

Analysis of additive and epistatic quantitative trait loci underlying fatty acid concentrations in soybean seeds across multiple environments

Shengxu Fan · Bin Li · Fukuan Yu ·
Fenxia Han · Shurong Yan · Lianzheng Wang ·
Junming Sun

Received: 13 May 2014 / Accepted: 7 June 2015 / Published online: 20 June 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Soybean is one of the most important oilseed crops in the world. The soybean oil contains various fatty acids. Their concentrations determine the quality and nutritional value of soybean oil. On the other hand, quantitative trait loci (QTL) mapping for the concentrations of soybean predominant fatty acids could provide the genetic basis for soybean fatty acid composition. In this study, a soybean genetic linkage map was constructed based on 161 polymorphic SSR markers in recombinant inbred lines (RILs) derived from a cross of *cv.* Luheidou2 × Nanhuizao. By using this map, 35 additive QTLs underlying individual fatty acid concentrations were identified in single environment, while 17 additive QTLs were identified underlying specific fatty acids across multiple environments or underlying multiple fatty acids. Fifteen of the 52 loci were found to be novel loci, explaining 5–24 % of

phenotypic variation. Moreover, 25 epistatic QTLs were identified and explained a high phenotypic variation for the fatty acid concentrations, suggesting an essential role of epistatic effect for fatty acid concentrations. The identification of additive and epistatic QTLs suggested a complex network for soybean fatty acid concentrations, and will facilitate the understanding for fatty acid accumulation.

Keywords Soybean · Fatty acid · Quantitative trait locus (QTL) · QTL interaction

Introduction

Soybean [*Glycine max* (L.) Merri] is the major oilseed crop in the world (Yaklich and Vinyard 2004). Fatty acids are predominant components of soybean oil. Fatty acids consist of saturated fatty acids (palmitic acid and stearic acid) and unsaturated fatty acids (oleic acid, linoleic acid, and linolenic acid). Different soybean fatty acids participate in different physiological functions (Mostofsky et al. 2001). Therefore, different compositions of fatty acids are desired depending on the end uses of the soybean oil (Panthee et al. 2006). For instance, vegetable oil with high concentration of unsaturated fatty acids is preferred in human diet for health reason. For industrial applications, however, soybean oil with high concentration of saturated fatty acids is more suitable due to its stability

Shengxu Fan, Bin Li and Fukuan Yu have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10681-015-1491-3) contains supplementary material, which is available to authorized users.

S. Fan · B. Li · F. Yu · F. Han · S. Yan ·
L. Wang · J. Sun (✉)

The National Key Facility for Crop Gene Resources and Genetic Improvement, NFCRI, MOA Key Laboratory of Soybean Biology (Beijing), Institute of Crop Science, Chinese Academy of Agricultural Sciences, 12 Zhongguancun South Street, Beijing 100081, China
e-mail: sunjunming@caas.cn

to oxidation (Henderson 1991; Hu et al. 1997; Spencer et al. 2003). Unsaturated fatty acids could decrease detrimental cholesterol in blood and lower the risk of cardiovascular disease (Mensink and Katan 1992; Connor 2000). Nevertheless, the polyunsaturated fatty acids, particularly linolenic acid, are prone to oxidation by lipoxygenase isozymes (Hildebrand et al. 1993) and negatively affect the flavor and shelf-life of soybean products (Robinson et al. 1995). Therefore, the improvement of fatty acid composition and the increase of oxidative stability have become to major goals of soybean breeding program for decades (Ha et al. 2010; Oliva et al. 2006).

In recent years, QTL mapping for main fatty acid components and identifying molecular markers closely linked to specific fatty acids had been performed in soybean breeding programs. In these studies, simple sequence repeat marker (SSR) (Bachlava et al. 2009; Panthee et al. 2006), single nucleotide polymorphism marker (SNP) (Wang et al. 2012a, b; Xie et al. 2012), and other types of marker (Diers and Shoemaker 1992; Li et al. 2011; Reinprecht et al. 2006) were employed in soybean genetic linkage map construction with different types of mapping populations [e.g. recombinant inbred lines (RILs), doubled haploid lines (DHs), and backcross population (BC)]. Based on these genetic maps, preliminary mapping of QTLs for the fatty acid concentrations was conducted. To date, 165 QTLs for individual fatty acid components were recorded in USDA-ARS Soybean Genetics and Genomics Database (SoyBase, <http://soybase.org>) according to previous studies (Alrefai et al. 1995; Bachlava et al. 2009; Brummer et al. 1997; Diers and Shoemaker 1992; Hyten et al. 2004a, b; Kim et al. 2010; Li et al., 2011; Shibata et al. 2008; Spencer et al. 2003; Panthee et al. 2006; Reinprecht et al. 2006; Wang et al. 2012a, b; Xie et al. 2012). These QTLs are useful for MAS in breeding program to alter soybean fatty acid composition.

Although numerous QTLs associated with fatty acid components have been identified, low repeatability and stability, as well as inconvenience of integration among numerous QTLs are still problems for their application due to the significant differences among various genetic backgrounds and environments. Therefore, it is necessary to identify and validate stable QTLs for individual fatty acid components across multiple environments. In this study, RIL populations derived from a cross of *cv.* Luheidou

2 × Nanhuizao were planted in Beijing over 2009–2011, and were used to identify QTLs underlying individual fatty acid concentrations. Amounts of consistent additive QTLs were identified across multiple environments. Moreover, the epistatic interactions of QTLs underlying fatty acids were also analyzed. Taken together, this study provided some new knowledge on the genetic basis of soybean fatty acid composition.

Materials and methods

Plant materials and field design

The mapping populations were initially developed from a cross between *cv.* Luheidou2 (distributed in HuangHuaiHai valley region) and Nanhuizao (distributed in South region of China). The RIL population was advanced by using single-seed descent of F₂ lines up to F₅ and then establishing 200 F₅:7 and F₅:8 populations. From 2009 to 2011, the 200 RILs together with their parents were planted at Shunyi Experimental Stations (N40°13' and E116°34') in Beijing. The three environments of 2009, 2010 and 2011 were designated as E1, E2 and E3, respectively. In this study, the F₅:7 and F₅:8 population were planted in rows of 2 m long, 0.5 m apart and with a space of 0.1 m between two plants. Three replicates were conducted with a randomized complete block design.

Fatty acid extraction and determination

The procedure was followed as described by Kamal-Eldin and Andersson (1997). About 20 g soybean seeds of each RIL were ground to fine powder with a Sample Preparation Mills (Retsch ZM100, $\Phi = 1.0$ mm, Rheinische, Germany) and stored at -20 °C before used. Three hundred milligrams of each soybean powder was weighted with analytic balance (SartoriusBS124S) and transferred to a 2-ml centrifuge tube preloaded with 1.5 ml *n*-hexane. After mixing, the mixture was stored at 4 °C for 12 h. Then the samples were centrifuged at 5000×*g* for 10 min. The supernatant was collected into a new 2-ml centrifuge tube with 350 μ l sodium methoxide solution and shaken for 1 h on the twist mixer (TM-300, ASONE, Japan) for full methyl esterification. Then the

mixture was centrifuged at $5000 \times g$ for 10 min again and supernatant was collected to detect the fatty acid composition (Kamal-Eldin and Andersson 1997).

Fatty acid composition was determined using gas chromatography (GC-2010, SHIMADZU, Japan). Operation instrumental conditions were as follows (Ma et al. 2015): chromatographic column RTX-Wax ($30 \text{ m} \times 0.25 \times 0.25$); auto injection $1 \mu\text{l}$; split ratio 40:1; injection port temperature $250 \text{ }^\circ\text{C}$; the carrier gas nitrogen 54 ml min^{-1} ; hydrogen 40 ml min^{-1} ; air 400 ml min^{-1} ; take the temperature programmed mode ($180 \text{ }^\circ\text{C}$ keep for 1.5 min, up to $210 \text{ }^\circ\text{C}$ by $10 \text{ }^\circ\text{C min}^{-1}$ and keep 2 min, up to 220 by $5 \text{ }^\circ\text{C min}^{-1}$ and 5 min). Detector FID temperature was $300 \text{ }^\circ\text{C}$. The area normalization method was employed to calculate the percentage of five predominant fatty acids (palmitic acid, stearic acid, oleic acid, linoleic acid and linolenic acid) on GC2010 workstation (Shimadzu, Japan).

Polymorphic SSR marker selection and genotyping for RILs

One hundred F5:7 RILs were selected randomly from the 200 F5:7 RILs. For each line of the 100 RILs, the young leaves collected from approximately 10 seedlings were ground to powder in liquid nitrogen with mortar and pestle. Total genomic DNA of each line was extracted separately using the CTAB method (Doyle 1990). The resultant DNA was dissolved in $200 \mu\text{l}$ ddH₂O, and quantified with NanoDrop ND-100 spectrophotometer. Then, the genomic DNA of each sample was diluted to a concentration of $100 \text{ ng } \mu\text{l}^{-1}$, and used in the subsequent genotyping analysis.

In this study, a total of 530 pairs of SSR primers selected from the SoyBase (<http://soybase.org>) were synthesized in Shanghai ShengGong biological engineering technology service co., LTD. The polymorphism of these SSR markers was tested between two parents of RIL population. Towards the end, 161 polymorphic SSR markers were identified. The polymorphic markers were used in genotyping for 100 RILs. Polymerase chain reaction (PCR) was performed as follow: $94 \text{ }^\circ\text{C}$ for 3 min followed by 35 cycles at $94 \text{ }^\circ\text{C}$ for denaturation for 30 s; $47\text{--}55 \text{ }^\circ\text{C}$ for annealing for 30 s; $72 \text{ }^\circ\text{C}$ for 45 s for extension, and the last step at $72 \text{ }^\circ\text{C}$ for final extension for 10 min.

Construction of the genetic map and data analysis

The genotyping of 100 RILs was performed based on the band type of the polymorphic SSR markers. The maternal band type was marked with '2', the paternal band type was marked with '0', the hybrid band type was marked with '1' and the missing ones marked with '-1'.

Construction of the genetic map was performed using the MAP model in the software of QTL IciMapping v3.3 (Cui et al. 2011). According to concentrations of five main fatty acid components in soybean seeds, the QTLs were detected for palmitic acid, stearic acid, oleic acid, linoleic acid and linolenic acid using inclusive composite interval mapping (ICIM) method in BIP model of QTL IciMapping v3.3. The threshold of LOD scores for evaluating the statistical significance of QTL effects was determined using 1000 permutations. Based on these permutations, a LOD score of 3.1 was used as a minimum to declare the presence of a QTL in genomic region. The epistatic effect of QTL was analyzed by ICIM-EPI in BIP model of QTL IciMapping v3.3 (P value < 0.0005 ; LOD > 5.0). This method has been described in detail by Li et al. (2008), and applied to epistatic QTL mapping for amounts of essential traits (Ding et al. 2014; Wang et al. 2012a, b). The genes within QTLs were annotated and analyzed via the database of Phytozome v9.1 (www.phytozome.net) and NCBI (www.ncbi.nlm.nih.gov).

Analysis of variance (ANOVA) was performed to determine the significance of genotypic differences between the RILs and environments. Correlation analysis among five components of fatty acid in soybean seeds was conducted using CORR procedure of SAS version 9.2 and the frequency distribution was analyzed by Microsoft Excel 2013.

Results

Phenotypic analysis of fatty acid in soybean seeds

The determination of fatty acid composition was performed by gas chromatography analysis. As a result, five predominant fatty acid components including stearic acid, palmitic acid, oleic acid, linoleic acid, and linolenic acid were identified and quantified in

soybean seeds. The characters of fatty acid compositions of the RIL population were analyzed across multiple environments. As shown in Table 1, the two parents of RILs exhibited significant differences for almost all fatty acids except palmitic acid, implying different genetic backgrounds between them. Moreover, the RILs exhibited a broad range of variation in fatty acid concentrations. Of the five predominant fatty acid components, linoleic acid presented a minimum coefficient of variation ranging from 4.0 to 6.1 %, while stearic acid showed a maximum coefficient of variation ranging from 10.2 to 18.4 % (Table 1). According to the Kolmogorov–Smirnov test, the frequency distributions of five fatty acid components across various environments generally exhibited in a continuous and normal manner except the oleic acid distribution in E3 (Fig. S1 and Table S1), suggesting the fatty acid concentrations are typical quantitative

Fig. 1 Consistent additive QTLs associated with specific fatty acids across multiple environments or associated with multiple fatty acids. The *orange regions* indicate the location of these QTLs on LGs. The name of the QTL, shown near their location in *orange color*, is a composite of the influenced trait: palmitic acid (PA), stearic acid (SA), oleic acid (OA), linoleic acid (LA) and linolenic acid (LNA) followed by the chromosome number. For QTL underlying multiple fatty acid components, the name is a composition of fatty acid (FA) followed by the chromosome number. The SSR markers are shown on the right of LGs, and their positions on LGs are indicated on the *left* in centimorgan (cM). (Color figure online)

traits and suitable for QTL mapping. Noticeably, transgression segregations for individual fatty acid concentration were also observed in the RIL population (Fig. S1), which suggests that the favorable alleles for fatty acid concentrations were derived from both parents.

Table 1 The statistical analysis of fatty acid components in the RIL population

Fatty acid	E ^c	P1 ^d	P2 ^e	Max	Min	Mean	V ^f	Range	CV % ^g	S ^h	K ⁱ
Palmitic	2009	10.3	11.2	13.3	9.1	10.3	0.4	4.2	5.9	1.3	4.4
	2010			13.8	9.9	11.1	0.3	3.8	5.0	1.5	5.3
	2011			13.0	9.6	10.8	0.5	3.4	6.4	0.5	0.2
Stearic	2009	3.6 ^a	5.1 ^b	5.0	3.2	3.9	0.2	1.8	11.5	0.6	−0.4
	2010			4.9	3.1	3.9	0.2	1.8	10.2	0.3	0
	2011			7.1	2.7	3.7	0.5	4.4	18.4	1.5	5.0
Oleic	2009	25.4 ^a	30.8 ^b	29.9	20.3	23.9	5.4	9.7	9.8	0.4	−0.4
	2010			35.8	19.1	26.2	14.8	16.8	14.7	0.4	−0.4
	2011			30.9	18.4	22.3	6.2	12.5	11.2	1.1	1.3
Linoleic	2009	52.8 ^a	47.4 ^b	57.5	47.7	53.3	4.8	9.8	4.1	−0.2	−0.7
	2010			57.5	43.1	51.4	9.8	14.4	6.1	−0.4	−0.3
	2011			59.5	46.3	55.0	4.8	13.1	4.0	−1.0	2.0
Linolenic	2009	8.4 ^a	6.0 ^b	10.6	7.1	8.6	0.4	3.5	7.6	0.3	0.2
	2010			10.2	6.0	7.5	0.6	4.2	10.4	0.5	0.4
	2011			9.9	5.7	8.2	0.5	4.2	8.9	−0.4	0.3

The fatty acid concentrations are shown as the percentage of total fatty acids. The average values of individual fatty acid concentrations across 3 years were shown for P1 and P2

The letters a and b indicate the values differs significantly between two parents ($P < 0.05$)

^c E indicates environments used in this study

^d P1 indicates maternal parent, *cv.* Luheidou2

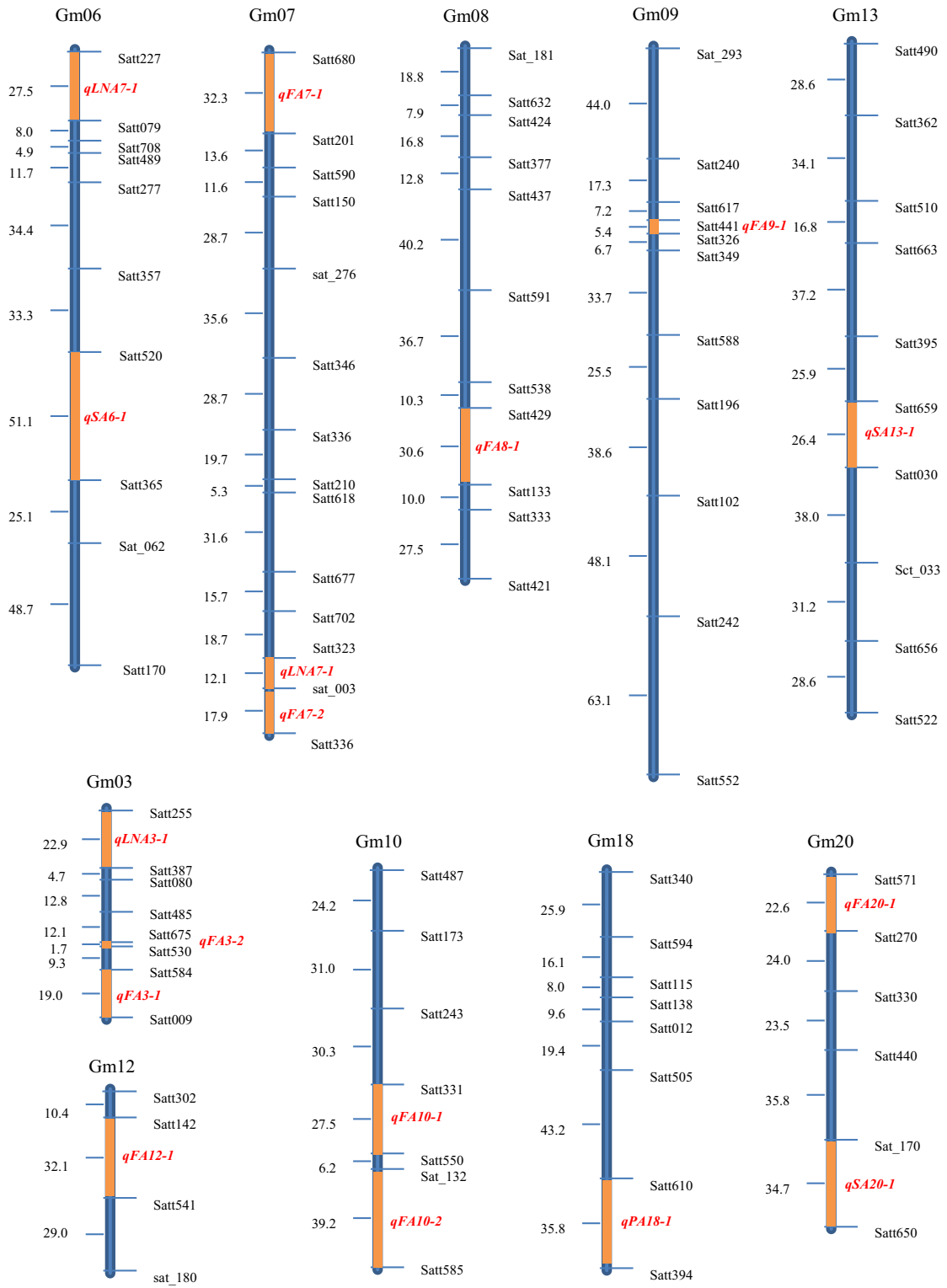
^e P2 indicates paternal parent, *cv.* Nanhuizao

^f V indicates the phenotypic variation

^g CV % indicates coefficient of variation within the segregation mapping population over environments

^h S indicates skewness

ⁱ K indicates kurtosis



10cM

Table 2 QTLs detected underlying specific fatty acids across various environments or underlying multiple fatty acids

QTL ^a	Effect ^b	Chr ^c	Interval ^d	LOD ^e	PVE ^f (%)	Add ^g
<i>qLNA3-1</i>	LNA-E1,E3	3	Satt387-Satt255	6.6	13.7	0.3
<i>qLNA6-1</i>	LNA-E1,E2,E3	6	Satt079-Satt227	8.4	22.8	-0.4
<i>qSA6-1</i>	SA-E2,E3	6	Satt365-Satt520	3.7	13.5	-0.2
<i>qLNA7-1</i>	LNA-E1,E2	7	Sat_003-Satt323	4.7	11.3	0.2
<i>qFA8-1</i>	LNA-E1,E2	8	Satt133-Satt429	4.7	17.6	-0.3
	SA-E2			16.0	40.0	0.3
<i>qSA13-1</i>	SA-E1,E2	13	Satt030-Satt659	5.8	11.9	-0.2
<i>qPA18-1</i>	PA-E1,E2	18	Satt394-Satt610	3.6	9.9	-0.2
<i>qSA20-1</i>	SA-E1,E2	20	Satt650-sat_170	9.3	23.6	0.2
<i>qFA3-1</i>	LNA-E1	3	Satt009-Satt584	3.5	8.7	-0.2
	SA-E3			11.4	28.3	0.4
<i>qFA3-2</i>	OA-E2	3	Satt530-Satt675	3.3	9.5	1.2
	LA-E2			3.6	11.5	-1.1
<i>qFA7-1</i>	SA-E2	7	Satt201-Satt680	10.9	20.8	-0.2
	LNA-E3			3.1	6.1	-0.2
<i>qFA7-2</i>	SA-E3	7	Satt336-sat_003	9.1	17.1	0.3
	LNA-E3			7.3	18.6	0.3
<i>qFA9-1</i>	LA-E1	9	Satt326-Satt441	3.9	9.4	0.7
	SA-E3			9.3	17.7	-0.3
<i>qFA10-1</i>	OA-E2	10	Satt550-Satt331	7.8	26.5	2.0
	LA-E2			8.8	32.3	-1.8
<i>qFA10-2</i>	OA-E2	10	Satt585-Sat_132	4.8	25.3	-1.9
	LA-E2			6.0	33.9	1.8
	SA-E3			15.6	35.5	-0.4
<i>qFA12-1</i>	LA-E1	12	Satt541-Satt142	7.7	18.0	-0.9
	SA-E3			3.2	5.3	-0.2
<i>qFA20-1</i>	PA-E2	20	Satt270-Satt571	3.1	9.1	0.2
	SA-E3			5.0	8.7	-0.2

The LOD scores, PVE and additive effect of QTLs underlying across various environments are presented as average values

^a The name of the QTL, is a composite of the influenced trait: stearic acid (SA), palmitic acid (PA), oleic acid (OA), linoleic acid (LA), linolenic acid (LNA), followed by the chromosome number. For QTLs underlying multiple fatty acids, the name is a composite of fatty acid (FA) followed by the chromosome number

^b The Effect of QTL indicates the concentrations of particular fatty acids (SA, PA, OA, LA, and LNA) in specific environments (E1, 2009; E2, 2010; E3, 2011)

^c Chr indicates chromosome

^d Interval indicates confidence intervals between two SSR markers

^e LOD indicates the logarithm of odds score

^f PVE indicates the phenotypic variance explained by individual QTL

^g Add indicates the additive effect value

Consistent with previous studies (Dornbos and Mullen 1992; Hou et al. 2006), the concentrations of almost all fatty acid components were significantly influenced by both genetic and environmental

factors (Table S2). Nevertheless, the significant correlations of fatty acid concentrations among various years (Table S3) suggested genetic effect still plays the most important role in soybean fatty

acid composition despite of the environmental effect.

QTL mapping of fatty acid components in soybean

Based on 161 polymorphic SSR markers, a soybean genetic linkage map was constructed. The genetic map spanned 3591.2 cM in genome sized with an average distance of 22.3 cM between adjacent markers. With this map, QTL mapping was conducted for the fatty acid concentrations. Finally, a total of 52 QTLs were identified on 19 linkage groups (LGs) except Gm04, which could explain 5–40 % of the phenotypic variation for the individual fatty acid concentrations in soybean seeds, with the LOD scores ranging from 3.1 to 16.0. Of these loci, 35 QTLs were detected in single environment, including nine QTLs for linolenic acid, eight for linoleic acid, one for oleic acid, six for stearic acid, and thirteen for palmitic acid (Table S4). The other 17 QTLs were mapped to 10 LGs

accounting for specific fatty acid components across multiple environments or underlying multiple fatty acid components (Fig. 1; Table 2). In consideration of the environmental effect on fatty acid composition, these 17 consistent QTLs may represent the major genetic basis for fatty acid composition, thereby were focused subsequently.

Of the 17 consistent QTLs, eight QTLs accounted for specific fatty acid components across various environments (Table 2). Specifically, four QTLs (*qLNA3-1*, *qLNA6-1*, *qLNA7-1*, and *qFA8-1*) contributed to linolenic acid concentration, accounting for 7–23 % of the phenotypic variation. The favorable alleles of *qLNA6-1* and *qFA8-1* were inherited from the paternal parent *cv.* Nanhuizao, whereas, the favorable alleles of *qLNA3-1* and *qLNA7-1* were derived from the maternal parent *cv.* Luheidou2. Noticeably, *qLNA6-1* could stably explain the a relative high phenotypic variation for linolenic acid concentration across three environments (with a mean

Fig. 2 The epistatic interactions of QTLs for individual fatty acids in soybean seeds among different environments. The figure was generated by ICIM-EPI in the QTL IciMapping v3.3 software using multi-environmental functionality. The 20 colors in the ring represent the soybean 20 chromosomes. The numbers in the ovals indicate the positions of markers on chromosomes. The different colors of dash lines connecting the loci represent the epistatic interaction between the two loci for different fatty acids in specific environment. The numbers on the dash lines indicate the LOD scores of the additive \times additive effects between two QTLs. (Color figure online)

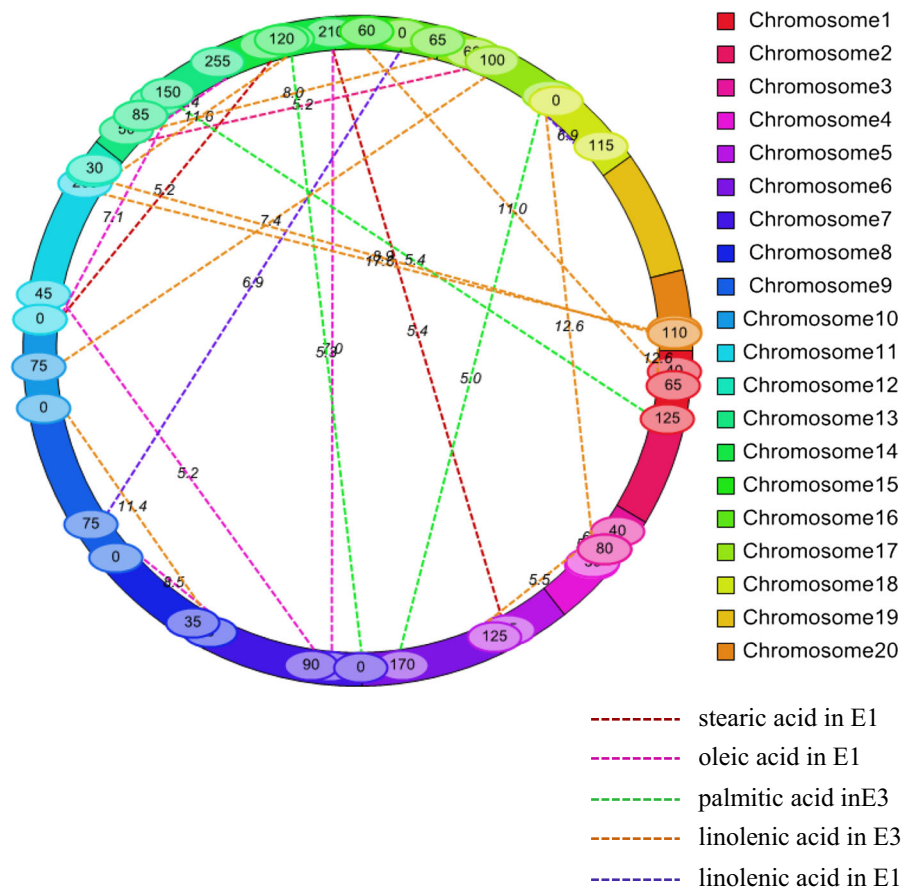


Table 3 Epistatic QTLs for the fatty acid components in soybean seeds among multiple environments

Effect ^a	Chr1 ^b	Interval 1 ^c	Chr2	Interval 2	LOD ^d	PVE ^e (%)	A × A ^f
LNA-E3	1	Satt267-Satt129	15	Satt651-Satt724	11.0	9.9	0.2
LNA-E3	1	Satt267-Satt129	20	Sat_170-Satt650	12.6	12.4	0.3
LNA-E3	3	Satt387-Satt255	5	Satt717-Satt545	5.5	5.0	-0.2
LNA-E3	3	Satt387-Satt255	18	Satt340-Satt594	12.6	10.4	0.2
LNA-E1	3	Satt675-Satt485	4	Satt361-Satt565	6.1	20.8	-0.3
LNA-E3	8	Satt591-Satt538	10	Satt585-Sat_132	11.4	10.9	--0.3
LNA-E1	9	Satt588-Satt240	16	Satt380-Sat_065	6.9	29.1	-0.4
LNA-E3	10	Satt331-Satt243	17	Satt256-Satt372	7.4	7.3	0.2
LNA-E3	11	Satt332-Satt583	20	Satt440-Sat_170	17.8	20.4	-0.3
LNA-E3	12	Sat_180-Satt541	20	Sat_170-Satt650	8.9	10.7	-0.3
LNA-E3	12	Satt541-Satt142	14	Satt070-Satt168	11.6	11.4	0.3
LNA-E3	13	Satt030-Sct_033	16	Sat_339-Satt431	8.0	10.9	-0.3
LNA-E1	18	Satt340-Satt594	18	Satt505-Satt610	6.9	28.2	0.4
OA-E1	7	Satt590-Satt346	14	Satt687-Satt726	7.0	19.8	-1.1
OA-E1	7	Satt590-Satt346	11	Satt251-Sat_247	5.2	12.7	-0.8
OA-E1	8	Satt591-Satt538	9	Satt552-Satt196	8.5	17.6	-1.0
OA-E1	11	Satt197-Satt251	13	Satt656-Satt522	7.1	17.8	-1.0
OA-E1	13	Satt030-Sct_033	13	Satt663-Satt510	5.4	13.4	0.9
PA-E3	1	Satt147-Sat_254	13	Satt522-Satt490	5.4	19.4	-0.3
PA-E3	6	Satt079-Satt708	17	Satt186-Satt615	5.0	19.4	0.3
PA-E3	7	Satt680-Satt201	14	Satt070-Satt168	5.3	25.3	0.3
SA-E3	3	Satt387-Satt255	4	Satt476-Satt361	5.3	23.1	-0.3
SA-E1	5	Satt276-Satt717	14	Satt687-Satt726	5.4	16.7	-0.2
SA-E1	11	Satt197-Satt251	14	Satt122-Satt070	5.2	12.5	0.2
SA-E1	13	Satt659-Satt030	17	Satt458-Satt256	5.2	22.4	0.2

^a The Effect of QTL indicates the concentrations of particular fatty acids: stearic acid (SA), palmitic acid (PA), oleic acid (OA), LA (linoleic acid), and linolenic acid (LNA) in specific environments (E1, 2009; E2, 2010; E3, 2011)

^b Chr indicates chromosome

^c Interval indicates confidence intervals between two SSR markers

^d LOD indicates the logarithm of odds score

^e PVE stands for phenotypic variation explained by epistatic QTL interaction

^f A × A indicates the epistatic effect of QTL interaction

of 23 %), with the average LOD score of 8.4, suggesting this QTL may be a major and stable locus for linolenic acid concentration. Three QTLs (*qSA6-1*, *qSA13-1* and *qSA20-1*) were associated with stearic acid concentration, accounting for 12 to 24 % of phenotypic variation. The favorable alleles of *qSA6-1*, *qSA13-1* were derived from *cv.* Nanhuizao, while the favorable alleles of *qSA20-1* were derived from *cv.* Luheidou2. In addition, Nanhuizao-derived *qPA18-1* explained 10 % of the phenotypic variation for palmitic acid concentration (Table 2).

On the other hand, of the 17 consistent QTLs, 10 were associated with multiple fatty acid components (Table 2). They could explain 5–40 % of phenotypic variation, with the LOD value from 3.1 to 16.0. Of these loci, *qFA3-1*, *qFA7-1*, *qFA7-2* and *qFA8-1* accounted for both linolenic and stearic acids. *qFA9-1* and *qFA12-1* were associated with both linoleic and stearic acids. *qFA10-1* and *qFA10-2* were detected for

both oleic and linoleic acids. *qFA3-2* contributed to oleic, linoleic, and stearic acids. Additionally, *qFA20-1* was detected for palmitic and stearic acids (Table 2).

Comparison of QTLs underlying fatty acid components

Subsequently, these QTLs were compared with previous studies. As shown in Table S4, of the all 52 QTLs, 37 were overlapped with the known QTLs for fatty acid concentration according to the Soybase Database. The other 15 QTLs, explaining 5–24 % of phenotypic variation for individual fatty acids, were novel loci (Table S4). For the 17 QTLs detected across multiple environments or associated with multiple fatty acids, 15 were reported previously, while two novel loci were observed (Table S4). *qSA20-1* contributed to stearic acid across multiple environments,

while *qFA12-1* was associated with both stearic and linoleic acid (Table 2).

In addition, with the completion of whole genome sequencing for soybean genome, genomic comparative analysis for QTL intervals could also be conducted. Therefore, the genomic information within the 52 QTL intervals was obtained. Toward the end, five QTL intervals (i.e. *qLNA3-1*, *qFA7-2*, *qLA2-1*, *qLA11-1* and *qLA14-1*) were found containing genes involved in fatty acid biosynthesis or accumulation (Table S4). These genes (*GmFAD2-2*, *GmFAD3-2b*, *GmSAD1*, *GmSAD2*, and *GmSACPD-C*) are responsible for the introduction of double bonds into fatty acyl chains, following the removal of two hydrogen atoms, and thereby are essential for biosynthesis of unsaturated fatty acid (Ohlrogge and Browse 1995; Fofana et al. 2004). The inclusion of these genes within QTL intervals suggested a relatively high accuracy of QTL mapping in this study.

Epistatic QTLs associated with fatty acid components

Finally, the epistatic interactions for fatty acid components were also analyzed. As a result, 25 pairwise QTLs with epistatic interactions were identified for five fatty acids among three environments (Fig. 2). These epistatic QTLs were mapped to almost all LGs except Gm02 and Gm19, and explained phenotypic variation between 7 and 29 % (Table 3). Nearly half of these epistatic QTLs were involved in linolenic acid, while that associated with saturated fatty acids were less. The high phenotypic variation explained by epistatic QTLs suggested the fatty acids, especially linolenic acid, were significantly influenced by epistatic interactions of QTLs. Unfortunately, the stable QTL \times QTL interaction was not detected across multiple environments in this study. This result suggested the fatty acids were also affected by epistatic QTLs and the epistatic interactions were considerably complex.

Discussions

Soybean is the leading oilseed crop in the world. The fatty acid composition determines the quality and nutritional value of soybean oil. Different soybean fatty acid compositions are desired depending on the

end uses of the soybean oil. However the soybean fatty acid composition was regulated by a complex network including genetic and environmental factors.

Fatty acids were influenced by both genetic and environmental factors

Soybean fatty acid concentrations are indicated as quantitatively inherited characters. They are subjected to an oscillation even in the same population under different environments (Hou et al. 2006). Therefore, multiple environments should be considered in QTL mapping. On the other hand, a high heritability in the broad-sense was observed for fatty acid components despite of the environmental influence (Gesteira et al. 2003), indicating their phenotypic variations are mainly controlled by genetic factor. In this study, QTL mapping was conducted in a RIL population over three environments. As expected, QTLs underlying specific fatty acids varied with different environments. However, 17 consistent additive QTLs were identified underlying specific fatty acids or underlying multiple fatty acids. These loci may represent the genetic basis of fatty acid biosynthesis.

The comparative analysis suggested stable and novel QTLs for fatty acid concentrations

The comparative analysis of QTLs suggested a consistency between our results and previous studies. A great part of QTLs identified in this study have been reported previously, suggesting a reliability and accuracy of our study to some extent. As to the 17 consistent QTLs underlying specific fatty acids across multiple environments or underlying multiple fatty acids, 15 of them have been reported. These loci are stable for fatty acid concentrations among multiple mapping populations, thereby may represent the common genetic basis for fatty acid composition. On the other hand, the 15 novel QTLs, especially the two novel consistent QTLs (*qSA20-1* and *qFA12-1*), provided new QTL information for fatty acid composition, therefore, should be emphasized in subsequent study.

By comparing the genomic regions of QTLs with soybean reference genomic sequences, five essential genes accounting for biosynthesis of unsaturated fatty acids were discovered within five QTL intervals. This result implied that these genes may represent the

genetic basis of related QTLs, although lots of work should be done in further study. In addition, a bZIP-transcription-factor-like gene (*Glyma20g01030.1*) was discovered within the genome interval of *qSA20-1*, which explained a great part of phenotypic variation for stearic acid concentration. It is reported the bZIP transcription factor (bZIP67) could regulate the Omega-3 fatty acid concentration by activating *FAD3* (Mendes et al. 2013). Therefore, this gene should be paid more attention in our further study.

The fatty acid concentrations were influenced significantly by epistatic interaction of QTLs

Genetic variation may be attributed by epistatic interaction of QTLs. Significant epistatic effect have been reported in soybean for yield (Lark et al. 1995), plant height (Orf et al. 1999), isoflavone contents (Gutierrez-Gonzalez et al. 2011) and fatty acid components (Li et al. 2011). In this study, the epistatic QTLs were also detected on almost all LGs, and explained a high phenotypic variation. Interestingly, half of these epistatic QTLs contributed to linolenic acid. This result, combined with previous studies, demonstrated fatty acids, especially linolenic acid, were significantly influenced by QTL \times QTL interaction. However, most of these epistatic interactions of QTLs were detected in single environment. Stable epistatic interaction across multiple environments was not detected in this study. That may be due to the limited population size and marker number in this study. Moreover, epistatic QTL \times environment interaction for fatty acids should also be considered in the subsequent study.

Taken together, the epistatic interaction for fatty acid components, combined with additive QTLs, suggested a complex network controlling fatty acid composition. That made it difficult to apply MAS for improving fatty acid composition in soybean breeding immediately. However, our result provided the solid evidence for the involvement of epistatic QTLs in fatty acid accumulation.

The limitation of efficiency and accuracy for QTL detection

The efficiency of QTL identification is determined by numerous factors such as the population size, the density of polymorphic markers, the mapping

algorithm, etc. (Darvasi et al. 1993; Li et al. 2010; Stange et al. 2013; Yu et al. 2011). In this study, although 100 RILs are sufficient to detect QTL explaining a relatively high phenotypic variation (PVE >10 %) with the existing marker density (Li et al. 2010), the QTLs with minor effect could be missed. For instance, only 17 consistent QTLs were identified for five predominant fatty acids in this study. Moreover, the fine mapping of these QTLs or MAS for fatty acid composition is also infeasible due to the large distance within QTL interval. Therefore, a large population size and high density of polymorphic markers were required for further study to detect more consistent QTLs and narrow down the QTL confident intervals.

Acknowledgments This work was supported by the National Science and Technology Pillar Program of China (2014BAD11B01-X02), the Genetically Modified Organisms Breeding Major Projects (No. 2014ZX08004-003), the National Nature Science Foundation of China (No. 31171576 and No. 31301345), and the CAAS Innovation Project.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alrefai R, Berke TG, Rocheford TR (1995) Quantitative trait locus analysis of fatty acid concentrations in maize. *Genome* 38(5):894–901
- 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(2):433–444
- Brummer EC, Graef GL, Orf J, Wilcox JR, Shoemaker RC (1997) Mapping QTL for seed protein and oil content in eight soybean populations. *Crop Sci* 37(2):370–378
- Connor WE (2000) Importance of n–3 fatty acids in health and disease. *Am J Clin Nutr* 71(1):171S–175S
- Cui F, Li J, Ding A, Zhao C, Wang L, Wang X, Li S, Bao Y, Li X, Feng D (2011) Conditional QTL mapping for plant height with respect to the length of the spike and internode in two mapping populations of wheat. *Theor Appl Genet* 122(8):1517–1536
- Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134(3):943–951

- Diers BW, Shoemaker RC (1992) Restriction fragment length polymorphism analysis of soybean fatty acid content. *J Am Oil Chem Soc* 69(12):1242–1247
- Ding G, Zhao Z, Lin W, Zhang D, Shi L, Xu F (2014) Identification and multiple comparisons of QTL and epistatic interaction conferring high yield under boron and phosphorus deprivation in *Brassica napus*. *Euphytica* 198(3):337–351
- Dornbos DL Jr, Mullen RE (1992) Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. *J Am Oil Chem Soc* 69(3):228–231
- Doyle JJ (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Fofana B, Duguid S, Cloutier S (2004) Cloning of fatty acid biosynthetic genes β -ketoacyl CoA synthase, fatty acid elongase, stearyl-ACP desaturase, and fatty acid desaturase and analysis of expression in the early developmental stages of flax (*Linum usitatissimum* L.) seeds. *Plant Sci* 166(6):1487–1496
- Gesteira AS, Schuster I, José IC, Piovesan ND, Viana JMS, Barros EGD, Moreira MA (2003) Biometrical analyses of linolenic acid content of soybean seeds. *Genet Mol Biol* 26(1):65–68
- Gutierrez-Gonzalez JJ, Vuong TD, Zhong R, Yu O, Lee JD, Shannon G, Ellersieck M, Nguyen HT, Slepser DA (2011) Major locus and other novel additive and epistatic loci involved in modulation of isoflavone concentration in soybean seeds. *Theor Appl Genet* 123(8):1375–1385
- Ha B-K, Monteros MJ, Boerma HR (2010) Development of SNP assays associated with oleic acid QTLs in N00-3350 soybean. *Euphytica* 176(3):403–415
- Henderson MM (1991) Correlations between fatty acid intake and cancer incidence. In: Nelson GJ (ed) Health effects of dietary fatty acids. The American Oil Chemists' Society, Champaign, pp 136–149
- Hildebrand DF, Deng W, Grayburn WS (1993) Use of genetic transformation to alter fatty acid metabolism in plants. In: Hiatt A (ed) Transgenic plants: fundamentals and applications. Marcel Dekker Inc., New York, pp 3–21
- Hou G, Ablett GG, Pauls KP, Rajcan I (2006) Environmental effects on fatty acid levels in soybean seed oil. *J Am Oil Chem Soc* 83(9):759–763
- Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Rosner BA, Hennekens CH, Willett WC (1997) Dietary fat intake and the risk of coronary heart disease in women. *New Engl J Med* 337(21):1491–1499
- Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, Schmidt ME (2004a) Seed quality QTL in a prominent soybean population. *Theor Appl Genet* 109(3):552–561
- Hyten DL, Pantalone VR, Saxton AM, Schmidt ME, Sams CE (2004b) Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. *J Am Oil Chem Soc* 81(12):1115–1118
- Kamal-Eldin A, Andersson R (1997) A multivariate study of the correlation between tocopherol content and fatty acid composition in vegetable oils. *J Am Oil Chem Soc* 74(4):375–380
- Kim HK, Kim YC, Kim ST, Son BG, Choi YW, Kang JS, Park YH, Cho YS, Choi IS (2010) Analysis of quantitative trait loci (QTLs) for seed size and fatty acid composition using recombinant inbred lines in soybean. *J Life Sci* 20(8):1186–1192
- Lark KG, Chase K, Adler F, Mansur LM, Orf JH (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. *Proc Natl Acad Sci USA* 92(10):4656–4660
- Li H, Ribaut J-M, Li Z, Wang J (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor Appl Genet* 116(2):243–260
- Li H, Hearne S, Bänziger M, Li Z, Wang J (2010) Statistical properties of QTL linkage mapping in biparental genetic populations. *Heredity* 105(3):257–267
- Li H, Zhao T, Wang Y, Yu D, Chen S, Zhou R, Gai J (2011) Genetic structure composed of additive QTL, epistatic QTL pairs and collective unmapped minor QTL conferring oil content and fatty acid components of soybeans. *Euphytica* 182(1):117–132
- Ma L, Li B, Han FX, Yan SR, Wang LZ, Sun JM (2015) Evaluation of the chemical quality traits of soybean seeds, as related to sensory attributes of soymilk. *Food Chem* 173:694–701
- Mendes A, Kelly AA, van Erp H, Shaw E, Powers SJ, Kurup S, Eastmond PJ (2013) bZIP67 regulates the omega-3 fatty acid content of *Arabidopsis* seed oil by activating *FATTY ACID DESATURASE3*. *Plant Cell* 25(8):3104–3116
- Mensink RP, Katan MB (1992) Effect of dietary fatty acids on serum lipids and lipoproteins. A meta-analysis of 27 trials. *Arterioscler Thromb Vasc* 12(8):911–919
- Mostofsky DI, Yehuda S, Salem N (2001) Fatty acids: physiological and behavioral functions. Humana Press, Totowa
- Ohlrogge J, Browse J (1995) Lipid biosynthesis. *Plant Cell* 7(7):957–970
- Oliva M, Shannon J, Slepser D, Ellersieck M, Cardinal A, Paris RL, Lee J (2006) Stability of fatty acid profile in soybean genotypes with modified seed oil composition. *Crop Sci* 46(5):2069–2075
- Orf JH, Chase K, Adler FR, Mansur LM, Lark KG (1999) Genetics of soybean agronomic traits: II. Interaction between yield quantitative trait loci in soybean. *Crop Sci* 39(6):1652–1657
- Panthee DR, Pantalone VR, Saxton AM (2006) Modifier QTL for fatty acid composition in soybean oil. *Euphytica* 152(1):67–73
- Reinprecht Y, Poysa VW, Yu K, 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(12):1510–1527
- Robinson DS, Wu Z, Domoney C, Casey R (1995) Lipoxygenases and the quality of foods. *Food Chem* 54(1):33–43
- Shibata M, Takayama K, Ujiie A, Yamada T, Abe J, Kitamura K (2008) Genetic relationship between lipid content and linolenic acid concentration in soybean seeds. *Breed Sci* 58(4):361–366
- Spencer MM, Pantalone VR, Meyer EJ, Landau-Ellis D, Hyten DL (2003) Mapping the *Fas* locus controlling stearic acid content in soybean. *Theor Appl Genet* 106(4):615–619
- Stange M, Utz HF, Schrag TA, Melchinger AE, Würschum T (2013) High-density genotyping: an overkill for QTL mapping? Lessons learned from a case study in maize and simulations. *Theor Appl Genet* 126(10):2563–2574

- Wang Z, Cheng J, Chen Z, Huang J, Bao Y, Wang J, Zhang H (2012a) Identification of QTLs with main, epistatic and QTL \times environment interaction effects for salt tolerance in rice seedlings under different salinity conditions. *Theor Appl Genet* 125(4):807–815
- Wang X, Jiang G-L, Green M, Scott RA, Hyten DL, Cregan PB (2012b) Quantitative trait locus analysis of saturated fatty acids in a population of recombinant inbred lines of soybean. *Mol Breed* 30(2):1163–1179
- Xie D, Han Y, Zeng Y, Chang W, Teng W, Li W (2012) SSR- and SNP-related QTL underlying linolenic acid and other fatty acid contents in soybean seeds across multiple environments. *Mol Breed* 30(1):169–179
- Yaklich RW, Vinyard BT (2004) Estimating soybean seed protein and oil concentration before harvest. *J Am Oil Chem Soc* 81(2):189–194
- Yu H, Xie W, Wang J, Xing Y, Xu C, Li X, Xiao J, Zhang Q (2011) Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS One* 6(3): e17595