Theor Appl Genet (2013) 126:1677–1687
DOI 10.1007/s00122-013-2083-z

ORIGINAL PAPER

Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield

Mehrzad Eskandari · Elroy R. Cober · Istvan Rajcan

Received: 25 August 2012 / Accepted: 28 February 2013 / Published online: 28 March 2013

Abstract Soybean [*Glycine max* (L.) Merrill] seed oil is the primary global source of edible oil and a major renewable and sustainable feedstock for biodiesel production. Therefore, increasing the relative oil concentration in soybean is desirable; however, that goal is complex due to the quantitative nature of the oil concentration trait and possible effects on major agronomic traits such as seed yield or protein concentration. The objectives of the present study were to study the relationship between seed oil concentration and important agronomic and seed quality traits, including seed yield, 100-seed weight, protein concentration, plant height, and days to maturity, and to identify oil quantitative trait loci (QTL) that are co-localized with the traits evaluated. A population of 203 $F_{4:6}$ recombinant inbred lines, derived from a cross between moderately high oil soybean genotypes OAC Wallace and OAC Glencoe, was developed and grown across multiple environments in Ontario, Canada, in 2009 and 2010. Among the 11 QTL associated with seed oil concentration in the population, which were detected using either single-factor ANOVA or multiple QTL mapping methods, the number of QTL that were co-localized with other important traits QTL were six for protein concentration, four for seed yield, two for 100-seed weight, one for days to maturity, and one for plant height. The oil-beneficial allele of the QTL tagged by marker Sat_020 was positively associated with seed protein concentration. The oil favorable alleles of markers Satt001 and GmDGAT2B were positively correlated with seed yield. In addition, significant two-way epistatic interactions, where one of the interacting markers was solely associated with seed oil concentration, were identified for the selected traits in this study. The number of significant epistatic interactions was seven for yield, four for days to maturity, two for 100-seed weight, one for protein concentration, and one for plant height. The identified molecular markers associated with oil-related QTL in this study, which also have positive effects on other important traits such as seed yield and protein concentration, could be used in the soybean marker breeding programs aimed at developing either higher seed yield and oil concentration or higher seed protein and oil concentration per hectare. Alternatively, selecting complementary parents with greater breeding values due to positive epistatic interactions could lead to the development of higher oil soybean cultivars.

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 one the most important sources of edible oil for humans worldwide and a reliable source of renewable and sustainable feedstock for biodiesel production (Clemente and Cahoon 2009; ASA 2011). Increasing the relative oil concentration in soybean seeds is complicated partly due to correlations with other agronomic and seed composition traits such as seed yield, size, and protein concentration (Burton 1987; Lee et al. 2007; Clemente and Cahoon 2009). As a result, improving

seed oil concentration, while maintaining a high level of protein concentration, has not been very successful using conventional breeding methods (Smith and Weber 1968; Burton and Brim 1981; Feng et al. 2004). The most significant progress in increasing total seed oil in soybean has been made through increasing overall seed yield, which translates into more oil production per hectare (Clemente and Cahoon 2009).

Molecular markers have made it possible for plant geneticists and breeders to identify, isolate, and even transfer beneficial genes into elite germplasm without transmitting unfavorable loci (Tanksley and McCouch 1997). Molecular markers have also been used in past two decades to discover quantitative trait loci (QTL) or chromosomal regions associated with seed oil concentration and other important agronomic and seed composition traits in soybean, which potentially could be used in marker-assisted selection (MAS) programs (Keim et al. 1997; Diers et al. 1992; Lark et al. 1995; Lee et al. 1996; Orf et al. 1999a; Qui et al. 1999; Csanádi et al. 2001; Specht et al. 2001; Chung et al. 2003; Kabelka et al. 2004; Hyten et al. 2004; Panthee et al. 2005; Reinprecht et al. 2006; Palomeque et al. 2009b; Qi et al. 2011).

Several seed oil QTL in soybean have been reported that were co-localized with other agronomic and seed composition characteristics, including seed yield, protein concentration, seed size, and maturity (Lark et al. 1995; Orf et al. 1999a; Qui et al. 1999; Csanádi et al. 2001; Specht et al. 2001; Chung et al. 2003; Kabelka et al. 2004; Hyten et al. 2004; Panthee et al. 2005; Reinprecht et al. 2006; Palomeque et al. 2009b). Most previous studies aimed at detecting QTL associated with important agronomic and seed traits used mapping populations that were derived from parental lines with large differences for target traits or from plant introductions and exotic germplasm (Hyten et al. 2004). While populations with large parental differences may be suitable for detecting major QTL for the trait under study, they may not be as useful in discovering minor QTL, which could be masked by major QTL (Asins 2002; Winter et al. 2007).

In this study, a recombinant inbred line (RIL) population derived from a cross between two moderately high oil soybean cultivars that also have high seed yield and protein concentration, OAC Wallace and OAC Glencoe, was used to address the following objectives: (1) to study genetic and phenotypic correlations between seed oil concentration and selected agronomic and seed quality traits, (2) to determine the co-localization of detected oil QTL in this population with QTL for other traits, and (3) to determine the effects of 2-way epistatic interactions of markers on target traits, where at least one of the two involved markers in a given interaction was individually associated with seed oil concentration.

Materials and methods

Description of the experimental designs and conditions as well as DNA extraction and genetic linkage map construction were provided in detail in a preceding companion paper (Eskandari et al. 2013). Briefly, a population of 203 $F_{4:6}$ RILs derived from two cultivars with higher than average seed oil concentration, OAC Wallace and OAC Glencoe, that also had high seed yield and moderately high protein concentration was developed, grown, and evaluated across three locations in Ontario, Canada, in 2009 and 2010. The population was evaluated in the field using randomized complete block designs (RCBD) with two replications and adjusting for spatial variation with the nearest neighbor analysis (NNA) in each of six environments.

Five hundred and fifty-five available SSR markers in the Rajcan's molecular lab at the University of Guelph, which were selected from the integrated soybean genetic map (Song et al. 2004), were used initially to screen for polymorphisms between the parental cultivars, OAC Wallace and OAC Glencoe. Diacylglycerol (*DGAT*) genes have been selected for this study because of their involvement in the sn-glycerol-3-phosphate pathway leading to triacylglycerol (TAG), i.e., oil synthesis. Three pairs of gene-based primers (Eskandari et al. 2013) 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).

Ninety selected polymorphic SSR markers, along with three gene-specific markers (GmDGAT1B, GmDGAT2B, and GmDGAT2C), were used in genotyping the entire population. A linkage map consisting of 80 markers distributed across linkage groups (LGs) was obtained using the QTL IciMapping software (Li et al. 2007). Markers were assigned to LGs based on a minimum likelihood of odds (LOD) $\geq$3 and recombination frequencies $\leq$0.45 cM. Map distances were estimated using the Kosambi's mapping function.

Phenotypic data collection

The phenotypic data of the agronomic and seed quality traits were evaluated and collected from each plot for all trials. Seed yield measurements were converted to kg ha^{-1} and adjusted to 130 g kg^{-1} moisture. "Days to maturity" was determined as the number of days after planting until approximately 95 % of the pods were matured (Fehr et al.

1971). Plant height (cm) was measured at maturity as the average distance from the soil surface to the tip of the main stem. Seed size (g) was determined by weighing 100 randomly selected seeds from each plot and adjusted to 130 g/kg moisture. Seed crude protein concentration (g kg^{-1}) of each line was measured and adjusted to 130 g kg^{-1} moisture using a Zeltex NIR analyzer (ZX-50 SRT, Zeltex, Inc., USA) on about 50 g whole bean sample from each plot. Seed oil concentration was measured on 5 g seed samples using a Minispec nuclear magnetic resonance (NMR) analyzer (Minispec Mq10, Bruker Inc., Germany) for all trials with the exception of the test at Ottawa in 2010, where it was measured using a near infrared transmission (NIR) machine (Infratec 1241 Grain Analyser, Foss Inc., Eden Prairie, MN) since NMR was not available at that site.

Statistical analysis

SAS version 9.2 (SAS Institute Inc., Cary, NC) was used to perform statistical analyses, including estimating LSMEANS (PROC GLM), performing single-factor ANOVA (PROC GLM), stepwise regression (PROC REG), and variance components (PROC VARCOMP). Single-factor ANOVA, with trait estimates as the dependent and markers genotypes as the independent variable, was used to test single marker effects. Stepwise regression, with backward elimination process, was used to ensure that individual significant markers that were retained in the model represented distinct QTL. The type-I error rate (α) was set at 0.05 for all analyses unless specified. Transgressive segregation for a given trait was considered significant if there was at least one RIL with a greater value than the threshold of $p_h + z_{(0.05)}\sigma_e$, where p_h, $z_{(0.05)}$, is the phenotypic value for the parent with the greater value for the trait at the critical Z value ($P = 0.05$) for normal distribution, and σ_e is standard error of the population.

The estimated variance components were used to calculate the broad sense heritability (H^2) of individual and combined environments (Hallauer and Miranda 1988; Nyquist and Baker 1991; Falconer and Mackay 1996). Broad sense heritability (H^2) and standard error (Se(H^2)) estimates were calculated as

$$H^2 = \sigma_G^2 / \left[\sigma_G^2 + \left(\sigma_{GE}^2 / e\right) + \left(\sigma^2 / re\right)\right]$$

and

$$\mathrm{Se}(H^2) = Se(\sigma_G^2) / \left(\sigma_G^2 + \sigma_{GE}^2/e + \sigma_e^2/re\right)$$

in which σ_G^2, σ_{GE}^2, σ_e^2, and Se(σ_G^2) refer to genotypic variance, genotype × environment variance, residual variance, and standard error of genotypic variance, respectively. Coefficients e and r refer to the number of environments and replications within environments, respectively.

Genetic and phenotypic correlations between seed oil concentration and the traits evaluated, and their standard errors, were estimated using multivariate REML implemented in the MIXED procedure (Holland 1998). Using the genotypic and phenotypic variance and co-variance component estimates, the genotypic (r_{Gij}) and phenotypic (r_{Pij}) correlation estimates between traits i and j are computed as

$$r_{Gij} = \sigma_{Gij} / \sqrt{\left(\sigma_{Gi}^2 \sigma_{Gj}^2\right)}$$

and

$$r_{Pij} = \sigma_{Pij} / \sqrt{\left(\sigma_{Pi}^2 \sigma_{Pj}^2\right)} = \left(\sigma_{Gij+}^2 \sigma_{GEij+}^2 \sigma_{eij}^2\right) / \sqrt{\left(\sigma_{Gi+}^2 \sigma_{GEi+}^2 \sigma_{ei}^2\right)} \sqrt{\left(\sigma_{Gj+}^2 \sigma_{GEj+}^2 \sigma_{ej}^2\right)}$$

in which σ_{Gij} and σ_{Pij} refer to the estimated genotypic and phenotypic co-variances between traits i and j, respectively; σ_G^2, σ_P^2, σ_{GE}^2, and σ_e^2 refer to the estimated genotypic variance, phenotypic variance, genotype × environment variance, and residual variance. Correlations were considered significantly different from zero if their approximate $1-\alpha$ confidence intervals did not include zero (Holland et al. 2003). Confidence intervals were estimated as $r \pm z(\alpha/2)\sigma_e$, where r is the correlation coefficient, $z(\alpha/2)$ is the standard normal distribution critical value at $P = \alpha$, and σ_e is the standard error of the correlation coefficients (Iqbal et al. 2007).

Two-way epistatic effects between each pair of markers and the magnitude of variation accounted for by the interactions (R^2) were calculated by EPISTACY 2.0 macro (Holland, 1998). To control 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 (Holland, 1998), and it was set at $\alpha = 0.0003$ for all pair-wise comparisons.

Simple interval mapping (IM) and composite interval mapping (CIM) were performed using MapQTL®6 software (van Ooijen 2009). The multiple QTL mapping (MQM) algorithm of MapQTL®6 software (van Ooijen 2009) was used to perform CIM. In MQM analyses, the significant markers resulting from simple IM analyses were exploited as co-factors. 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.

Results

In the previous study (Eskandari et al. 2013), a total of 11 genomic regions located on nine different chromosomes were identified and reported as associated with seed oil concentration in a RIL population derived from OAC

Wallace × OAC Glencoe across several environments. Among the 11 oil QTL, four genomic regions tagged by SSR markers Satt317, Satt001, Satt335, and Satt463 were also identified as associated with at least two additional agronomic or seed traits by either single-factor ANOVA (Table 1) or MQM methods (Table 2). Remarkably, the seed oil QTL tagged by Satt317 was co-localized with QTL for all the traits under study: seed yield, 100-seed weight, and seed protein concentration as well as days to maturity, and plant height. The oil QTL tagged by markers Satt001 and Satt335 were also co-localized with seed protein concentration and 100-seed weight QTL in several different environments and also in the combined analyses for the traits across environments. The QTL linked to Satt001 was also identified as affecting seed yield at Ottawa location in 2009 and 2010 (Tables 1, 2). The genomic region tagged by the SSR marker Satt463, which we reported earlier as associated with seed oil concentration (Eskandari et al. 2013), was associated with a seed yield QTL at Woodstock in 2009 using single-factor ANOVA (Table 1) and a seed protein QTL at Ottawa in 2010 using MQM (Table 2).

Among the markers associated with seed oil concentration in the previous study (Eskandari et al. 2013), some showed statistically significant ($P \leq 0.0003$) effects on other traits when interacting with other markers evaluated (Table 3). A total of seven 2-way epistatic interactions were identified as significantly associated with seed yield across different environments; four for days to maturity, two for 100-seed weight, one for seed protein concentration, and one for plant height (Table 3). The marker Sat_020 that was detected as associated with an oil QTL at Ottawa in 2009 (Eskandari et al. 2013), was also associated with days to maturity by interacting with the genomic region tagged by the marker Satt155 in two environments (Table 3). The oil QTL tagged by Satt182 that was identified as associated with seed oil concentration (Eskandari et al. 2013) was also significantly associated with days to maturity, plant height, and 100-seed weight while interacting with different genomic regions (Table 3).

Least square means and broad sense heritability estimates have been calculated for all agronomic and seed quality characteristics in each and across environments (Table 4). Seed yield had the lowest broad sense heritability (ranging from 0.32 to 0.48) among all the traits, whereas days to maturity had the greatest estimate of heritability (ranging from 0.84 to 0.91). Significant transgressive segregation ($P \leq 0.05$) was present in the RIL populations for all traits in each environment and across all environments.

Genetic and phenotypic coefficients of correlation between seed oil concentration and five agronomic and seed traits have been calculated for each environment individually and across combined environments (Table 5). Seed protein concentration and 100-seed weight showed significant negative correlations with seed oil concentration in all environments as well as in the combined data across environments. Days to maturity also showed significant negative correlation with seed oil concentration across most environments and in the combined data, except for Ottawa and Woodstock in 2010 where it was not significant. Seed yield was not significantly correlated with seed oil in most of the environments or in the combined analysis across environments. However, seed yield showed a significant negative phenotypic correlation with oil concentration (−0.24) at Woodstock in 2009 (Table 5).

Discussion

In a previous study (Eskandari et al. 2013), 11 QTL on nine different chromosomes have been identified as associated with seed oil concentration in a RIL population derived from a cross between two moderately high oil soybean cultivars, OAC Wallace and OAC Glencoe, using data from three locations in Ontario, Canada, in 2009 and 2010. To determine if the oil QTL were co-localized with other important agronomic and seed composition traits in that population, five traits including seed yield, size, and protein concentration as well as plant height and days to maturity have been evaluated and QTL analyses were performed for all the markers and traits across the environments using single-factor ANOVA and MQM methods.

Broad sense heritability estimates, which were calculated on a plot basis, were moderate to high for most of the traits with the exception of seed yield (Table 4). The estimates were similar to those previously reported in other soybean QTL mapping populations (Mansur et al. 1993; Orf et al. 1999a; Specht et al. 2001; Kabelka et al. 2004; Hyten et al. 2004; Guzman et al. 2007; Palomeque et al. 2009b; Du et al. 2009). The heritability estimates obtained in this study indicated that a large proportion of the phenotypic variation for the most of the traits was genetic suggesting that genetic gains could be achieved through phenotypic selection. However, high negative genotypic correlation between seed oil and protein concentration, which were detected in all individual environments as well as in the combined analysis across environments (ranging from −0.34 to −0.88), indicated that increasing seed oil composition using conventional selection may occur at the expense of protein concentration and vice versa (Schwender et al. 2003; Chung et al. 2003).

Highly significant negative relationship between seed oil and protein concentrations obtained in this study is well documented in the literature (Wilcox and Shibles 2001; Schwender et al. 2003; Chung et al. 2003; Ray et al. 2006; Bellaloui et al. 2009; Ramteke et al. 2010). It is suggested

Table 1 Putative QTL for selected agronomic and seed traits identified by single-factor ANOVA in a RIL population of OAC Wallace × OAC Glencoe at Ottawa, Ridgetown, and Woodstock in 2009 and 2010

Marker trait	Ch (Pos[a])	R^{2b}	*P* value	Add. effect[c]	Environment
Satt317[d] (OTT09, RID10, WST10, combined)	12 (89.5)				
Yield (kg ha^{-1})[e]		0.06	0.0034	−205.7	WST09
Protein (g kg^{-1})		0.07	0.0006	−5.2	OTT10
		0.11	0.0000	−9.6	Combined
Days to maturity (days)		0.06	0.0023	−2.9	OTT09
		0.06	0.0021	−3.2	WST09
		0.06	0.0013	−2.9	WST10
		0.05	0.0050	−3.4	Combined
		0.05	0.0040	−3.7	OTT09
		0.05	0.0059	−5.8	WST09
Satt335 (OTT10)	13 (77.7)				
Protein (g kg^{-1})		0.05	0.0061	7.1	WST09
		0.05	0.0059	5.3	OTT09
		0.07	0.0010	5.5	OTT10
		0.08	0.0006	7.1	Combined
100-seed weight (g)		0.07	0.0005	0.54	OTT09
		0.05	0.0070	0.50	OTT09
		0.09	0.0002	0.84	WST09
		0.08	0.0002	0.69	WST10
		0.08	0.0003	0.58	Combined
Satt001 (OTT09)	9 (50.6)				
Yield (kg ha^{-1})		0.11	0.0000	176.2	OTT09
		0.04	0.0043	107.8	OTT10
Protein (g kg^{-1})		0.16	0.0000	−10.9	OTT09
		0.09	0.0002	−12.2	WST09
		0.07	0.0029	−6.1	OTT10
		0.12	0.0000	−11.4	RID10
		0.13	0.0000	−11.9	Combined
Sat_020 (OTT09)	9 (103.1)				
Protein (g kg^{-1})		0.06	0.0040	7.9	WST09
		0.05	0.0091	4.8	Combined
Satt463 (combined)	7 (50.1)				
Yield (kg ha^{-1})		0.05	0.0066	88.5	WST09

OTT09 Ottawa 2009, *OTT10* Ottawa 2010, *RID09* Ridgetown 2009, *RID10* Rigdetown 2010, *WST09* Woodstock 2009, *WST10* Woodstock 2010 *Combined* combined environments

[a] Chromosome designation and position as per Song et al. (2004)

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

[c] Additive effect 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 estimates indicates that the higher mean was obtained for the alternate, OAC Glencoe allele

[d] Marker shown in bold indicate those linked to a putative oil QTL as reported by Eskandari et al. (2013). Environments in which markers were associated with seed oil concentration are provided in the parentheses

[e] The values are expressed on a 13 % seed moisture basis

that 1 unit increase in oil concentration will lead to about 2 U reduction in seed protein concentration (Schwender et al. 2003; Chung et al. 2003). This relationship could be due to tightly linked loci governing oil and protein concentration s separately, or because of pleiotropic effects of certain loci (Chung et al. 2003).

Highly significant negative genotypic correlation between oil and 100-seed weight and days to maturity had

Table 2 Putative QTL for selected agronomic and seed traits identified by multiple QTL mapping (MQM) in a RIL population of OAC Wallace × OAC Glencoe at Ottawa, Ridgetown, and Woodstock in 2009 and 2010

Interval trait	Ch (Pos[a])	2009						2010									Combined environments		
		Ottawa			Woodstock			Ottawa			Ridgetown			Woodstock					
		R^2 [b]	LOD[c]	Add. effect[d]	R^2	LOD	Add. effect	R^2	LOD	Add. effect	R^2	LOD	Add. effect	R^2	LOD	Add. effect	R^2	LOD	Add. effect
Satt317-Satt302 [e] (OTT09, OTT10, RID09, WST10, combined env's)	12 (4.3–0.0)																		
Protein (g kg^{-1})[f]		0.12	7.3	−6.60	–	–	–	0.06	3.5	−4.10	0.06	3.0	−3.80	–	–	–	0.08	3.0	−3.80
Days to maturity (days)		0.05	2.6	−1.14	0.05	2.2	−1.85	–	–	–	–	–	–	0.06	3.0	−1.58	–	–	–
Plant height (cm)		–	–	–	0.07	3.5	−4.48	–	–	–	–	–	–	–	–	–	–	–	–
100-seed weight (g)		–	–	–	–	–	–	0.09	4.9	−0.69	–	–	–	–	–	–	0.05	3.0	−0.35
Satt510-Satt335 (OTT10)	13 (63.5–73.5)																		
Protein (g kg^{-1})		0.06	2.8	3.40	–	–	–	0.04	2.5	2.20	–	–	–	–	–	–	0.06	3.0	2.70
100-seed weight (g)		0.08	3.5	0.34	0.06	2.9	0.38	0.09	4.9	0.52	–	–	–	0.08	3.8	0.39	0.07	3.7	0.31
Satt001-Satt273 (OTT09)	9 (34.7–42.3)																		
Yield (kg ha^{-1})		0.07	3.6	96.26	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Protein (g kg^{-1})		0.04	6.4	−3.40	0.07	3.4	−6.70	0.05	2.9	−3.60	0.06	2.9	−0.35	–	–	–	0.07	3.4	−3.00
GmDGAT2B-Satt132 (WST09)	16 (20.8–0.0)																		
Yield (kg ha^{-1})		–	–	–	0.11	5.8	−102.3	–	–	–	–	–	–	–	–	–	0.06	2.9	−46.1
Satt182-Satt523 (OTT10)	19 (0.0–11.5)																		
Protein (g kg^{-1})		–	–	–	–	–	–	0.05	3.0	−3.70	–	–	–	–	–	–	–	–	–
Satt323-Satt463 (combined env's)	7 (24.8–19.7)																		
Protein (g kg^{-1})		–	–	–	–	–	–	0.05	2.8	2.60	–	–	–	–	–	–	–	–	–

[a] As per our previous study (Eskandari et al. 2013)

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

[c] All LOD values were greater than the empirical LOD threshold values in their corresponding test

[d] The estimates of additive effect are based on the OAC Wallace allele. A negative value for the estimates indicates that the higher mean was obtained for the alternate, OAC Glencoe allele

[e] Genomic regions shown in bold indicate those that contain a putative oil QTL according to Eskandari et al. (2013). Environments in which markers were associated with seed oil concentration are provided in the parentheses

[f] The values are on a 13 % seed moisture basis

Table 3 Markers with significant epistatic effects on selected agronomic and seed traits and the amount of phenotypic variation accounted for by each interaction in a RIL population of OAC Wallace × OAC Glencoe at Ottawa, Ridgetown, and Woodstock in 2009 and 2010

Trait	Interaction						
	$Locus_i$	Ch_i	$Locus_j$	Ch_j	R^{2a}_{ij}	*P* value	Environment
Seed yield	Satt301	17	**Satt302**[b]	12	0.12	0.0000	Combined
Seed yield	GmDGAT1B	17	**Satt335**	13	0.11	0.0002	Combined
Seed yield	Satt042	5	**Satt132**	16	0.11	0.0001	RID09
Seed yield	**Satt066**	14	Satt569	13	0.11	0.0002	RID09
Seed yield	Satt260	9	**Satt335**	13	0.12	0.0000	RID10
Seed yield	**Satt302**	12	Satt313	19	0.15	0.0001	WST10
Seed yield	Satt363	6	**Satt510**	13	0.10	0.0001	WST10
Seed protein	**Satt302**	12	Satt712	16	0.11	0.0002	Combined
Days to maturity	**Sat_020**	9	Satt155	5	0.13	0.0001	RID10
Days to maturity	**Satt182**	19	Satt646	4	0.12	0.0001	RID10
Days to maturity	**Sat_020**	9	Satt544	9	0.12	0.0002	WST10
Days to maturity	**Sat_020**	9	Satt155	5	0.11	0.0002	WST10
Plant height	**Satt182**	19	Satt485	3	0.11	0.0001	WST10
100-seed weight	**GmDGAT2B**	16	Satt712	16	0.11	0.0001	OTT09
100-seed weight	Satt150	7	**Satt182**	19	0.12	0.0002	RID09

RID09 Ridgetown 2009, *RID10* Ridgetown 2010, *WST10* Woodstock 2010; *OTT09* Ottawa 2009, *Combined* combined across environments

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

[b] Markers shown in bold indicate the markers individually associated with seed oil concentration as identified previously by either single-factor ANOVA or MQM analysis at any environment in Eskandari et al. (2013)

Table 4 Least square mean (top values), heritability (bottom values), and standard error (in parentheses) for selected agronomic and seed traits in a RIL population of OAC Wallace × OAC Glencoe at Ottawa, Ridgetown, and Woodstock in 2009 and 2010

Trait[a]	Environment						Combined
	2009			2010			
	Ottawa	Ridgetown	Woodstock	Ottawa	Ridgetown	Woodstock	
Yield (kg ha^{-1})	3,633 (22.5)	3,927 (29.6)	3,629 (21.0)	3,813 (18.1)	3,631 (21.2)	3,798 (18.0)	3,735 (9.1)
	0.32 (0.06)	0.48 (0.09)	0.41 (0.10)	0.39 (0.06)	0.35 (0.11)	0.45 (0.09)	0.39 (0.02)
Protein (g kg^{-1})	396.3 (0.51)	396.3 (0.53)	399.4 (0.49)	405.9 (0.69)	400.3 (0.51)	392.8 (0.58)	398.7 (0.34)
	0.92 (0.01)	0.84 (0.03)	0.73 (0.10)	0.86 (0.02)	0.91 (0.01)	0.78 (0.03)	0.81 (0.03)
100-seed weight (g)	18.9 (0.04)	20.1 (0.09)	16.0 (0.08)	20.6 (0.13)	17.1 (0.13)	17.7 (0.09)	19.3 (0.03)
	0.89 (0.02)	0.84 (0.02)	0.73 (0.04)	0.81 (0.02)	0.79 (0.03)	0.73 (0.04)	0.79 (0.03)
Days to maturity (days)	122 (0.2)	92 (0.7)	122 (0.3)	140 (0.8)	97 (0.5)	116 (0.1)	115 (0.1)
	0.91 (0.01)	0.86 (0.01)	0.84 (0.02)	0.88 (0.03)	0.90 (0.01)	0.86 (0.03)	0.85 (0.03)
Plant height (cm)	91.7 (0.4)	90.8 (0.5)	92.1 (0.4)	92.9 (0.4)	94.6 (0.5)	95.9 (0.6)	93.0 (0.3)
	0.85 (0.02)	0.83 (0.03)	0.79 (0.06)	0.68 (0.03)	0.78 (0.02)	0.49 (0.08)	0.76 (0.02)

[a] The values for yield, protein and 100-seed weight are on a 13 % seed moisture basis

been seen in most environments and also in the combined data (Table 5). Negative correlation between oil concentration and maturity was also reported in previous studies (Kabelka et al. 2004; Bellaloui et al. 2009). No significant genotypic correlation was found between seed oil and yield in any of the environments, indicating that this population would be desirable for selection of both high oil and high yield genotypes separately across several environments (Burton 1987; Scott and Kephart 1997).

The QTL analyses identified seven previously reported oil-associated QTL in this population (Eskandari et al. 2013) as co-localized with QTL for either agronomic or seed quality traits (Tables 1, 2). The putative oil QTL in the interval of Satt317-Satt302, located on Chromosome 12

Table 5 Genetic (top values) and phenotypic (bottom values) correlation coefficients between seed oil concentration and five agronomic and seed quality traits evaluated in a RIL population of OAC Wallace × OAC Glencoe at Ottawa, Ridgetown, and Woodstock in 2009 and 2010

Trait	Environment						
	2009			2010			Combined
	Ottawa	Ridgetown	Woodstock	Ottawa	Ridgetown	Woodstock	
Yield (kg ha^{-1})	−0.21	−0.23	−0.16	0.19	0.16	0.21	0.10
	−0.20	−0.10	−0.24*	0.11	0.21	0.18	0.04
Protein (g kg^{-1})	−0.68**	−0.71**	−0.51**	−0.66**	−0.88**	−0.34**	−0.68**
	−0.49**	−0.48**	−0.32**	−0.35**	−0.56**	−0.25**	−0.39**
100-seed weight (g)	−0.28**	−0.25**	NA[a]	−0.26**	−0.23**	−0.32**	−0.29**
	−0.25**	−0.23**	NA	−0.25**	−0.19*	−0.16**	−0.11*
Days to maturity (days)	−0.41**	−0.23*	−0.24**	−0.15	−0.53**	−0.09	−0.30**
	−0.34**	−0.13*	−0.16*	−0.13	−0.26**	−0.05	−0.19**
Plant height (cm)	−0.12*	0.10	0.06	0.23	−0.10	0.22	0.01
	−0.29**	0.05	0.13*	0.14	−0.07	0.17*	0.04

* Represents significance at $P = 0.05$

** Represents significance at $P = 0.01$

[a] Not available

(LG H), was co-localized with seed yield, size, and protein concentration as well as days to maturity and plant height. This oil-enhancing QTL allele, which was inherited from OAC Wallace, was negatively associated with all the co-localized traits. This genomic region was previously reported as carrying putative QTL associated with seed yield per plant (Du et al. 2009) and seed yield per hectare (Kabelka et al. 2004). Kabelka et al. (2004) also found that this genomic region was associated with shorter plants. In another study, Specht et al. (2001) also reported this genomic region as associated with maturity, plant height, and lodging. The current study along with the previous studies (Specht et al. 2001; Kabelka et al. 2004; Du et al. 2009) showed that either several tightly linked genes or a pleiotropic gene in this region is affecting several traits. However, further genetic investigation and fine mapping of the region is suggested to determine whether gene linkage or pleiotropy or a combination of both phenomena caused the relationships.

The putative oil QTL placed between markers Satt510 and Satt335 (Eskandari et al. 2013) was also detected as associated with seed protein concentration and 100-seed weight in several environments as well as across environments (Tables 1 ,2). The oil positive allele of this QTL from OAC Glencoe showed negative impact on both seed size and protein concentration. This genomic region was previously reported to be associated with seed protein content in two different studies (Hyten et al. 2004; Kabelka et al. 2004). Hyten et al. (2004) also reported a seed size QTL in the interval of Satt335-Satt144. The SSR marker Satt144 is 24.4 cM away from Satt335 (Song et al. 2004). Orf et al. (1999a) detected a seed weight QTL tagged by RFLP marker L050_14 (Satt510) in a RIL population derived from Noir 1 × Archer. The results of the present study were in agreement with the previous studies (Orf et al. 1999a; Hyten et al. 2004; Kabelka et al. 2004) that identified a multi-trait QTL within this genomic region as associated with seed size and protein concentration across different environments and genetic backgrounds.

The putative oil QTL in the interval of Satt001-Satt273, which was tagged by Satt001 and detected at Ottawa in 2010 (Eskandari et al. 2013), was identified as associated with seed protein concentration across five of six environments, including Ottawa 2010, and in the combined analysis across environments. This QTL was also associated with seed yield at Ottawa in both years. While the oil-enhancing allele from OAC Wallace was negatively associated with seed protein concentration at Ottawa in 2010, it was positively correlated with seed yield at that location. This putative QTL seemed to be the same as previously reported seed yield QTL by Yuan et al. (2002), Guzman et al. (2007), and Du et al. (2009). Hyten et al. (2004) also reported a seed size QTL in close proximity to this QTL, which was located between Satt518 and Satt273. There are two more previously reported protein-associated QTL close to this QTL: one was linked to markers A065_3, which is positioned within the Satt001-Satt273 interval (Soybase 2011) as reported by Lee et al. (1996), and the other one was tagged by Satt178, which is located 9.7 cM from Satt001 (Song et al. 2004) as reported by Specht et al. (2001). The results indicate that this multi-trait QTL could be used in marker-assisted selections to improve both seed oil and yield simultaneously in specific environments such as Ottawa.

Another putative QTL associated with seed oil located on Chromosome 9 and tagged by marker Sat_020 (Eskandari et al. 2013) was co-localized with seed protein concentration at Woodstock in 2009 and in the combined analysis (Table 1). This QTL is more than 50 cM away from another QTL on the same chromosome within the Satt001-Satt273 interval (Eskandari et al. 2013), indicating that they represent distinct QTL. The oil-enhancing allele from OAC Wallace was also positively correlated with protein concentration. This QTL could be exploited in MAS to increase both oil and protein concentration simultaneously. Csanádi et al. (2001) identified an oil/protein QTL within the genomic region between Sat_020 and Satt196, tagged by Satt196, using single-factor ANOVA and simple interval mapping in an F_2 population of Ma.Belle × Proto. However, their oil positive QTL allele was negatively associated with seed protein content (Csanádi et al. 2001).

The putative oil QTL linked to the gene-specific marker GmDGAT2B on Chromosome 16 (LG J) (Eskandari et al. 2013) was also identified as associated with seed yield at Woodstock in 2009 and in the combined environments (Table 2). The oil-enhancing allele of this gene, which came from OAC Glencoe, caused also an increase in seed yield suggesting its potential use in MAS to elevate both seed oil and yield in new soybean cultivars. This QTL seemed to be in the same genomic region as the seed yield QTL previously reported associated with markers Satt529 and Satt414 (Guzman et al. 2007; Li et al. 2007) being in close proximity to GmDGAT2B. Fine mapping of this region with more molecular markers is suggested to investigate whether these markers represent the same or distinct QTL.

The putative oil QTL on Chromosome 19 (LG L) anchored by markers Satt182 and Satt523 (Eskandari et al. 2013) was also detected as associated with seed protein at Ottawa in 2010. The positive oil QTL allele coming from OAC Glencoe was negatively associated with protein concentration. The closest protein QTL to the current oil QTL on this chromosome that has been reported is the QTL associated with the RFLP marker A023_1 (Diers et al. 1992), which is located at least 8.8 cM from our oil QTL (Song et al. 2004).

In the genomic region between SSR markers Satt323 and Satt463 on Chromosome 7 (LG M), in which an oil QTL tagged by Satt463 was identified in the combined analysis of the environments (Eskandari et al. 2013), we also identified a yield QTL at Woodstock in 2009 and a protein QTL at Ottawa in 2010, both of which were tagged by marker Satt463 (Tables 1 ,2). While for the oil QTL in this genomic region the oil-enhancing allele was contributed by OAC Glencoe, the yield- and protein-enhancing QTL alleles came from OAC Wallace. In this genomic region, their QTL have been previously reported for protein (Hyten et al. 2004), seed yield, plant height, and maturity (Wang et al. 2004), but not for seed oil concentration.

Using simple models for explaining genetic control of complex quantitative traits, it is assumed that individual loci act in additive and independent manners (Falconer and Mackay 1996; Lark et al. 1995). However, the importance of epistatic interactions among loci on polygenic traits was investigated and confirmed in different crops, including soybean (Lark et al. 1995; Orf et al. 1999a, b; Palomeque et al. 2009a, b). The importance of two-way epistatic effects between molecular markers on soybean seed oil accumulation has been also shown in our companion paper (Eskandari et al. 2013). Our results have established that epistatic effects between different chromosomal regions could be important to explain the correlation between seed oil and other agronomic and seed traits. In particular, it was determined that the oil-associated QTL linked to Sat_020 (Eskandari et al. 2013), which was co-localized with a protein concentration QTL was also associated with days to maturity in interaction with different markers across different environments. Or, the oil QTL tagged by Satt182, which was also associated with protein concentration individually, was associated with plant height, seed size, and days to maturity in interaction with different markers in different environments.

In conclusion, we identified an oil-enhancing QTL on Chromosome 9 (LG K) tagged by Sat_020 at Woodstock in 2009, which was also positively associated with protein concentration in the same environment and also the combine analysis of the environments. This QTL could be used in marker-assisted allele introgression to improve both oil and protein concentration of soybean seeds simultaneously, which is one of the goals of soybean breeding. In the current study, we also discovered two oil QTL on Chromosome 9 (LG K, tagged by Satt001) and Chromosome 16 (LG J, tagged by gene-specific marker GmDGAT2B) that were also associated with increased seed yield in the same environments. These QTL could be exploited in molecular breeding programs aimed at elevating both seed oil and yield together. The results of two-way epistatic interactions among molecular markers on agronomic and seed traits, where one of the interacting markers had been associated with seed oil concentration (Eskandari et al. 2013), revealed that co-segregation of seed oil and some other traits in soybean such as seed yield and protein could be caused in part due to epistatic interactions between genomic regions. The results of this study may be helpful in selecting complementary parental lines that could result in the development of high oil cultivars without a penalty on protein concentration and with higher seed yield.

Acknowledgments The authors would like to thank Drs G. R. Ablett (in memoriam), K. P. Pauls, L. R. Erikson, and Y. Kakuda (University of Guelph) for their valuable suggestions on this research. The authors are also grateful to Wade Montminy, Chris Grainger, Ron Guillemette, Bryan Stirling, Dennis Fischer and the entire soybean crew at the University of Guelph for their excellent technical 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

ASA (2011) The American Soybean Association. [online] Available at: http://www.soygrowers.com/

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

Bellaloui N, Smith JR, Ray JD, Gillen AM (2009) Effect of maturity on seed composition in the early soybean production system as measured on near-isogenic soybean lines. Crop Sci 49:608–620

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

Burton JW, Brim CA (1981) Recurrent selection in soybeans: III. Selection for increased percent oil in seeds. Crop Sci 21:31–34

Chung J, Babka HL, Graef GL, Staswick PE, Lee DJ, Cregan PB, Shoemaker RC, Specht JE (2003) The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci 43:1053–1067

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

Csanádi 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

Du W, Wang M, Fu S, Yu D (2009) Mapping QTLs for seed yield and drought susceptibility index in soybean (*Glycine max* L.) across different environments. J Genet Genom 36:721–731

Eskandari M, Cober ER, Rajcan I (2013) 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. Theor Appl Genet 126:483–495

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

Fehr WR, Caviness CE, Burmood DT, Pennington JS (1971) Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. Crop Sci 11:929–931

Feng L, Burton JW, Carter TE Jr, Pantalone VR (2004) Recurrent half-sib selection with testcross evaluation for increased oil content in soybean. Crop Sci 44:63–69

Guzman PS, Diers BW, Neece DJ, Martin SKS, Leroy AR, Grau CR, Hughes TJ, Nelson RL (2007) QTL associated with yield in three backcross-derived populations of soybean. Crop Sci 47:111–122

Hallauer AR, Miranda JB (1988) Quantitative genetics in maize breeding. 2nd edn. Iowa University Press, Ames

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

Holland JB, Nyquist WE, Cervantes-Martinez CT (2003) Estimating and interpreting heritability for plant breeding: an update. Plant Breed Rev 22:9–111

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

Iqbal M, Navabi A, Salmon DF, Yang RC, Spaner D (2007) Simultaneous selection for early maturity, increased grain yield and elevated grain protein content in spring wheat. Plant Breed 126:244–250

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

Keim P, Schupp JM, Travis SE, Clayton K, Zhu T, Shi L, Ferreira A, Webb DW (1997) A high-density soybean genetic map based on AFLP markers. Crop Sci 37:537–543

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 speciWc 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 and Nguyen HT (2007) Soybean, P 1-3. In: Kole C (ed) Genome mapping and molecular breeding in plants. vol 2, Oilseeds, pp 1–3

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

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

Nyquist WE, Baker RJ (1991) Estimation of heritability and prediction of selection response in plant populations. Crit Rev Plant Sci 10(3):235–322

Orf JH, Chase K, Adler FR, Mansur LM, Lark KG (1999a) Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. Crop Sci 39:1652–1657

Orf JH, Chase K, Jarvik T, Mansur LM, Cregan PB, Adler FR, Lark KG (1999b) 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 (2009a) 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

Palomeque L, Liu L, Li W, Hedges B, Cober E, Rajcan I (2009b) QTL in mega-environments: II. Agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines. Theor Appl Genet 119:429–436

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

Ramteke R, Kumar V, Murlidharan P, Agarwal DK (2010) Study on genetic variability and traits Pinterrelationship among released soybean varieties of India [*Glycine max* (L.) Merrill]. Electron J Plant Breed 1(6):1483–1487

Ray JD, Fritschi FB, Heatherly LG (2006) Large application of fertilizer N at planting affects seed protein and oil concentrations

in the early soybean production system. Field Crops Res 99:67–74

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

Schwender J, Ohlrogge JB, Shachar-Hill Y (2003) A flux model of glycolysis and oxidative pentosephosphate pathway in developing *Brassica napus* embryos. J Biol Chem 278:29442–29453

Scott RA, Kephart KD (1997) Selection for yield, protein, and oil in soybean crosses between adapted and introduced parents. Field Crops Res 49:177–185

Smith RR, Weber CR (1968) Mass selection by specific gravity for protein and oil in soybean populations. Crop Sci 8:373–377

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. Available at: http://soybeanbreederstoolbox.org/ (Verified on 20 February, 2012) [Online]

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

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066

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

Wang D, Graef GL, Procopiuk AM, Diers BW (2004) Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor Appl Genet 108:458–467

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

Yuan J, Njiti VN, Meksem K, Iqbal MJ, Triwitayakorn K, Kassem MA, Davis GT, Schmidt ME, Lightfoot DA (2002) Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. Crop Sci 42(1):271–277