Euphytica (2021) 217:24
https://doi.org/10.1007/s10681-020-02755-y

Association studies and QTL mapping for soybean oil content and composition

Luiz Cláudio Costa Silva · Loreta Buuda da Matta · Guilherme Ramos Pereira · Rafael Delmond Bueno · Newton Deniz Piovesan · Andrea J. Cardinal · Pedro Ivo Vieira Good God · Cleberson Ribeiro · Maximiller Dal-Bianco

Received: 31 March 2020 / Accepted: 18 December 2020 / Published online: 20 January 2021

Abstract Soybean oil is one of the most important vegetable oils in world. Increasing oil content and improve its composition is an important aim of many soybean breeding programs. In this context, the objective of this study was to identify QTLs controlling oil and fatty acid contents in soybean seed in different locations. For this, we developed F_2-derived populations by crossing CS303TNKCA and FA22 lines, and analyzed oil and fatty acid content in four locations in Brazil. We evaluated the correlations of traits in each location and between locations and used 1536 SNPs to molecular characterize this population. QTLs associated with the average values of four environments were identified by using simple interval mapping. A linkage map was generated using 534 SNP markers, and 25 linkage groups were formed. A total of 20 QTLs controlling oil, palmitic, stearic, oleic, linoleic and linolenic contents were found, varying from one (linolenic content) to six (palmitic content), and explaining from 7.02% (*qSte-13*) to 70.37% (*qLin-14*) of phenotypic variation. We could not associate eight QTLs (*qPal-02, qOle-02, qLol-02, qOle-04, qLol-04, qPal-13.2, qSte-15 and qSte-17*) with reported QTLs and genes, so these could be set as new QTLs controlling fatty acid composition. The results found in the present work can help understand the genetic basis of these traits and help breeders to

Luiz Cláudio Costa Silva and Lorêta Buuda da Matta have contributed equally to this work.

Supplementary Information The online version contains supplementary material available at (https://doi.org/10.1007/s10681-020-02755-y).

L. C. C. Silva (✉)
Departamento de Ciências Biológicas, Universidade Estadual de Feira de Santana, Feira de Santana, BA 44036-900, Brazil
e-mail: lccsilva@uefs.br

L. B. da Matta
BASF, Soybean Breeding, Rio Verde, GO 75903-970, Brazil

G. R. Pereira · N. D. Piovesan
Instituto de Biotecnologia Aplicada à Agropecuária, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil

R. D. Bueno · M. Dal-Bianco
Departamento de Bioquímica, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil

A. J. Cardinal
Crop Science Department, North Carolina State University, Raleigh, NC 27695, USA

Present Address:
A. J. Cardinal
Syngenta Vegetable Seeds, Raleigh-Durham, NC 27709, USA

modify oil content and composition in soybean seed, especially in tropical conditions.

Keywords Soybean oil · Fatty acid content · Soybean composition · Soybean breeding · Plant breeding

Introduction

Soybean is the major crop cultivated in Brazil, which is the second largest producer worldwide. In the 2018/2019 harvest, Brazil reached the production of 115 million grain tons, corresponding for about 47.6% of the national grain production (CONAB 2019).

The principal compounds in soybean grain are protein (about 40%) and oil (about 20%). Soybean oil is the second vegetable oil produced in the world, corresponding to about 28% of vegetable oil consumption (USDA 2019). In Brazil, soybean is historically used by food purposes, but in recent years the demand for soybean oil for biodiesel production has increased. This was mainly due to the government efforts to reduce the dependency of fossil fuels by increasing about 228% biodiesel production in the last 10 years (Brazil 2016; ANP 2019).

Soybean oil is composed by 13% of palmitic acid (16:0), 4% of stearic acid (18:0), 20% of oleic acid (18:1), 55% of linoleic acid (18:2) and 8% of linolenic acid (18:3), on average (Goettel et al. 2014; Silva et al. 2018a; Zhang et al. 2018). The content of these fatty acids is determinant for the quality of soybean oil, influencing its physicochemical, functional and nutritional properties (Silva et al. 2018a; Kim et al. 2019). Saturated fatty acids, such as palmitic and stearic acids, are more stable to oxidative degradation but their consumption is associated with low-density lipoprotein (LDL) cholesterol increase, coronary disease and some types of cancer (Wang et al. 2012a, b; Kim et al. 2019). However, recent results indicate that stearic acid does not have this cholesterolemic effect in humans (Hunter et al. 2009; Heim and Gillman 2017). Oils with high levels of monounsaturated fatty acids like oleic acid are oxidatively stable and useful for industrial applications, such as biodiesel, paints, lubricants and cosmetics production, and food proposals (Maheshwari and Kovalchuk 2016; Woyann et al. 2019). Linoleic and linolenic acids are polyunsaturated fatty acids (PUFA) that are considered essential for humans because the body cannot synthesize them (Patterson et al. 2012). However, a high concentration of PUFA in soybean oil is the main cause of low oxidative stability and rancidification, altering the flavor and reducing storage time of food products (Warner and Fehr 2008).

Fatty acid desaturases (FAD) are the main enzymes that determine the content of oleic, linoleic and linolenic acids in soybean seed. The omega-6-desaturases *GmFAD2-1A* (*Glyma.10g278000*) and *GmFAD2-1B* (*Glyma.20g111000*) convert oleic acid into linoleic acid, and the omega-3-desaturases *GmFAD3A* (*Glyma.14g194300*), *GmFAD3B* (*Glyma.02g227200*) and *GmFAD3C* (*Glyma.18g062000*) convert linoleic acid into linolenic acid (Baud and Lepiniec 2010). Other enzymes associated with fatty acid content in soybean seed are delta-9-stearoyl-acyl carrier protein desaturases (*GmSACPD-A-Glyma.07G207200*; *GmSACPD-B-Glyma.02G138100*; *GmSACP-C-Glyma.14G121400*), which convert stearic acid into oleic acid (Gillman et al. 2014; Kim et al. 2019), and fatty acyl-ACP thioesterases (*GmFATB1a-Glyma.05g012300*; *GmFAT1B-Glyma.17g120400*; *GmFATB2a-Glyma.04g151600*; *GmFATB2b-Glyma.06g211300*), which hydrolyze acil-ACPs and produce free fatty acids (Weselake et al. 2009; Vogel et al. 2019).

Increasing oil content and improving its composition are two of the main goals of soybean breeding programs (Warner and Fehr 2008; Wang et al. 2012a, b; Pham et al. 2014, Silva et al. 2018a). In this context, molecular markers have been used successfully to increase the content of oleic acid and reduce the contents of PUFA to produce oil with increased oxidative stability (Pham et al. 2010, 2012). This is a healthier alternative to the hydrogenation, a process that converts PUFA into oleic acid but generates *trans* isomers of fatty acids (molecules which are associated with some heart diseases, high

P. I. V. G. God
Universidade Federal de Viçosa, Campus Rio Paranaíba, Rio Paranaíba, MG 38810-000, Brazil

C. Ribeiro
Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil

cholesterol levels and development of diabetes type 2) (Hu et al. 1997; Pham et al. 2014).

The oil content in soybean seed and its composition are considered quantitative traits (Burton et al. 1983; Bachlava et al. 2009), and thus controlled by quantitative trait loci (QTLs). Therefore, QTL mapping is widely used to identify markers associated with the content and composition of soybean oil (Wang et al. 2012a, b; Li et al. 2015; Heim and Gillman 2017; Zhang et al. 2018; Priolli et al. 2019). Although SSR and SCAR markers continue to be used for QTL mapping in soybean (Shu et al. 2013; Leite et al. 2016; Yu et al. 2018), most recent QTL mapping studies in soybean have used SNP markers (Heim and Gillman 2017; Akond et al. 2018; Zhang et al. 2018; Priolli et al. 2019; Seo et al. 2019). The results of QTL mapping analysis can be used to identify genes controlling important traits as well as identify molecular markers associated with these traits and to estimate their effects on the traits. Heim and Gillman (2017) evaluated a RIL population in three environments and found that a large genomic deletion of A6 soybean line is associated with a dramatic increase of stearic acid levels (about 24–28%), and suggested that *GmSACPD-C* and at least one other gene in this region are responsible for the variation of the trait. Zhang et al. (2018) used genome-wide scan to find many candidate genes for the variation of various soybean seed composition traits. Priolli et al. (2019) found 19 SNPs associated with palmitic, oleic and oil contents, and suggested various candidate genes to controlling oil and palmitic deposition in soybean seed.

Previously, we developed a soybean variety line CS303TNKCA which contains low linolenic acid content (3.5–4%), absence of lipoxygenases and the Kunitz protease inhibitor, and characterized a mutation on its *GmFAD3A* gene that is responsible for the reduced linolenic acid content of CS303TNKCA. Here, we used 534 SNP markers to detect QTLs associated with oil and fatty acid content in a segregating population derived from CS303TNKCA variety and evaluated in four environments. We found 20 QTLs associated with these traits, including five QTLs that was not associated with previously identified QTLs or genes. The results found in the present work can help understand the genetic basis of these traits and help breeders to modify oil content and composition in soybean seed, especially in tropical conditions.

Materials and methods

Plant material

The population used in this study was generated by crossing two contrasting lines, CS303TNKCA and FA22. The CS303TNKCA variety was developed by Programa de Melhoramento da Qualidade da Soja from Universidade Federal de Viçosa (PMQS/BIOAGRO/UFV), and contains middle oil levels in the seed (20–23%), low linolenic acid content (about 3.5–4%) and absence of lipoxygenases (*lox1*, *lox2* and *lox3*) and Kunitz protease inhibitor. CS303TNKCA was developed by backcrossing BARC-12 (Leffel 1994) into the recurrent variety Monarca (COOPADAP, Brazil). The FA22 line has middle oleic acid content (about 50%), about 19% of oil content, and was developed in the Iowa State University by chemical mutagenesis (Alt et al. 2005).

To generate the segregating population, CS303TNKCA and FA22 soybean lines were crossed in a greenhouse at Universidade Federal de Viçosa (Viçosa, Minas Gerais (MG), Brazil; 20°45′14″S, 42 52′55″W). Two hundred and eleven $F_{2:3}$ families were cultivated in the experimental field at Visconde do Rio Branco-MG, Brazil (VRB; 21 00′37″S, 42 50′26″W), in February to June 2010. The experiment was performed in a completely randomized design with two replications and an experimental plot of 25 plants per 1.5-m row and a row spacing of 0.5 meters. $F_{2:4}$ families were evaluated in three experimental fields, one in Viçosa-MG, Brazil (VIC), in December 2010 to April 2011, and two in São Gotardo-MG, Brazil (SG1 and SG2; 19°18′39″S, 46°02′56″W), in November 2010 to March 2011. These experiments were performed in a completely randomized block design with four replications and an experimental plot of 25 plants per 1.5-m row and a row spacing of 0.5 meters.

Phenotypic analysis

The oil content of CS303TNKCA, FA22 and each $F_{2:3}$ and $F_{2:4}$ family was determined by nuclear magnetic resonance spectrometry (NM—Resonance Instruments, Witney, Oxfordshire, UK), in dry basis and expressed in g kg^{-1}. To evaluate the fatty acid composition, lipids from soybean seeds were extracted using the following method: 15 mg of powdered soybean seed were mixed with 1 mL of hexane using

N_2 and stored at 4 °C for 16 h. The hexane solution was collected into another tube and evaporated using N_2. Then, 0.4 mL of 1 M sodium methoxide was added to the tube, shaken and incubated at 30 °C for 1 h. Then, 1 mL of milliQ water and 1 mL of hexane was added in this order, shaken and incubated at room temperature for 1 h. Finally, 0.75 mL of the organic phase was collected, added to anhydrous sodium sulfate to remove the moisture and transferred to a vial tube. The fatty acid composition was performed by gas chromatography in a GC-2010 Plus chromatograph (Shimadzu), as reported by Burkey et al. (2007). The content of each fatty acid was expressed as a proportion of total fatty acids in g kg^{-1}.

Genotypic analysis

Genomic DNA from CS303TNKCA, FA22, and each F_2 plant was extracted by using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI). The concentration of extracted DNA was determined by using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE), and the quality was checked by the 260/280 and 260/230 ratios, as well as by 0.8% agarose gel electrophoresis. Genomic DNA was diluted to 100 ng/μL.

Genotyping of F_2 population was performed by using Illumina GoldenGate from USDA/ARS Soybean Genomics and Improvement Laboratory (Beltsville, MD), using 1536 SNP from Universal Soy Linkage Panel (USLP 1.0), as described by Hyten et al. (2010). The analysis of alleles in each locus was performed using GenomeStudio software (Illumina Inc., San Diego, CA).

Segregating test and map construction

The goodness-of-fit of markers to the expected segregation ratio of 1:2:1 in the F_2 population was checked by Chi-Square at 5% probability. The linkage map was constructed using a minimum log-of-odds (LOD) threshold of 3.0 and a maximum recombination frequency of 30%. Recombination frequencies were converted into genetic distances in centiMorgans (cM) by using the Kosambi function (Kosambi 1944). In case of more than one linkage group per chromosome was generate, the linkage groups were analyzed together. Segregating test and map construction were performed in Genes software (Cruz 2013). The average distance between markers was calculated for any soybean linkage group by using only non-fully linked markers.

QTL detection

Descriptive statistics were calculated for each environment and the average of all experiments (joint analysis). Frequency distribution graphs were generated for the average content of oil and fatty acids, and the normality of the data was checked by the Lilliefors Test (Lilliefors 1967). Correlations between traits in each environment and for the same trait in all environments were calculated by the Pearson coefficient.

Analysis of deviance was performed by restricted maximum likelihood (REML) method to evaluate de significance of the genotype, environment and genotype by environment interaction effects using software Selegen (Resende 2016) and R (R Core Team 2017). Heritabilities based on means were calculated using the formulas:

$$h_g^2 = \sigma_g^2 / \left(\sigma_g^2 + \sigma_e^2 / r \right), \text{ for individual analysis.}$$

$$h_g^2 = \sigma_g^2 / \left(\sigma_g^2 + \sigma_e^2 / r + \sigma_{ge}^2 / re \right), \text{ for joint analysis.}$$

where h_g^2 is the heritability coefficient, σ_g^2 is the genotypic variation, σ_e^2 is the residual variation, σ_{ge}^2 is the genotype by environment interaction variation, e is the number of environments and r is the number of repetitions. Due to the variation in the number of repetitions between and among environments, we used a harmonic mean of the repetition number for calculating the heritabilities.

QTL detection was performed in software Genes (Cruz 2013) for each environment and for the average values of the four environments using only markers that segregated as expected for the F_2 population. Simple interval mapping (Lander and Botstein 1989) was performed by linear regression model, at distance intervals of 0.1 cM. QTL positions were determined using a LOD threshold of 3.0. The coefficient of determination (R^2) was used to evaluate the proportion of phenotypic variation that is explained by each marker. Dominant and additive effects were estimated. Only QTLs that were identified in the joint analysis (LOD > 3.0) were considered in this study. QTLs were named using an abbreviation of the trait

followed by the number of the chromosome and a sequential number when more than one QTL was identified in the same chromosome for the same trait.

The position of the SNPs flanking the QTLs was identified in Soybase (Grant et al. 2009) and used to position each QTLs in the soybean genome. Using this information, we searched in Soybase and Phytozome (Goodstein et al. 2011) databases and previous literature for nearby genes and QTLs that could be associated with the QTLs found in this study.

Results and discussion

Phenotypic analysis

Descriptive statistics (minimum, maximum, average and standard deviation) and values of kurtosis and skewness for each environment and joint analysis are shown in Table 1. There was a substantial amplitude for each trait in each environment, varying from 86.71 to 121.12 g kg^{-1} for palmitic acid, 22.21–51.39 g kg^{-1} for stearic acid, 184.01–577.42 g kg^{-1} for oleic acid, 272.46–618.07 g kg^{-1} for linoleic acid, 27.06–84.05 g kg^{-1} for linolenic acid and 180.06–236.31 g kg^{-1} for oil content. Transgressive segregation was observed for all traits in all environments except for palmitic acid in the joint analysis. Transgressive segregation has been reported in soybean segregating populations for oil and fatty acid contents, which matches to the concept that the traits are controlled by multiple genes with small effects (Cao et al. 2017; Heim and Gillman 2017; Smallwood et al. 2017; Bueno et al. 2018; Silva et al. 2019).

Kurtosis and skewness indicated normal distribution for all traits in joint analysis, except linolenic content (positive skewness at 5% probability). However, significant kurtosis and skewness were observed in individual analysis. Linolenic and also oleic content did not show normal distribution in joint analysis at 1% and 5% probability, respectively (Fig. 1). For linolenic acid, this fact could be explained by a gene (*GmFAD3A-Glyma.14g194300*) that was previously described having a major effect in the trait (Bilyeu et al. 2005; Pinto et al. 2013). Indeed, we recently described a mutation in *GmFAD3A* of CS303TNKCA variety, and individuals carrying this mutation had 41.4 g kg^{-1} of linolenic acid on average (Silva et al. 2018a), a value that matches the second peak in linolenic acid distribution (Fig. 1).The presence of QTL with major effect can distort the normal distribution curve since the values will tend to approximate the averages of each QTL genotype and not only to the mean of the trait.

The traits showed significant genotypic effect at 1% probability in all environments and in joint analysis except oil content in Vicosa environment (Table 2), indicating possibility of gains with selection for all traits. There also was significant effect of environment for all traits in joint analysis. Significant genotype by environment effect was observed for all fatty acid contents at 1% probability but not for oil content. Fatty acids in soybean are typical quantitative traits and are affected by environment and genotype by environment interaction effects (Matei et al. 2018; Zhao et al. 2019). Indeed, Brazil has a markable environment variability and it can change the behavior of soybean crop, affecting the quantity and composition of seed oil. For example, temperature variations can alter the balance of saturated and unsaturated fatty acids and the balance of protein and oil content. Generally, as the average temperature increases the quantity of oil and saturated fatty acid increases, while in lower average temperatures there is an increase of protein and unsaturated fatty acids content (Baud and Lepiniec 2010; Matei et al. 2018).

Heritability values for fatty acid content varied from 53.86% (oleic content/VIC) to 81.46% (linolenic content/VRB) in the individual analysis, and from 77.82% (stearic content) to 89.62% (linolenic content) in joint analysis. These are considered good values for quantitative traits and point to good perspectives of selection gains. Heritability values for fatty acid content have varied in many studies, usually ranging from 40 to 90% (Hyten et al. 2004; Leamy et al. 2017; Li et al. 2017; Priolli et al. 2019; Woyann et al. 2019). For oil content, heritability values in each environment varied from 16.51% (VIC) to 56.70% (VRB) and joint analysis showed a higher value (66.60%). Oil content is also a quantitative trait and different values of heritability are showed in QTL mapping studies such as values below 50% (Leamy et al. 2017) and above 80% (Priolli et al. 2019; Woyann et al. 2019). These results show that the evaluation of oil content and composition in multiple environments can lead to a higher values of heritability, increasing the selection gains.

Table 1 Descriptive statistics, kurtosis and skewness for oil and fatty acid contents of F_2-derived populations from CS303TNKCA × FA22 crossing, evaluated in four environments in Brazil

Trait	Minimum	Maximum	Average	SD	Kurtosis	Skewness	CS303TNKCA	FA22
VRB								
Palmitic	90.59	115.92	102.58	4.47	2.943	− 0.080	111.20	89.15
Stearic	33.10	51.39	41.02	3.16	2.979	0.245	35.72	37.46
Oleic	214.71	411.56	280.31	34.93	3.775*	0.746**	198.28	396.90
Linoleic	405.19	590.74	519.57	31.71	3.586	− 0.549**	615.54	414.14
Linolenic	38.60	80.22	56.51	9.66	2.145*	0.097	39.26	62.35
Oil	191.07	236.31	217.45	8.96	2.883	−.338*	231.21	188.10
VIC								
Palmitic	87.90	116.19	100.33	5.35	2.730	0.077	120.99	88.34
Stearic	24.32	40.10	31.53	2.96	3.151	0.413*	39.55	27.21
Oleic	256.86	577.42	418.03	62.74	2.745	− 0.002	216.26	560.94
Linoleic	272.46	548.92	409.46	54.47	2.800	0.030	580.82	277.31
Linolenic	27.06	63.46	40.65	6.69	3.282	0.570**	42.38	46.20
Oil	203.79	231.53	217.96	5.24	2.881	− 0.147	196.33	195.20
SG1								
Palmitic	86.71	121.12	103.32	5.76	3.155	− 0.049	110.84	83.94
Stearic	26.75	43.81	33.97	3.01	3.263	0.472**	33.70	26.43
Oleic	192.50	523.70	315.60	64.44	2.952	0.459**	193.67	503.64
Linoleic	308.31	609.73	493.53	56.85	2.886	− 0.423*	623.78	326.33
Linolenic	36.58	84.05	53.58	9.76	2.616	0.520**	38.01	59.66
Oil	180.66	229.90	210.12	7.18	4.475**	− 0.479**	211.59	192.99
SG2								
Palmitic	88.05	118.62	102.97	5.80	2.780	0.107	114.99	85.29
Stearic	22.21	47.04	34.25	3.45	3.958**	0.065	34.20	26.94
Oleic	184.01	505.05	317.08	63.90	2.603	0.291	212.09	502.98
Linoleic	344.33	618.07	492.57	55.15	2.547	− 0.239	601.33	325.45
Linolenic	34.03	80.54	53.13	9.62	2.958	0.469**	37.39	59.35
Oil	191.00	225.46	210.30	6.00	3.125	− 0.356*	206.47	183.76
JOA								
Palmitic	91.77	114.15	102.30	4.46	2.586	− 0.076	114.51	86.68
Stearic	29.31	43.13	35.20	2.48	3.059	0.191	35.79	29.51
Oleic	221.68	480.34	332.75	46.75	2.884	0.297	205.08	491.12
Linoleic	345.41	581.70	478.78	41.46	3.006	− 0.232	605.37	335.80
Linolenic	36.26	71.69	50.97	7.96	2.598	0.363*	39.26	56.89
Oil	201.20	227.45	213.96	4.93	2.799	− 0.039	211.40	190.01

Values in g kg^{-1}; *SD* Standard deviation; *VRB* Visconde do Rio Branco-MG, Brazil; *VIC* Vicosa-MG, Brazil; *SG1* Sao Gotardo-MG, Brazil (1); *SG2* Sao Gotardo-MG, Brazil (2); *JOA* Joint analysis

Pearson coefficients of correlation between traits in each environment and between environments are shown in the supplementary information (Table S1). Oil content was the trait that varied most in different environments, with an average coefficient of correlation of 0.3388 for the trait in different locations. On the other hand, linolenic acid content had a lesser variation between environments, with an average coefficient of correlation of 0.7205 between the trait in the four locations. Regarding the correlation

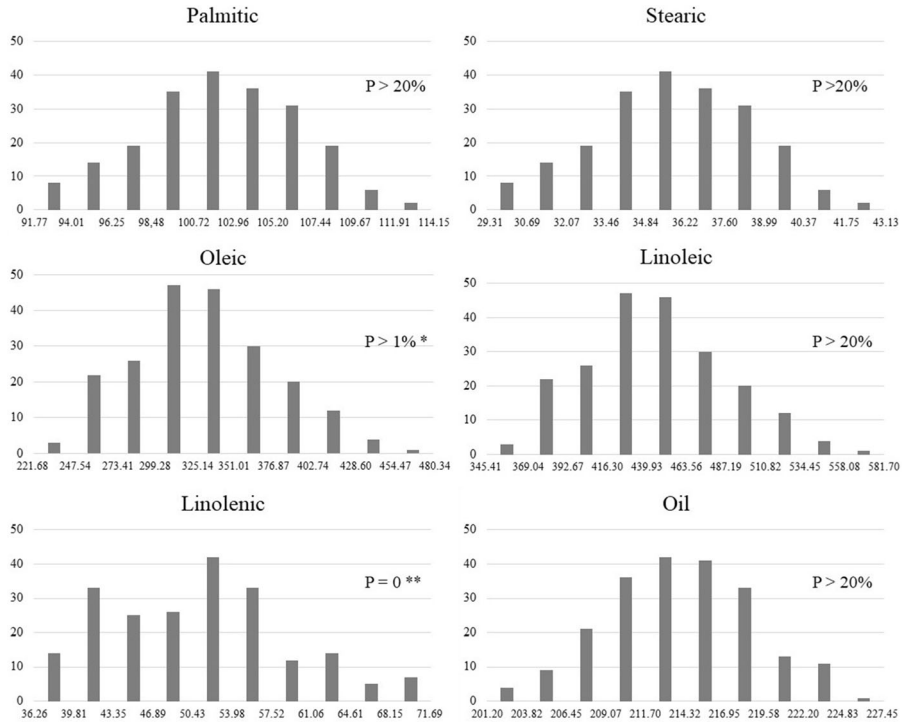

Fig. 1 Distribution of average oil and fatty acid contents in F_2-derived populations from CS303TNKCA × FA22 crossing. The *P* value indicates the probability of following normal distribution

between traits in each environment and the average values of the same traits (joint analysis), the coefficients of correlation varied from 0.5780 (oil content in VIC location) to 0.9150 (linolenic content in SG1 location).

In all environments and joint analysis, we observed a strong negative correlation between oleic and linolenic contents, ranging from − 0.9572 (VRB) to − 0.9928 (VIC). Negative correlations between these traits were well reported, likely due to genes that present pleiotropic effects on both traits (Bachlava et al. 2008; Leamy et al. 2017; Li et al. 2017; Smallwood et al. 2017; Silva et al. 2019). We also observed negative correlations between palmitic and oleic acids (− 0.4828 to − 0.7687), between oleic and linolenic acids (− 0.2901 to − 0.5157), and between oleic acid and oil content (− 0.1739 to − 0.5903). Positive correlations were found between palmitic and linoleic acids (0.3770 to 0.7272), and between linoleic and oil contents (0.2024 to 0.5789).

Correlations between fatty acids are expected since they are at the same metabolic pathway. Palmitic acid is converted to stearic acid by the addition of two carbons, a process that involves the enzymes very-long-chain 3-oxoacyl-CoA reductase 1, very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase and very-long-chain enoyl-CoA reductase (Kanehisa and Goto 2000). The production of unsaturated fatty acids is mediated by desaturases, enzymes that add double bounds in specific positions of the carbon chain. Stearoyl-ACP desaturases convert stearic into oleic acid, omega-6-fatty acid desaturases convert oleic into linoleic acid, and omega-3-fatty acid dessaturases convert linoleic into linolenic acid, by the addition of a double bounds after carbons 9, 12 and 15, respectively (Kanehisa and Goto 2000). Since the production of these fatty acids depends on the amount of the fatty acids upstream in the metabolic pathway, they will be correlated, as was observed in our results and by many authors (Leamy et al. 2017; Li et al. 2017; Smallwood et al. 2017; Silva et al. 2019).

SNP selection and map construction

We genotyped 211 soybean F_2 segregating plants using 1536 SNPs from Illumina GoldenGate Assay

Table 2 Analysis of deviance for oil and fatty acid contents of F_2-derived populations from CS303TNKCA x FA22 crossing, evaluated in four environments in Brazil

	Pal	Ste	Ole	Lol	Lin	Oil
VRB						
DEV-G	91.56**	59.85**	82.94**	92.6**	110.26**	32.43**
Vg	14.84	6.65	875.96	743.18	74.20	44.96
Ve	9.95	6.69	625.90	476.29	38.33	67.69
h^2g	74.63%	66.19%	73.40%	75.47%	79.24%	56.70%
VIC						
DEV-G	122.9**	58.80**	119.18**	120.90**	219.32**	3.72ns
Vg	20.45	5.07	2759.88	2088.12	36.30	5.35
Ve	22.89	11.13	3151.57	2362.42	21.18	69.42
h^2g	69.60%	53.86%	69.18%	69.38%	81.46%	16.51%
SG1						
DEV-G	142.54**	58.21**	136.67**	159.08**	176.90**	41.25**
Vg	22.93	4.63	2930.59	2365.46	70.22	23.43
Ve	28.04	10.96	3894.18	2751.55	70.63	73.73
h^2g	70.17%	54.84%	68.41%	71.21%	74.10%	47.76%
SG2						
DEV-G	133.83**	121.63**	131.04**	141.30**	168.35**	29.20**
Vg	22.97	7.76	2815.55	2146.72	66.65	15.64
Ve	29.26	9.99	3749.80	2684.18	67.62	64.14
h^2g	67.87%	67.65%	66.90%	68.28%	72.62%	39.61%
JOA						
DEV-G	359.33**	228.27**	290.38**	319.65**	525.32**	122.57**
DEV-GxE	14.26**	24.81**	23.37**	23.10**	12.72**	3.17ns
E (F-value)	29.88	596.24	436.62	397.11	400.66	130.50
Vge	2.40	1.31	400.80	288.20	4.70	3.21
Vg	17.64	4.70	1966.30	1566.90	56.14	15.75
Ve	24.49	10.02	3173.60	2292.20	52.70	70.22
h^2g	85.16%	77.82%	82.37%	83.77%	89.62%	66.60%

Values in $g.kg^{-1}$; *Pal* palmitic content; *Ste* stearic content; *Ole* oleic content; *Lol* linoleic content; *Lin* linolenic content; *DEV-G* deviance for genotype effect; *DEV-GxE* deviance for genotype by environment interaction effect; *E* environment effect; *Vg* genotypic variance; *Ve* residual variance; *Vge* genotype by environment interaction effect variance; h^2g heritability based on averages; *VRB* Visconde do Rio Branco-MG, Brazil; *VIC* Vicosa-MG, Brazil; *SG1* Sao Gotardo-MG, Brazil (1); *SG2* Sao Gotardo-MG, Brazil (2); *JOA* Joint analysis

and found 595 polymorphic SNPs. Illumina GoldenGate Assay was created with SNPs on evenly distributed in each of the 20 consensus linkage groups of soybean, having a wide range of allele frequencies in different germplasm and its SNPs were mapped in the integrated molecular linkage map (Gutierrez-Gonzalez et al. 2011; Hyten et al. 2010). The polymorphic SNPs percentage (38.74%) was higher than that found by Vuong et al. (2016; 31.5%), but lower than that found by Gutierres-Gonzalez et al. (2011; 55.33%); both studies used the same set of SNPs.

Chi square test revealed that 58 polymorphic markers (9,7%) did not segregated as expected for F_2 population and was removed from the subsequent analysis. Of these markers, 42 markers (72.4%) had predominant alleles from parent CS303TNKCA while 16 markers (27.6%) had predominant alleles from

parent FA22. Segregation distortions is a common biological phenomenon that can be caused by small population size as well as by genetic factors affecting the inheritance of the markers (Nodar et al. 1993; Wang et al. 2012a, b; Silva et al. 2018b). According to Wang et al. (2012a, b), the skewed segregation ratio is higher in RIL populations than in F_2 or backcross populations.

In the mapping analysis, 537 markers were used to construct the linkage map. A total of 534 markers were distributed in 25 linkage groups (Supplementary Fig. S1) which were positioned in the 20 soybean chromosomes, while the remaining three markers were not included in any linkage groups (BARC-035199-07136, BARC-029937-06757, and BARC-042281-08231). The lack of polymorphic markers closely flanking these three markers may be the cause of them had not being included in linkage groups because this could have led to recombination frequencies higher than 30% between these three markers and the markers flanking. BARC-035199-07136 marker is located at position 1,030,857 of Gm01 chromosome but the first marker included at D1a linkage group (BARC-060833-16926) is located at 4,831,739 position of Gm01 (Grant et al. 2009). Similarly, BARC-042281-08231 marker is located at position 343,168 of Gm20 but the first marker included at I linkage group (BARC-042281-08231) is located at 16,731,974 position of Gm20 (Grant et al. 2009). Gm06 was formed by two linkage groups in this study (5 and 25). BARC-029937-06757 marker (10,900,876 position of C2 likage group) is placed in Gm06 between the last marker of linkage group 5 (BARC-065139-19125, position 4,422,517) and the first marker of linkage group 25 (BARC-021735-04194, position 15,980,506) (Grant et al. 2009).

Table 3 shows information about the distribution of polymorphic SNPs in the 20 chromosomes of soybean. Five soybean linkage groups were composed of two linkage groups in this study (C2/Gm06, A2/Gm08, B1/Gm11, F/Gm13, and G/Gm18) and fifteen chromosomes were composed by one linkage group. The number of markers per chromosome varied from 12 (Gm12) to 45 (Gm02), covering from 52.69 cM (Gm11) to 154.93 cM (Gm02), totaling 2067.17 cM, an average of 103.36 cM per chromosome and 5.10 cM between markers. Recent linkage map studies in soybean have found similar overall lengths of 2054.50 cM (Cao et al. 2017), 2201.4 cM (Heim and Gillman 2017), 2534.42 cM (Li et al. 2017), but overall lengths for soybean genome greater than 3000 cM were also found (Fan et al. 2015; Seo et al. 2019). Gm06 had the shortest average distance between markers (3.05 cM) and Gm12 had the longest average distance (12.12 cM). Gm01 and Gm17 were the only chromosomes that did not presented full linked markers (only bins containing a single marker), and the percentage of bins with two or more markers varied from zero (Gm01 and Gm17) to 47.22% (Gm13).

QTL mapping

A total of 20 QTLs controlling oil content or fatty acid content were found in the joint analysis (Table 4), varying from one (linolenic content) to six (palmitic content) per trait. The number of QTLs per chromosome was one (Gm05, Gm06 and Gm15), two (Gm14 and Gm17), three (Gm04, Gm 13 and Gm19) and four (Gm02). These QTLs explained from 7.02% (*qSte-13*) to 70.37% (*qLin-14*) of phenotypic variation in the joint analysis. The coefficients of determination for individual environments are also shown in Table 4. All QTLs for fatty acid and oil content had predominantly additive effects, agreeing with other studies (Monteros et al. 2008; Reinprecht et al. 2009; Wang et al. 2012a, b; Pinto et al. 2013, Heim and Gillman 2017; Bueno et al. 2018; Silva et al. 2018b).

The interval between flanking markers ranged from 0.20 cM (*qSte-06* and *qPal-17*) to 18.73 cM (*qLin-14*). The occurrence of monomorphic markers was observed within the intervals of 15 QTLs, and markers that showed segregation distortion were observed in three QTLs. These could be the cause of the wide intervals observed for some flanking markers, notably for *qPal-05* (13.51 cM) and *qLin-14* (18.73 cM). Despite presenting the largest interval, *qLin-14* showed the highest R^2 (70.37%).

Six QTLs were found to be associated with palmitic acid content explaining 7.03–12.56% of the trait variation in five chromosomes, totaling 58.26% of phenotypic variation. These QTLs were also identified in four (*qPal02*) three (*qPal-05* and *qPal-19*), two (*qPal-13.1* and *qPal-17*) and one (*qPal-13.2*) environments. All the alleles that reduce palmitic content were from FA22 line.

The *qPal-05* QTL presented the highest coefficient of determination (12.83%) and had a significant effect

Table 3 Coverage of the twenty chromosomes of soybean from a linkage map derived from 534 SNP markers

Chromosome	Soybean LG	LG	NM	L (cM)	AD (cM)	LD	D ≤ 5 cM (%)	PB1 (%)	PB2 (%)
1	D1a	23	16	81.21	5.41	16.03	66.67	100.00	0.00
2	D1b	1	45	154.93	3.87	16.46	72.50	82.22	17.78
3	N	2	32	104.35	4.35	21.18	75.00	59.38	40.63
4	C1	3	33	109.20	4.37	22.87	80.00	60.61	39.39
5	A1	4	29	123.88	5.16	34.35	79.17	79.31	20.69
6	C2	5 and 25	36	79.60	3.06	22.14	84.62	61.11	38.89
7	M	6	17	61.04	5.09	30.07	66.67	52.94	47.06
8	A2	7 and 22	19	96.03	6.86	29.93	64.29	68.42	31.58
9	K	8	40	123.85	3.54	11.82	74.29	80.00	20.00
10	O	9	35	136.07	4.39	21.3	79.41	85.71	14.29
11	B1	10 and 24	15	53.69	4.88	19.42	54.55	80.00	20.00
12	H	11	12	108.11	10.81	32.82	40.00	80.00	20.00
13	F	12 and 13	36	79.46	3.31	15.25	75.00	52.78	47.22
14	B2	14	20	101.32	5.63	32.97	66.67	90.00	10.00
15	E	15	23	135.22	6.76	29.51	55.00	82.61	17.39
16	J	16	24	108.27	6.02	24.61	61.11	62.50	37.50
17	D2	20	24	130.14	5.66	33.27	73.91	100.00	0.00
18	G	17 and 21	35	96.04	3.20	18.16	83.33	85.71	14.29
19	L	18	22	106.60	5.33	16.59	60.00	90.91	9.09
20	I	19	21	78.16	4.34	15.03	66.67	80.95	19.05

Soybean LG Soybean linkage groups; *LG* Linkage groups from 534 SNP markers; *NM* Number of markers; *L* Coverage for each soybean chromosome; *AD* Average distance between markers; *LD* Longest distance between markers; *D ≤ 5* cM Percentage of distances between markers less than 5 cM; *PB1* Percentage of bins containing a single marker; *PB2* Percentage of bins containing two or more markers

in three environments. This QTL was placed in a wide region (432,530–7,943,521 bp) where QTLs and genes controlling palmitic content were found. Li et al. (2015) evaluated 421 soybean accession and mapped a QTL controlling palmitic acid content between positions 1,127,438 and 1,131,632 bp of Gm05 (R^2 = 4.0%), a region where is located *GmFATB1a* (*Glyma.05g016500*), a gene that codifies for strictosidine synthase and have been associated with palmitic content (Cardinal et al. 2007; Thapa et al. 2016). Previously, Cardinal et al. (2007) found that a deletion in *GmFATB1a* is responsible for the reduced palmitic content in soybean lines N97-3681-11 and N97-3708-13. Bachleda et al. (2016) developed a TaqMan marker to target a mutation in *GmFATB1a* which could reduce palmitic content from 110 g kg^{-1} to 66 g kg^{-1} (R^2 = 57.3%) in an F_2 population derived from low palmitic acid N87-2122-4 soybean line (Burton et al. 1994). Close to *GmFATB1a* (1,048,878 bp), Zhang et al. (2018) found a QTL controlling palmitic content associated with the candidate gene *Glyma.05g012300*, another *GmFATB* gene. Besides these, Heim and Gillman (2017) mapped a QTL associated with palmitic acid in position 1,420,686 bp (R^2 = 13.06%). In another region of *qPal-05* interval is located *Glyma.05g015400* (7,703,169–7,710,844 bp) that codifies for phospholipid-translocating ATPase. This gene was identified by sequence comparison of known gene families in *Arabidopsis* as associated with palmitic acid content with R^2 = 7.7% (Li et al. 2015). All these findings indicate to the importance of this region of Gm05 in the control of palmitic acid.

We could not localize the position of one of the flanking markers of *qPal-13.1* in Soybase (BARC-041141-07916) but the other flanking marker (BARC-041671-08065) is placed at 30,268,680 bp. This QTL was significant in two environments and explained

Table 4 QTLs associated with oil and fatty acid average contents in F_2-derived populations from CS303TNKCA × FA22 crossing, evaluated in four environments in Brazil

Trait	QTL name	Chr	LG	SNP_name	Position (bp)	I (cM)	LOD	m	a	d	R^2	R^2-VRB	R^2-VIC	R^2-SG1	R^2-SG2
Palmitic	*qPal-02*	Gm02	D1b	BARC-050661-09809	4,189,847	0.80	6.48	102.17	2.14	0.03	12.56%	6.24%	11.30%	9.65%	9.70%
				BARC-020481-04638	4,032,973						LOD	3.00	5.75	4.82	4.85
Oleic	*qOle-02*	Gm02	D1b	BARC-019149-03315	7,473,425	2.60	4.43	332.95	− 18.52	2.65	8.93%	2.30%	9.45%	6.85%	5.55%
				BARC-037223-06748	6,976,258						LOD	1.06	4.71	3.32	2.65
Linoleic	*qLol-02*	Gm02	D1b	BARC-019149-03315	7,473,425	2.60	3.9	478.48	15.37	− 2	7.95%	1.80%	9.12%	5.97%	5.23%
				BARC-037223-06748	6,976,649						LOD	0.83	4.53	2.87	2.49
Oil	*qOil-02*	Gm02	D1b	BARC-048815-10726	14,106,445	16.36	3.92	214.12	− 1.95	− 0.28	7.98%	5.92%	2.15%	3.95%	6.50%
				BARC-030665-06919	10,233,770						LOD	2.84	0.99	1.86	3.14
Oleic	*qOle-04*	Gm04	C1	BARC-031733-07217	4,054,945	17.24	6.61	341.96	− 24.77	− 17.62	12.78%	11.23%	6.63%	7.03%	11.90%
				BARC-044691-08761	6,322,237						LOD	5.71	3.21	3.41	6.10
Linoleic	*qLol-04*	Gm04	C1	BARC-060869-16937	23,716,593	1.00	6.84	478.35	21.07	− 0.12	13.15%	15.60%	11.11%	6.24%	8.01%
				BARC-058991-15550	36,725,586						LOD	8.35	5.64	3.00	3.93
Oil	*qOil-04*	Gm04	C1	BARC-030765-06943	1,891,503	1.00	3.57	214.18	1.81	− 0.47	7.32%	6.41%	0.40%	4.72%	5.05%
				BARC-039239-07481	2,507,409						LOD	3.09	0.18	2.24	2.4
Palmitic	*qPal-05*	Gm05	A1	BARC-053497-11882	7,943,521	13.51	6.65	101.88	2.4	1.09	12.83%	14.67%	8.59%	9.12%	6.04%
	sd			BARC-064775-18801	432,530						LOD	7.76	4.24	4.53	2.90
Stearic	*qSte-06*	Gm06	C2	BARC-023517-05442	27,940,205	0.20	8.92	35.28	− 1.5	− 0.1	16.50%	7.37%	6.14%	20.13%	10.66%
				BARC-051071-10973	21,977,088						LOD	3.59	2.95	11.38	5.39

Table 4 continued

Trait	QTL name	Chr	LG	SNP_name	Position (bp)	I (cM)	LOD	m	a	d	R^2	R^2-VRB	R^2-VIC	R^2-SG1	R^2-SG2
Palmitic	*qPal-13.1*	Gm13	F	BARC-041671-08065	30,268,680	0.50	3.56	102.33	1.75	− 0.1	7.31%	4.46%	2.96%	6.83%	6.80%
				BARC-041141-07916	**						LOD	2.11	1.38	3.31	3.29
Palmitic	*qPal-13.2*	Gm13	F	BARC-031169-07006	1,456,933	6.84	3.75	102.25	1.82	0.02	7.67%	7.84%	5.13%	4.17%	5.19%
				BARC-065403-19439	2,776,574						LOD	3.84	2.44	1.96	2.47
Stearic	*qSte-13*	Gm13	F	BARC-030359-06859	32,171,107	7.56	3.41	35.57	0.82	− 0.73	7.02%	8.14%	2.57%	2.73%	5.33%
				BARC-013633-01184	30,771,524						LOD	4	1.19	1.27	2.54
Stearic	*qSte-14*	Gm14	B2	BARC-052789-11619	44,292,942	4.41	23.17	35.3	2.03	− 0.23	33.91%	43.94%	13.04%	16.50%	17.48%
				BARC-062647-17964	42,946,740						LOD	35.39	6.77	8.92	9.57
Linolenic	*qLin-14*	Gm14	B2	BARC-013273-00464	46,713,845	18.73	107.27	52.9	− 10.13	0.87	70.37%	71.71%	52.22%	54.20%	46.07%
				BARC-052789-11619	44,292,942						LOD	114.46	49.36	53.45	38.58
Stearic	*qSte-15*	Gm15	E	BARC-027480-06591	8,262,792	4.41	3.9	34.81	− 0.86	0.77	7.95%	4.89%	2.72%	6.89%	5.64%
				BARC-052309-11428	9,125,708						LOD	2.32	1.26	3.34	2.70
Palmitic	*qPal-17*	Gm17	D2	BARC-064695-18779	8,477,484	0.20	3.42	102.48	1.78	− 0.11	7.03%	14.92%	6.59%	2.52%	1.82%
				BARC-062955-18179	8,585,017						LOD	7.92	3.19	1.17	0.84
Stearic	*qSte-17*	Gm17	D2	BARC-051885-11289	38,811,019	7.35	3.85	34.18	− 1.29	0.02	7.86%	6.01%	3.26%	7.26%	6.38%
				BARC-049255-10878	39,399,316						LOD	1.55	1.52	3.54	3.07
Palmitic	*qPal-19*	Gm19	L	BARC-026069-05243	44,658,492	4.41	5.27	102.45	2.27	− 0.16	10.86%	1.21%	8.35%	13.64%	11.57%
				BARC-055739-13676	42,048,082						LOD	0.55	4.11	7.13	5.91

Table 4 continued

Trait	QTL name	Chr	LG	SNP_name	Position (bp)	I (cM)	LOD	m	a	d	R^2	R^2-VRB	R^2-VIC	R^2-SG1	R^2-SG2
Oleic	*qOle-19*	Gm19	L	BARC-026069-05243	44,658,492	4.41	6.23	327.01	− 25.07	9.04	12.13%	0.27%	6.16%	17.14%	15.05%
				BARC-055739-13676	42,048,082						LOD	0.12	2.96	9.34	8.00
Linoleic	*qLol-19*	Gm19	L	BARC-026069-05243	44,658,492	4.41	6.58	484.64	22.57	− 9.34	12.72%	0.00%	5.93%	18.28%	15.99%
				BARC-055739-13676	42,048,082						LOD	0.01	2.85	10.10	8.60

The position of marker was not available in Soybase. *Chr* Soybean chromosome; *LG* Soybean linkage groups; *I* Interval between markers; log-of-odds; *AA* Allele with higher average (C:CS303TNKCA; F: FA22) *m* average; *a* additive effect; *d* dominance effect; R^2: coefficient of determination; *VRB* Visconde do Rio Branco-MG; *VIC* Viçosa-MG; *SG* São Gotardo-M

7.31% of the palmitic content variation. Close to it, Priolli et al. (2015) associated marker Satt586 (29,609,521 bp) to palmitic acid content by evaluating 95 soybean accession. This marker could reduce palmitic content from 10.74% to 10.12%. In interval 8,477,484–8,585,017 bp of Gm17, *qPal-17* was found significant in two environments (R^2 = 7.03%). Reinprecht et al. (2006) found a QTL associated with Sat_092 marker (9,687,613 bp) that explained 11% of the palmitic content variation of a RIL population in one of three environments. Similarly, Smallwood et al. (2017) found a QTL close to this region (7,525,398 bp) that explained 3% of the palmitic content variation in a RIL population. The same authors found another QTL located at 44,515,446 bp position of Gm19 (R^2 = 4.0%), which matches to the interval of *qPal-19* (42,048,082–44,658,492 bp), a QTL significant in three environments which explained 10.86% of palmitic content variation. Besides theses QTLs, we could not associate *qPal-02* (4,032,973–4,189,847 bp; R^2 = 12.56%) and *qPal-13.2* (1,456,486–2,776,574 bp; R^2 = 7.67%) with QTLs or genes found in other previous studies, so these could be new QTLs controlling palmitic acid.

We identified five QTLs associated with stearic acid content that explained 7.02–33.91% of phenotypic variation, totaling 73.24% of phenotypic variation. Alleles from CS303TNKCA cultivar were responsible for reducing stearic content, except for *qSte-14*. The major QTL found was *qSte-14* (R^2 = 33.91%), which was significant in all environments and was located in the interval of 42,946,740–44,292,942 bp. In this interval, there are two serine-threonine protein kinases, *Glyma14g34560* (43,135,438–43,135,773 bp) and *Glyma14g34570* (43,138,158–43,139,368 bp). Li et al. (2015) considered another serine-threonine protein kinase gene located in Gm15 as a candidate gene for controlling stearic acid content in soybean seed. Thus, *Glyma14g34560* and *Glyma14g34570* are possible candidates to explain the effect of this QTL on stearic content. In addition, Heim and Gillman (2017) found a QTL in position 42,206,409 bp that explained 56.76% of stearic acid content in a RIL population, which increases the hypothesis of the importance of this region of Gm14 in controlling stearic acid content.

In Gm06, we found *qSte-06* (R^2 = 16.50%), located in the interval 21,977,088–27,940,205 bp and significant in all environments. In Soybase, BARC-023517-

05442 marker (27,940,205 pb) is placed at 112.87 cM, close to the estimated position of a QTL found by Hyten et al. (2004; 117.8–125.3 cM), which explained 13.1% of the stearic content variation in a RIL population. In Gm13, we found a QTL controlling stearic content (*qSte-13*; R^2 = 7.02%) that is located at the interval 30,771,524–32,171,107 bp. This region is close to other QTL found controlling palmitic content (*qPal-13.1*), indicating the possibility that the same region is controlling both traits. Two other QTLs were found controlling stearic acid content, *qSte-15* (8,262,792–9,125,708 bp; R^2 = 7.95%), and *qSte-17* (38,811,019–39,399,316 bp; R^2 = 7.86%), both significant in two environments, but no QTLs or genes associated with stearic content were found in these regions in the literature. As discussed above, Li et al. (2015) also found a QTL controlling stearic acid content in Gm15, but in another region of the chromosome. Therefore, *qSte-15* and *qSte-17* can be new QTLs associated with stearic content.

Three QTLs controlling oleic acid content were identified, explaining 8.93–12.78% of trait variation, and all alleles that increase oleic content were from FA22 line. Identified in two environments, *qOle-19* (R^2 = 12.13%) was located between 42,048,082 bp and 44,658,492 bp positions. Close to this interval, Kim et al. (2010) found a QTL associated with sct_010 marker (41,380,304 bp) with R^2 = 7.48% by evaluating RIL population in various environments. In the same region, Monteros et al. (2008) found the association between Satt561 marker (43,022,543 bp) to oleic content, explaining 25.0% of trait variation. In Gm04, *qOle-04* presented the highest coefficient of determination QTLs (12.78%) and was placed between 4,054,945 bp and 6,322,237 bp, being significant in all environments. In Gm 02, *qOle-02* (6,976,528–7,473,425; R^2 = 8.93%) was also significant in two environments. We could not identify other QTLs or genes associated to oleic content in these two regions in previous studies, so it can be new regions controlling oleic content.

Despite the wide amplitude in the content of oleic acid in all environments (184.01–577.42 g kg^{-1}), we did not find a major QTL controlling this trait (the QTLs found explained 33.84% of phenotypic variation). Thus, we could not completely elucidate the genetic bases of the middle oleic content of FA22 line (about 50%). Alternatively, we sequenced the entire region of *GmFAD2-1A* and *GmFAD2-1B* genes from FA22 line but found no mutations in them (data not shown). These two genes code for the main omega-6-desaturases expressed in soybean seed, and mutations in these genes are generally the cause of the increase of oleic content in soybean lines (Alt et al. 2005; Pham et al. 2010; Pham et al. 2011). We also evaluated the expression of *GmFAD2-1A* and *GmFAD2-1B* genes in FA22 line by Real-time PCR, but its expression pattern was similar to other soybean lines with normal values of oleic acid (data not shown). These results indicate that the middle oleic content of FA22 line could be explained by other mechanisms such as post-translational mechanisms, besides the QTLs discussed above.

Regarding the linoleic acid content, we found *qLol-04* in all environments, in the interval 23,716,593–36,725,586 bp (R^2 = 13.15%), *qLol-19*, significant in two environments and placed between 42,048,082 bp and 44,658,492 bp positions (R^2 = 12.72%) and *qLol-02* (R^2 = 7.95%), found in one environment. The QTLs *qLol-02* and *qLol-19* were located at the same intervals as *qOle-02*, and *qOle-19*, respectively, and had similar coefficients of determination, suggesting that there is a same locus controlling these two traits. In fact, highly negative correlations between oleic and linoleic acid content were reported in all environments as well as in other studies, suggesting a pleiotropic effect for these loci, as discussed before (Leamy et al. 2017; Li et al. 2017; Smallwood et al. 2017). Heim and Gillman (2017) and Kim et al. (2010) also found QTLs controlling linoleic content in the same regions, as discussed above. Moreover, the interval of *qOle-19* and *qLol-19* was also associated with palmitic acid (*qPal-19*). Pleiotropic effect for oleic, linolenic and palmitic contents were also reported by Smallwood et al. (2017). As expected, as FA22 line alleles were responsible for increasing oleic content, the same alleles reduced linoleic content.

Only one QTL was found to be associated with linolenic content, *qLin-14*, located in the interval 44,292,942–46,713,845 bp, explaining 70.37% of phenotypic variation. This region houses *Glyma.14g194300* (*GmFAD3A*), a major effect gene which codes for an omega-3-fatty-acid desaturase, an enzyme that converts linoleic to linolenic acid (Bilyeu et al. 2005; Pinto et al. 2013). We previously reported a mutation in *GmFAD3A* of CS303TNKCA variety that causes a premature stop codon, reduces

linolenic content to 33–44 g kg^{-1} and explained 50.83–73.70% of phenotypic variation (Silva et al. 2018a). Mutations in *GmFAD3A* associated with low linolenic acid were also reported by many authors (Bilyeu et al. 2005; Reinprecht et al. 2009; Pham et al. 2014).

In relation to oil content, we found two QTLs, both also significant in one environment. The QTL *qOil-02* (10,233,770–14,106,445 bp) explained 7.98% of oil variation. We could not find genes or QTLs associated to oil content close to this interval but a QTL associated with protein content in RIL population was found in the interval 14,764,422–15,058,707 bp Seo et al. (2019). Considering that negative correlation between oil and protein in soybean seed due to pleiotropy have been well reported (Reinprecht et al. 2006; Leite et al. 2016), a QTL controlling protein content is a target to controlling oil content as well. In Gm04 we found *qOil-04* (R^2 = 7.32%) in the interval 1,891,503–2,507,409 bp. Close to this region, the Satt274 marker (5,286,777 bp) was associated to oil content in two RIL populations (R^2 = 10.1–13.1) by Wang et al. (2014).

Other QTLs controlling oil content were found in individual analysis, but at most in two locations (data not shown). These results are in according to the low correlations between oil content in different environments (0.3397, on average). In fact, oil content is a complex trait that is controlled by both genetic and environmental factors (Priolli et al. 2015; Seo et al. 2019), and genotype x environment interaction for oil content were reported by many authors, including in Brazilian conditions (Bueno et al. 2013; Cao et al. 2017; Matei et al. 2018). Some studies have been reported that the oil content increases as the average temperature increases (Naeve and Huerd 2008; Matei et al. 2018). Rocha et al. (2009) evaluated the effect of many factors in soybean oil content, and found a pronounced effect of the interaction between genotype and year, overcoming the effects of planting location and sowing date. It is possible that different genes controlling oil content can be activated in different environment conditions, but more studies are necessary.

Oil and fatty acid contents are important traits for soybean to both industrial and food uses. In this study, we evaluated 211 segregating F_2-derived families in four environments and identify QTLs controlling oil and fatty acid contents. This population was derived from CS303TNKCA, a soybean cultivar developed by PMQS/BIOAGRO/UFV which has low linolenic acid and middle oil content and is used in our breeding program as allele donor for producing high oleic low linolenic acid lines. Of the 20 QTLs found in this study, eight could not be associated with previously reported QTLs or genes (*qPal-02, qOle-02, qLol-02, qOle-04, qLol-04, qPal-13.2, qSte-15 and qSte-17*), and can be configured as new QTLs. The results indicate FA22 line as an allele donor for reducing palmitic and linoleic acids and increasing oleic acid, while CS303TNKCA cultivar was identified as a source of alleles for reducing linolenic acid, as previously published (Silva et al. 2018a), and for reducing stearic content.

Further studies are needed to confirm the effect of these QTLs and reduce its intervals, such as fine QTL mapping and DNA sequencing, making it possible to find candidate genes that are related to fatty acid contents, increasing the knowledge about their genetic basis, including to identify new markers for marker-assisted selection. The results generated in this study can help understand the genetic basis of these traits and help breeders to modify oil content and composition in soybean seed, especially in tropical conditions.

Acknowledgements The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico for granting Doctoral Scholarships and Fundação de Amparo à Pesquisa do Estado de Minas Gerais for financial support (Grant APQ-01416-16 and PPM-640-18 to MD).

References

Akond M, Yuan J, Liu S, Kantartzi SK et al (2018) Detection of QTL underlying seed quality components in soybean [*Glycine max* (L.) Merr.]. Can J Plant Sci 98:881–888

Alt JL, Fehr WR, Welke GA, Shannon JG (2005) Transgressive segregation for oleate content in three soybean populations. Crop Sci 45:2005–2007

ANP – Agência Nacional de Petróleo (2019) Matérias-primas utilizadas na produção de biodiesel (B100) no Brasil – 2009-2018. Tabela 4.13. http://www.anp.gov.br/publicacoes/anuario-estatistico/5237-anuario-estatistico-2019

Bachlava E, Burton JW, Brownie C, Wang S, Auclair J, Cardinal AJ (2008) Heritability of oleic acid content in soybean seed oil and its genetic correlation with fatty acid and agronomic traits. Crop Sci 48:1764–1772

Bachlava E, Dewey RE, Burton JW, Cardinal AJ (2009) Mapping and comparison of quantitative trait loci for oleic acid

seed content in two segregating soybean populations. Crop Sci 49:433–442
Bachleda N, Pham A, Li Z (2016) Identifying FATB1a deletion that causes reduced palmitic acid content in soybean N87-2122-4 to develop a functional marker for marker-assisted selection. Mol Breed 36:45
Baud S, Lepiniec L (2010) Physiological and developmental regulation of seed oil production. Prog Lipid Res 49:235–249
Bilyeu K, Palavalli L, Sleper D, Beuselinck P (2005) Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. Crop Sci 45:1830–1836
Brazil (2016) Lei n° 13.263, de março de 2016
Bueno RD, Arruda KMA, Bhering LL, Gonccedil E et al (2013) Genetic parameters and genotype × environment interaction for productivity, oil and protein content in soybean. Afr J Agric Res 8:4853–4859
Bueno RD, God PIVG, Prata IO, Pereira PHS et al (2018) Association of candidate genes for fatty acid content in soybean by temperature-switch PCR (TSP) genotyping. Crop Breed Appl Biotechnol 18:244–251
Burkey KO, Booker FL, Pursley WA, Heagle AS (2007) Elevated carbon dioxide and ozone effects on peanut: II. Seed yield and quality. Crop Sci 47:1488–1497
Burton JW, Wilson RF, Brim CA (1983) Recurrent selection in soybeans. IV. Selection for increased oleic acid percentage in seed oil 1. Crop Sci 23:744–747
Burton JW, Wilson RF, Brim CA (1994) Registration of N79-2077-12 and N87-2122-4, two soybean germplasm lines with reduced palmitic acid in seed oil. Crop Sci 34:313
Cao Y, Li S, Wang Z, Chang F et al (2017) Identification of major quantitative trait loci for seed oil content in soybeans by combining linkage and genome-wide association mapping. Front Plant Sci 8:1222
Cardinal AJ, Burton JW, Camacho-Roger AM, Yang JH et al (2007) Molecular analysis of soybean lines with low palmitic acid content in the seed oil. Crop Sci 47:304–310
CONAB – Companhia Nacional De Abastecimento (2019) Acompanhamento Da Safra Brasileira de Grãos. Safra 2018/2019 5:11-20
Cruz CD (2013) Genes: a software package for analysis in experimental statistics and quantitative genetics. Acta Sci Agron 35:271–276
Fan S, Li B, Yu F, Han F et al (2015) Analysis of additive and epistatic quantitative trait loci underlying fatty acid concentrations in soybean seeds across multiple environments. Euphytica 206:689–700
Gillman JD, Stacey MG, Cui Y, Berg HR et al (2014) Deletions of the SACPD-C locus elevate seed stearic acid levels but also result in fatty acid and morphological alterations in nitrogen fixing nodules. BMC Plant Biol 14:143
Goettel W, Xia E, Upchurch R, Wang ML et al (2014) Identification and characterization of transcript polymorphisms in soybean lines varying in oil composition and content. BMC Genom 15:299
Goodstein DM, Shu S, Howson R, Neupane R et al (2011) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:1178–1186
Grant D, Nelson RT, Cannon SB, Shoemaker RC (2009) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res 38:843–846
Gutierrez-Gonzalez JJ, Vuong TD, Zhong R, Yu O et al (2011) Major locus and other novel additive and epistatic loci involved in modulation of isoflavone concentration in soybean seeds. Theor Appl Genet 123:1375–1385
Heim CB, Gillman JD (2017) Genotyping-by-sequencing-based investigation of the genetic architecture responsible for a sevenfold increase in soybean seed stearic acid. G3 Genes Genom Gene 7:299–308
Hu FB, Stampfer MJ, Manson JE, Rimm E et al (1997) Dietary fat intake and the risk of coronary heart disease in women. N Engl J Med 337:1491–1499
Hunter JE, Zhang J, Kris-Etherton PM (2009) Cardiovascular disease risk of dietary stearic acid compared with trans, other saturated, and unsaturated fatty acids: a systematic review. Am J Clin Nutr 91:46–63
Hyten DL, Pantalone VR, Saxton AM, Schmidt ME et al (2004) Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. J Am Oil Chem Soc 81:1115–1118
Hyten DL, Choi IY, Song Q, Specht JE et al (2010) A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. Crop Sci 50:960–968
Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30
Kim HK, Kim YC, Kim ST, Son BG et al (2010) Analysis of quantitative trait loci (QTLs) for seed size and fatty acid composition using recombinant inbred lines in soybean. J Life Sci 20:1186–1192
Kim SH, Ryu J, Kim WJ, Kang R et al (2019) Identification of a new *GmSACPD-C* allele in high stearic acid mutant Hfa180 derived from gamma-ray irradiation. Mol Breed 39:17
Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175
Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199
Leamy LJ, Zhang H, Li C, Chen CY et al (2017) A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*). BMC Genom 18:18
Leffel RC (1994) Registration of BARC-12, a low linolenic acid soybean germplasm line. Crop Sci 34:1426–1427
Leite DC, Pinheiro JB, Campos JB, Di Mauro AO et al (2016) QTL mapping of soybean oil content for marker-assisted selection in plant breeding program. Genet Mol Res 15:01–11
Li YH, Reif JC, Ma YS, Hong HL et al (2015) Targeted association mapping demonstrating the complex molecular genetics of fatty acid formation in soybean. BMC Genom 16:841
Li B, Fan S, Yu F, Chen Y et al (2017) High-resolution mapping of QTL for fatty acid composition in soybean using specific-locus amplified fragment sequencing. Theor Appl Genet 130:1467–1479
Lilliefors HW (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. J Am Stat Assoc Association 62:399–402

Maheshwari P, Kovalchuk I (2016) Genetic transformation of crops for oil production. In: Industrial oil crops, pp 379–412

Matei G, Meneguzzi C, Woyann LG, Todeschini MH et al (2018) Oil, protein and fatty acid profiles of Brazilian soybean cultivars in multi-environmental trials. Aust J Crop Sci 12:686–698

Monteros MJ, Burton JW, Boerma HR (2008) Molecular mapping and confirmation of QTLs associated with oleic acid content in N00-3350 soybean. Crop Sci 48:2223–2234

Naeve SL, Huerd SC (2008) Year, region, and temperature effects on the quality of Minnesota's soybean crop. Agron J 100:690–695

Nodari RO, Tsail SM, Gilbertson RL, Gepts P (1993) Towards an integrated linkage map of common bean 2. Development of an RFLP-based linkage map. Theor Appl Genet 85:513–520

Patterson E, Wall R, Fitzgerald GF, Ross RP et al (2012) Health implications of high dietary omega-6 polyunsaturated fatty acids. J Nutr Metab 2012

Pham AT, Lee JD, Shannon JG, Bilyeu KD (2010) Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biol 10:195

Pham AT, Lee JD, Shannon JG, Bilyeu KD (2011) A novel FAD2-1 A allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. Theor Appl Genet 123:793–802

Pham AT, Shannon JG, Bilyeu KD (2012) Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. Theor Appl Genet 125:503–515

Pham AT, Bilyeu K, Chen P, Boerma HR et al (2014) Characterization of the fan1 locus in soybean line A5 and development of molecular assays for high-throughput genotyping of FAD3 genes. Mol Breed 33:895–907

Pinto MDO, Good-God PIV, Moreira MA, Barros EGD (2013) Associação de marcadores moleculares SNP com o conteúdo de ácido linolênico em sementes de soja. Pesqui Agropecu Bras 48:263–269

Priolli RHG, Campos JB, Stabellini NS, Pinheiro JB et al (2015) Association mapping of oil content and fatty acid components in soybean. Euphytica 203:83–96

Priolli RHG, Carvalho CRL, Bajay MM, Pinheiro JB et al (2019) Genome analysis to identify SNPs associated with oil content and fatty acid components in soybean. Euphytica 215:54

R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

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

Reinprecht Y, Luk-Labey SY, Larsen J, Poysa VW et al (2009) Molecular basis of the low linolenic acid trait in soybean EMS mutant line RG10. Plant Breed 128:253–258

Resende MDVD (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. Crop Breed Appl Biotechnol 16:330–339

Rocha MDM, Vello NA, Lopes ADA, Maia MCC (2009) Estabilidade e adaptabilidade produtiva em linhagens de soja de ciclo médio. Rev Ceres 56:764–771

Seo JH, Kim KS, Ko JM, Choi MS et al (2019) Quantitative trait locus analysis for soybean (*Glycine max*) seed protein and oil concentrations using selected breeding populations. Plant Breed 138:95–104

Shu YJ, Yu DS, Wang D, Bai X et al (2013) Genomic selection of seed weight based on low-density SCAR markers in soybean. Genet Mol Res 12:2178–2188

Silva LC, Souza TLPO, Cruz CD, Carneiro PCS et al (2018a) Linkage fine-mapping and QTLs affecting morpho-agronomic traits of a Mesoamerican × Andean RIL common bean population. Euphytica 214:221

Silva LCC, Bueno RD, da Matta LB, Pereira PHS et al (2018b) Characterization of a new *GmFAD3A* allele in Brazilian CS303TNKCA soybean cultivar. Theor Appl Genet 131:1099–1110

Silva LCC, Mota LM, Fonseca LABV, Bueno RD et al (2019) Effect of a mutation in Raffinose Synthase 2 (*GmRS2*) on soybean quality traits. Crop Breed Appl Biotechnol 19:62–69

Smallwood CJ, Gillman JD, Saxton AM, Bhandari HS et al (2017) Identifying and exploring significant genomic regions associated with soybean yield, seed fatty acids, protein and oil. J Crop Sci Biotech 20:243–253

Thapa R, Carrero-Colón M, Hudson KA (2016) New alleles of *FATB1A* to reduce palmitic acid levels in soybean. Crop Sci 56:1076–1080

USDA (2019) Oilseeds: World markets and trade. August 2019

Vogel PA, de Noyer BS, Park H, Nguyen H et al (2019) Expression of the Arabidopsis WRINKLED 1 transcription factor leads to higher accumulation of palmitate in soybean seed. Plant Biotechnol J 17:1369–1379

Vuong TD, Walker DR, Nguyen BT, Nguyen TT et al (2016) Molecular characterization of resistance to soybean rust (Phakopsora pachyrhizi Syd. & Syd.) in soybean cultivar DT 2000 (PI 635999). PLoS ONE 11:e0164493

Wang W, Huang S, Liu Y, Fang Z et al (2012a) Construction and analysis of a high-density genetic linkage map in cabbage (Brassica oleracea L. var. capitata). BMC Genom 13:523

Wang X, Jiang GL, Green M, Scott RA et al (2012b) Quantitative trait locus analysis of saturated fatty acids in a population of recombinant inbred lines of soybean. Mol Breed 30:1163–1179

Wang X, Jiang G-L, Green M, Scott RA et al (2014) Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. Mol Genet Genomics 289:935–949

Warner K, Fehr W (2008) Mid-oleic/ultra low linolenic acid soybean oil: a healthful new alternative to hydrogenated oil for frying. J Am Oil Chem Soc 85:945–951

Weselake RJ, Taylor DC, Rahman MH, Shah S et al (2009) Increasing the flow of carbon into seed oil. Biotechnol Adv 27:866–878

Woyann LG, Meira D, Zdziarski AD, Matei G et al (2019) Multiple-trait selection of soybean for biodiesel production in Brazil. Ind Crop Prod 140:111721

Yu M, Liu Z, Jiang S, Xu N et al (2018) QTL mapping and candidate gene mining for soybean seed weight per plant. Biotechnol Biotec Eq 32:908–914

Zhang J, Wang X, Lu Y, Bhusal SJ (2018) Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. Mol Plant 11:460–472

Zhao X, Jiang H, Feng L, Qu Y et al (2019) Genome-wide association and transcriptional studies reveal novel genes for unsaturated fatty acid synthesis in a panel of soybean accessions. BMC Genom 20:68