Lee et al. 1996; Qi et al. 2011). They are located at 3–9 cM from the marker Satt523. However, due to the low density of markers on this linkage group in the current study, we could not determine if these QTL were the same as ours.

One oil QTL on Chromosome 17 (LG D2) was identified as linked to the SSR marker Sat_282 in Ridgetown in 2010 and also in the combined environments using single-marker ANOVA. The marker Sat_282 is located about 6.3 cM away from the SSR marker Satt458, which was reported to be associated with oil concentration by Hyten et al. (2004). There are also four more oil QTL on this chromosome which are located at more than 40 cM away from this QTL (Lee et al. 1996; Hyten et al. 2004). The oil QTL linked to marker Sat_282 was not associated with seed protein concentration in this study (data not shown).

The putative oil QTL on Chromosome 7 (LG M), located in or close to the interval of Satt323–Satt463 was detected in the combined environments using both single-marker ANOVA and MQM. The oil beneficial allele was

from OAC Glencoe and explained 6 % of the total phenotypic variation. The closest previously reported oil QTL to this QTL was linked to RFLP marker R079_1 (Lark et al. 1995) and SSR marker Satt540 (Hyten et al. 2004), which are 10.9 and 16.6 cM away from Satt463, respectively (Song et al. 2004).

The putative oil QTL associated with marker Satt129, located on Chromosome 1 (LG D1a), was detected in the combined environments using single-marker ANOVA. This QTL was also reported by Hyten et al. (2004) as associated with SSR marker Satt147, which is located less than 1 cM away from Satt129 (Song et al. 2004). There are three more oil putative QTL on this linkage group that are more than 50 cM away from our putative QTL (Hyten et al. 2004).

This study also discovered novel oil QTL. The oil QTL on Chromosome 16 (LG J) tagged by the gene-specific marker GmDGAT2B was identified using both MQM and single-factor ANOVAs in three different environments. The GmDGAT2 marker was designed based on a 2-bps indel mutation on the second exon of the DGAT2 isoform on chromosome 16 (Eskandari and Rajcan, in preparation). This mutation generated an immature stop codon that resulted in a truncated polypeptide with only 141-amino acids, which might be ineffective. The protein encoded by the wild type gene has a sequence of 350-amino acid. Since the wild type allele from OAC Glencoe contributed to an increase in seed oil concentration in the RILs, this gene is suggested to be involved in the oil biosynthesis process.

A substantial role for DGAT2 genes in triacylglycerol (TAG) formation has been previously reported in soybean and other species (Zou et al. 1999; Lardizabal et al. 2008). Moreover, Lardizabal et al. (2008) reported that they could increase seed oil concentration up to 1.5 % in soybean by introducing a fungal DGAT2 with no significant impact on seed yield or protein concentration. In the current study, this QTL did not have any association with protein concentration, which was in agreement with Lardizabal et al. (2008). DGAT2 genes, however, have been reported in some plant species to only affect the quality of the seed storage oils due to their effect on unusual fatty acid accumulation (Shockey et al. 2006; Burgal et al. 2008; Li et al. 2010). There are two previously identified QTL for oil concentration on Chromosome 16 (LG J), which were reported earlier by Lee et al. (1996) and Kabelka et al. (2004). Due to a limited number of markers on this linkage group in this study, we could not determine if the previously discovered QTL were also linked to this gene.

Within the genomic region flanked by SSR markers Satt001 and Satt273 on Chromosome 9 (LG K), a new oil QTL has been detected using MQM and single-marker ANOVAs in two environments. The closest previously reported oil QTL to this genomic region was the QTL associated with the RFLP marker A315-1 (Mansur et al., 1993) which is located about 21.9 and 27.9 cM away from the Satt001 and Satt273, respectively. Brummer et al. (1997) also discovered an oil QTL on this linkage group, which was associated with the RFLP marker K387-1 and located about 42.3 cM far away from the marker Satt273. The putative QTL associated with the marker Satt001 was also identified associated with oil concentration in a RCAT Angora × OAC Wallace RIL population in the current study.

On Chromosome 9 (LG K), we also detected an oil QTL tagged by Sat_020 at the Ottawa location in 2009. The favorable oil allele for this QTL came from OAC Wallace and also showed a positive significant association with seed protein composition at Ridgetown in 2010 and in the combined analysis. The closest previously reported oil QTL on this chromosome is the QTL linked to the RFLP marker K387-1 (Brummer et al. 1997), which is located about 24.1 cM away from SSR marker Sat_020 (Song et al. 2004).

The region on Chromosome 14 (LG B2), linked to SSR marker Satt066 where we mapped an oil QTL, was not previously reported as associated with seed oil concentration, but rather seed yield (Concibido et al. 2003). However, this marker did not show any association with either seed yield or protein concentration in this study (data not shown). This QTL was also confirmed as associated with oil concentration in the RCAT Angora × OAC Wallace validation population. The closest oil QTL to the marker Satt066 was reported by Csanádi et al. (2003), which was tagged by Satt020 and is located 6.7 cM away from our QTL, which means that it may represent the same QTL. Two more oil QTL on this chromosome have been reported (Diers et al. 1992; Brummer et al. 1997), which were at least 39 cM away from the marker Satt066.

Epistatic interactions have not received much attention in QTL analysis of seed oil concentration in soybean. This could be due in part to the difficulty of exploiting them in plant breeding programs especially when a large number of interactions are involved and associated with the trait of interest (Bernardo 2010). It could also be because the amount of total phenotypic variances explained by each interaction tends to be small (Bernardo 2010). However, in our study, using a population of 203 RILs, the proportion of phenotypic additive × additive variance accounted for by each interaction was relatively high and ranged from 9 to 12 % of the total phenotypic variation for seed oil. In this study, 71 % of the epistatic interactions included a marker that was individually associated with oil concentration whereas in the rest neither marker was. The results of the current study show the importance of two-way interaction effects on oil concentration in this population. These oil-associated interaction effects could be fixed through selecting appropriate RILs from this population to establish new RILs populations that will probably have

pure lines with greater breeding values for the oil concentration due to the conversion of a portion of epistatic variance into additive variance (Lark et al. 1995; Bernardo 2010).

The results of best-fit models with and without epistatic interactions (Tables 2, 7) provided further support for the existence of "environmentally stable" oil QTL on different chromosomes. QTL linked to markers Sat_120, Satt317, and GmDGAT2B were present in their corresponding best-fit model in three or more different environments.

In conclusion, we identified an oil QTL on Chromosome 9 (LG K) which also had a significant positive effect on seed protein composition, and three oil QTL on Chromosomes 14 (LG B2), 16 (LG J), and 17 (LG D2) that had no significant effects on seed protein concentration in any of the environments. These QTL could be used in breeding programs targeted at increasing the oil concentration without affecting protein concentration. Our detection of new oil QTL in this study also demonstrated the importance of using moderately high-oil elite soybean cultivars as valuable sources for detecting minor QTL, which could be masked by major ones in populations derived from parents with large differences in the oil concentration (Asins 2002; Winter et al. 2007). These QTL along with the epistatic interactions identified in this study could be used in either marker-assisted selection or allele introgression to increase the frequency of favorable QTL alleles by pyramiding the beneficial high-oil alleles using from novel sources while maintaining the current ones. The partial genetic map for population OAC Wallace × OAC Glencoe developed in the current study provided insight into genomic regions that govern seed oil concentration in soybean; however, the development of a more saturated genetic linkage map would enhance the chance of identification of more oil QTL, especially within the gaps in our current genetic map.

Acknowledgments The authors would like to thank the late Dr. G.R. Ablett, and Drs. K.P. Pauls, L.R. Erickson, Y. Kakuda, and A. Navabi (University of Guelph) for their valuable suggestions on this research. Also, the authors are grateful to Wade Montminy, Chris Grainger, Ron Guillemette, Bryan Stirling, Dennis Fischer and the entire soybean crew at the University of Guelph in Guelph and Ridgetown, Ontario, for their excellent technical assistance and support. Generous funding to conduct this research was provided by the Alternative Renewable Fuels II Program of the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) and by the Grain Farmers of Ontario.

References

Asins MJ (2002) Present and future quantitative trait locus analysis in plant breeding. Plant Breed 121:281–291

Bernardo R (2010) Breeding for quantitative traits in plants, 2nd edn. Stemma Press, Woodbury, p 73, 141

Bowley S (1999) A hitchhiker's guide to statistics in plant biology. Any Old Subject Books, Guelph, p 115

Brummer EC, Graef GL, Orf J, Wilcox JR, Shoemaker RC (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci 37:370–378

Burgal J, Shockey J, Lu CF, Dyer J, Larson T, Graham I, Browse J (2008) Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. Plant Biotechnol J 6:819–831

Burton JW (1987) Quantitative genetics: results relevant to soybean breeding. In: Wilcox JR (ed) Soybeans: improvement, production, and uses, 2nd edn. ASA, CSSA, and SSSA, Madison, pp 211–242

Clemente TE, Cahoon EB (2009) Soybean oil: genetic approaches for modification of functionality and total oil. Plant Physiol 151: 1030–1040

Concibido VC, La Vallee B, Mclaird P, Pineda N, Meyer J, Hummel L, Yang J, Wu K, Delannay X (2003) Introgression of a quantitative trait locus for yield from glycine soja into commercial soybean cultivars. Theor Appl Genet 106:575–582

Csanadi G, Vollmann J, Stift G, Lelley T (2001) Seed quality QTLs identified in a molecular map of early maturing soybean. Theor Appl Genet 103:912–919

Diers BW, Keim P, Fehr WR, Shoemaker RC (1992) RFLP analysis of soybean seed protein and oil content. Theor Appl Genet 83:608–612

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. 4th edn. Longman, New York, pp 122–144

Field A (2005) Discovering statistics using SPSS, 2nd edn. SAGE Publication Ltd, London, p 364

Holland JB (1998) EPISTACY: a SAS program for detecting two locus epistatic interactions using genetic marker information. J Heredity 89:374–375

Hwang T, Sayama T, Takahashi M, Takada Y, Nakamoto Y, Funatsuki Y, Hisano H, Sasamoto S, Sato S, Tabata S, Kono I, Hoshi M, Hanawa M, Yano C, Xia Z, Harada K, Kitamura K, Ishimoto M (2009) High-density integrated linkage map based on SSR markers in soybean. DNA Res 16(4):213–225

Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, Schmidt ME (2004) Seed quality QTL in a prominent soybean population. Theor Appl Genet 109:552–561

Kabelka EA, Diers BW, Fehr WR, LeRoy AR, Baianu IC, You T, Neece DJ, Nelson RL (2004) Putative alleles for increased yield from soybean plant introductions. Crop Sci 44:784–791

Lardizabal K, Effertz R, Levering C, Mai J, Pedroso MC, Jury T, Aasen E, Gruys K, Bennett K (2008) Expression of Umbelopsis ramanniana DGAT2A in seed increases oil in soybean. Plant Physiol 148:89–96

Lark KG, Chase K, Adler F, Mansur LM, Orf JH (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc Natl Acad Sci USA 92:4656–4660

Lee SH, Bailey MA, Mian MAR, Carter TE, Shipe ER, Ashley DA, Parrott WA, Hussey RS, Boerma HR (1996) RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor Appl Genet 93:649–657

Lee GJ, Wu X, Shannon JG, Sleper DA, Nguyen HT (2007) Soybean. In: Kole C (ed) Genome mapping and molecular breeding in plants, volume 2, oilseeds. Springer, Berlin, pp 1–3

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

Li R, Yu K, Hildebrand DF (2010) DGAT1, DGAT2 and PDAT expression in seeds and other tissues of epoxy and hydroxy fatty acid accumulating plants. Lipids 45:145–157

Mansur LM, Orf JH, Lark KG (1993) Determining the linkage of qualitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). Theor Appl Genet 86:914–918

Navabi A, Mather DE, Bernier J, Spaner DM, Atlin GN (2009) QTL detection with bidirectional and unidirectional selective genotyping: marker-based and trait-based analysis. Theor Appl Genet 118:347–358

OOPSCC (2012) Ontario oil and protein seed crop committee. http://www.oopscc.org/index.php, reports 2000, 2001, and 2010

Orf JH, Chase K, Jarvik T, Mansur LM, Cregan PB, Adler FR, Lark KG (1999) Genetics of soybean agronomic traits: I. comparison of three related recombinant inbred populations. Crop Sci 39: 1642–1651

Palomeque L, Liu L, Li W, Hedges B, Cober E, Rajcan I (2009) QTL in mega-environments: I. Universal and speciWc seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines. Theor Appl Genet 119:417–427

Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE (2005) Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Sci 45:2015–2022

Qi Z, Wu Q, Han X, Sun Y, Du X, Liu C, Jiang H, Hu G, Chen Q (2011) Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. Euphytica 179:499–514

Qui BX, Arelli PR, Sleper DA (1999) RFLP markers associated with soybean cyst nematode resistance and seed composition in a 'Peking' × 'Essex' population. Theor Appl Genet 98:356–364

Reinprecht Y, Poysa VW, Yu KF, Rajcan I, Ablett GR, Pauls KP (2006) Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. Genome 49:1510–1527

Shan DP, Qi ZM, Qiu HM, Shan CY, Liu CY, Hu H, Chen QS (2008) Epistatic effects of QTLs and QE interaction effects on oil content in soybean. Acta Agron Sin 34(6):952–957

Shockey JM, Gidda SK, Chapital DC, Kuan JC, Dhanoa PK, Bland JM, Rothstein SJ, Mullen RT, Dyer JM (2006) Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. Plant Cell 18:2294–2313

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

Soybase (2011) Soybean breeder's toolbox genetic map information. http://soybeanbreederstoolbox.org/ (Verified on 20 Feb 2012)

Soystats (2011) Soy Stats: a reference guide to important soybean fact and figures. http://www.soystats.com/2011/Default-frames.htm (Verified on 25 Jun 2011)

Specht JE, Chase K, Macrander M, Graef BL, Chung J, Markwell JP, Germann M, Orf JH, Lark KG (2001) Soybean response to water: a QTL analysis of drought tolerance. Crop Sci 41:493–509

Van Ooijen JW (2009) MapQTL® 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma B. V, Wageningen

Wilcox JR, Shibles RM (2001) Interrelationships among seed quality attributes in soybean. Crop Sci 41:11–14

Winter SMJ, Shelp BJ, Anderson TR, Welacky TW, Rajcan I (2007) QTL associated with horizontal resistance to soybean cyst nematode in Glycine soja PI464925B. Theor Appl Genet 114: 461–472

Zou J, Wei Y, Jako C, Kumar A, Selvaraj G, Taylor DC (1999) The Arabidopsis thaliana TAG1 mutant has a mutation in a diacylglycerol acyltransferase gene. Plant J 19:645–653