



Effect of marker segregation distortion on high density linkage map construction and QTL mapping in Soybean (*Glycine max* L.)

Jian-Fang Zuo¹ · Yuan Niu² · Peng Cheng¹ · Jian-Ying Feng³ · Shi-Feng Han³ · Ying-Hao Zhang¹ · Guoping Shu⁴ · Yibo Wang⁴ · Yuan-Ming Zhang¹

Received: 20 January 2019 / Revised: 16 May 2019 / Accepted: 17 May 2019 / Published online: 31 May 2019
© The Author(s) 2019. This article is published with open access

Abstract

Marker segregation distortion is a natural phenomenon. Severely distorted markers are usually excluded in the construction of linkage maps. We investigated the effect of marker segregation distortion on linkage map construction and quantitative trait locus (QTL) mapping. A total of 519 recombinant inbred lines of soybean from orthogonal and reciprocal crosses between LSZZH and NN493-1 were genotyped by specific length amplified fragment markers and seed linoleic acid content was measured in three environments. As a result, twenty linkage groups were constructed with 11,846 markers, including 1513 (12.77%) significantly distorted markers, on 20 chromosomes, and the map length was 2475.86 cM with an average marker-interval of 0.21 cM. The inclusion of distorted markers in the analysis was shown to not only improve the grouping of the markers from the same chromosomes, and the consistency of linkage maps with genome, but also increase genome coverage by markers. Combining genotypic data from both orthogonal and reciprocal crosses decreased the proportion of distorted markers and then improved the quality of linkage maps. Validation of the linkage maps was confirmed by the high collinearity between positions of markers in the soybean reference genome and in linkage maps and by the high consistency of 24 QTL regions in this study compared with the previously reported QTLs and lipid metabolism related genes. Additionally, linkage maps that include distorted markers could add more information to the outputs from QTL mapping. These results provide important information for linkage mapping, gene cloning and marker-assisted selection in soybean.

Introduction

Soybean (*Glycine max* L. Merr., $2n = 40$) is an important source of protein and oil for food and feed. Linkage map

plays an important foundational role in soybean genetics and molecular biology, such as quantitative trait locus (QTL) mapping and candidate gene mining, map-based cloning, whole-genome *de novo* assembly, marker-assisted selection, and genome selection (Xia et al. 2007). As we know, marker segregation distortion is a common phenomenon in linkage map construction and QTL mapping. In most research and applications, significantly distorted markers are either simply discarded or included without notice and their impact on high density linkage map construction and QTL mapping is largely unknown.

To date, many studies on linkage map construction in soybean have been reported. Initially, the linkage maps were constructed using morphology and isoenzyme markers, e.g., the first soybean genetic map was constructed by 57 classical markers (Palmer and Kilen 1987). Clearly, the number of markers was too small. With the development of more advanced technologies, molecular markers have been used to construct linkage maps. These markers have the potential to “saturate” the genome and this increases the probability of QTL detection (Keim et al. 1990). Although these markers are

Supplementary information The online version of this article (<https://doi.org/10.1038/s41437-019-0238-7>) contains supplementary material, which is available to authorized users.

✉ Yuan-Ming Zhang
soyzzhang@mail.hzau.edu.cn

- ¹ Crop Information Center, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China
- ² College of Life Sciences and Food Engineering, Huaiyin Institute of Technology, Huaian 223003, China
- ³ State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China
- ⁴ Center of Molecular Breeding and Biotechnology, Beijing Lantron Seed Corp., Beijing 100081, China

useful in the construction of high density linkage maps, large gaps occur frequently (Wu et al. 2001). Thus, integrated linkage maps have been often reported in soybean (Cregan et al. 1999; Yamanaka et al. 2001; Song et al. 2004; Xia et al. 2007), e.g., all the 165 RFLP, 25 RAPD and 650 AFLP markers in Keim et al. (1997), and all the 1015 SSR, 709 RFLPs, 73 RAPDs, 6 AFLPs and 46 other markers in Song et al. (2004). With the development of genomic sequencing technology in recent years, several thousand single nucleotide polymorphism (SNP) markers have become available to construct linkage maps (Choi et al. 2007; Hyten et al. 2010; Song et al. 2016). Recently, Collard and Mackill (2008) established, and Sun et al. (2013) described in detail, specific length amplified fragment sequencing (SLAF-seq) technology, as a high-resolution strategy, for high-throughput SNP genotyping. These SLAF markers have been frequently used to construct linkage maps. In soybean, Zhang et al. (2013), Qi et al. (2014) and Li et al. (2014) used 1233, 5308 and 5785 SLAF markers, respectively, to construct their high-density linkage maps. In the above studies, however, severely distorted markers were not included in their linkage maps.

Marker segregation distortion, defined as the significant deviation of the observed segregation ratio from the Mendelian ratio expected from the mating design of the studied population, is a common biological phenomenon (Lyttle 1991). To date, there have been many studies on this research topic (Wang et al. 2005; Zhu et al. 2007; McMullen et al. 2009; Gardner et al. 2016). Hackett and Broadfoot (2003) showed that segregation distortion has very little effect on marker order or length when the distances between adjacent markers are approximately 10 cM. However, the average genetic distance in the currently reported linkage maps is frequently less than 1 cM. Thus, this issue needs to be further studied.

In this study, five hundred and nineteen recombinant inbred lines (RILs) from orthogonal and reciprocal crosses of soybean cultivars LSZZH (P_1) with NN493-1 (P_2) were genotyped by SLAF markers and the seed linoleic acid content was measured in three environments. All the SNP markers, including very significantly distorted markers, were used to construct a high-density linkage map. In addition, the effect of these distorted markers on linkage map construction and QTL mapping was investigated. To validate the effectiveness of constructed linkage maps in this study, we evaluated the quality of the linkage maps, and detected QTLs for seed linoleic acid content in soybean.

Materials and methods

Plant materials and DNA extraction

Five hundred and nineteen recombinant inbred lines (RILs) derived from the orthogonal (242, OC) and reciprocal (277,

RC) crosses between two parents LSZZH (P_1) and NN493-1 (P_2), together with their parents, were planted in three-row plots in a completely randomized design at the Jiangpu experimental station of Nanjing Agricultural University in 2015 (NJ2015), and at the Wuhan and Ezhou experimental stations of Huazhong Agricultural University, respectively, in 2014 (WH2014) and 2015 (EZ2015). The plots were 1.5 m wide and 2.0 m long, and approximately 15 plants were planted in each row. Five plants in the middle row for each line were randomly harvested, and 10 g of seeds were prepared for the measurement of linoleic acid content.

The young healthy leaves from 519 RILs as well as from their two parents were collected and frozen in liquid nitrogen. Total genomic DNA was extracted from each leaf sample using the cetyltrimethyl ammonium bromide (CTAB) method (Doyle 1990). All the information obtained from 519 RILs and their parents was used for genotyping and mapping analyses.

Genotyping the RIL population with the SLAF-seq method

SLAF-seq technology, developed by the Beijing Biomarker Technology Corporation (Collard and Mackill 2008), was used to genotype 519 RILs and the two parents in this study. The sequencing and preliminary bioinformatics analyses were as described by Sun et al. (2013) with minor modifications. The soybean reference genome (*Glycine max* Wm82.a1.v1) was used to align the SLAF markers for SNP discovery. The SLAF library construction was performed accordingly using the predesigned scheme. Genomic DNA from each sample was digested by two restriction enzymes, *RsaI* (NEB, Ipswich, MA, USA) and *HaeIII* enzyme. Then, restriction-ligation samples were diluted and mixed with dNTP, Taq DNA polymerase (NEB) and primer for PCR reactions. The PCR productions were purified using E.Z.N.A.® Cycle Pure Kit (Omega) and pooled. The pooled samples were incubated with *MseI*, T4-DNA ligase, ATP and Solexa adapter at 37°C, then purified with a Quick Spin column (Qiagen, Hilden, Germany), and electrophoresed on a 2% agarose gel. After gel purification, DNA fragments 314–414 bp in length with indices and adaptors, defined as SLAF tags, were excised and diluted for pair-end sequencing on an Illumina high-throughput sequencing platform (Illumina, Inc.; San Diego, CA, US).

The SLAF-seq data grouping and genotyping were described in detail by Sun et al. (2013). Based on sequence similarity, all the SLAF pair-end reads with clear index information were clustered. To reduce computational intensity, identical reads were merged together, and sequence similarity was detected using one-to-one alignment by BLAT (tileSize = 10, stepSize = 5) (Kent 2002). Sequences with over 90% identity were grouped to one

SLAF locus. Through the minor allele frequency (MAF) evaluation, alleles were defined in each SLAF marker. Because soybean is a diploid species, one locus can contain no more than four SLAF tags; groups containing over four tags were considered to be repetitive SLAFs and were filtered out. Alleles of each SLAF locus were then defined according to the average sequence depths of SLAF markers, which were greater than 10-fold in parents and greater than 3-fold in RILs. High-quality SLAF markers for the genetic mapping were filtered by the following criteria: (i) average sequence depths should be > 10-fold in the parents, (ii) markers with more than 25% missing data were excluded, and (iii) the loci containing more than four SLAF tags were excluded.

Linkage map construction

The program MSTmap (Wu et al. 2008) was used to group and order the SLAF markers. Considering the existence of marker segregation distortion, the software DistortedMap of Xie et al. (2014) was used to detect marker segregation distortion. The quality of the constructed linkage maps was evaluated by the collinearity between the linkage maps in this study and the soybean reference genome, the heat map and the uniform distribution of recombination fractions on the genome. The collinearity was measured by the Spearman correlation coefficient, which was calculated by the R function *cor.test*. The higher the Spearman correlation coefficient, the better the collinearity. The R program *pheatmap* (<https://cran.r-project.org/web/packages/pheatmap/index.html>) was used to construct the heat maps.

Linoleic acid content measurement

In WH2014, EZ2015 and NJ2015, approximately 10 g of seeds collected from five plants per RIL and the two parents were ground using a pulverizer, and the seed powder was filtered. Then, 30 mg soybean powder was used to extract fatty acid. Five fatty acids for each line were measured by gas chromatography with a flame ionization detector and a Permabond FFAP stainless steel column (50 m × 0.2 mm × 0.33 μm, ThermoFisher Scientific, Waltham, MA) at the Wuhan Research Branch of the National Rapeseed Genetic Improvement Center in 2014 and 2015, respectively, and the details were described by Zhou et al. (2016).

Genome-wide composite interval mapping of QTLs for linoleic acid content

The linoleic acid contents for each RIL in the three environments, WH2014, EZ2015 and NJ2015, were indicated by datasets I, II and III, respectively. The BLUP values (dataset

IV) were predicted by the R program *lme4* (<https://CRAN.R-project.org/package=lme4>).

All the above four datasets for seed linoleic acid content, along with marker genotypic information and linkage maps, were used to detect QTLs using genome-wide composite interval mapping (Wang et al. 2016), implemented by the QTL.gCIMapping.GUI program (<https://cran.r-project.org/web/packages/QTL.gCIMapping.GUI/index.html>). The covariate on the orthogonal and reciprocal crosses was included in the genetic model while all the 519 RILs were jointly analyzed. The walk speed for genome-wide scanning was set at 1 cM. The LOD score thresholds for significant QTLs at the 0.05 probability level were calculated based on 1000 permutations using the Windows QTL Cartographer v2.5 software (Wang et al. 2012). The QTLs with the LOD scores between the threshold and 2.5 were viewed as suggestive QTLs (Lander and Kruglyak 1995).

The QTLs, detected repeatedly across the above four datasets, were viewed as stable. If the QTLs identified across various datasets were within 5 cM, these QTLs were viewed as being the same (Song et al. 2004). If the QTL region in the linkage maps including distorted markers (case I) overlapped, in the physical positions of the genome, with ones excluding distorted markers (case II), these QTLs were also viewed as being the same. The nomenclature for detected QTL was denoted as *q* + trait name + chromosome + the number of QTL on the chromosome, such as “*qLAI-1*”, “*qLA*” indicated one QTL for linoleic acid content in soybean, and “*1-1*” indicates the first QTL on chromosome 1 (McCouch et al. 1997).

Previously reported QTLs from <https://www.soybase.org/> and 1123 lipid metabolism related genes in Zhang et al. (2016) were used to identify the true QTLs for seed linoleic acid content in soybean. These true QTLs were used to validate the correctness of the linkage maps in this study.

Results

Genotyping of RIL population using the SLAF-seq method

Using the SLAF-seq method, a total of 384.64 M of raw data was generated; this comprised 22,595,092 reads for the female parent, 31,069,592 reads for the male parent, and 635,267 reads on average for each RIL. After clustering of the reads with their reference, a total of 418,100 SLAF labels were obtained from all the RILs and two parents. More specifically, 392,904 SLAFs were generated from 21,133,891 reads for the male (average sequence depth 53.79-fold), 391,472 SLAFs from 14,542,086 reads for the female (37.15-fold), and 165,417 SLAFs from 427,897 reads on average for each RIL (2.57-fold). Among the

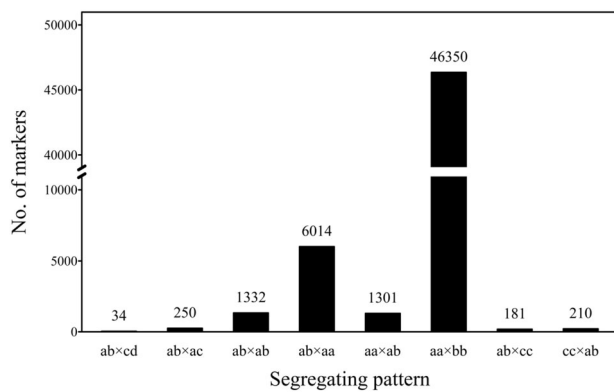


Fig. 1 Number of markers with various segregation patterns in soybean. ab × cd, ab × ac, ab × ab, ab × aa, aa × ab, aa × bb, ab × cc and cc × ab are segregation patterns of markers, where the English alphabets from a to d are the alleles for each marker

418,100 SLAF labels, 58,351 (13.96%) were polymorphic. All of the SLAF labels were mapped on the soybean reference genome (*Glycine max* Wm82.a1.v1) using SOAP software (Li et al. 2008), 411,262 were distributed on 20 chromosomes (Fig S1), and 6838 were on a scaffold. After filtering out the SLAF markers lacking parent information, 55,672 were retained and classified into eight types (Fig. 1). In the RIL population, only the 46,350 SLAF markers with the aa × bb segregation pattern were used to construct the linkage map (Fig. 1). In order to ensure the quality of our linkage map, the SLAF labels with an average sequence depth <10-fold in the parents, >25% missing data, and >4 SLAF tags, were filtered out. Among the 11,979 SLAF labels finally used in this study, 11,846 were mapped on 20 chromosomes, and 133 were on a scaffold.

Construction of a high-density genetic linkage map in soybean

Construction of linkage maps in the RILs from the orthogonal and reciprocal crosses

With OC RILs, all the 11,846 markers covered 2631.89 cM on 20 soybean chromosomes with an average marker-interval of 0.223 cM; total map length for each chromosome ranged from 68.13 cM (chr14) to 206.21 cM (chr2), and chromosomes 14 and 5 had the minimum (0.14 cM) and maximum (0.71 cM) average marker-intervals, respectively; the maximum (18.68 cM) and minimum (2.49 cM) marker-gaps were found on chromosomes 19 and 14, respectively (Table 1).

With RC RILs, all the 11,846 markers covered 2643.24 cM across 20 soybean chromosomes with an average marker-interval of 0.224 cM; total map length for each chromosome ranged from 77.27 cM (chr19) to 196.18 cM (chr9), and chromosomes 15 and 5 showed the

minimum (0.12 cM) and maximum (0.71 cM) average marker-intervals, respectively; the maximum (15.75 cM) and minimum (3.12 cM) marker-gaps were found on chromosomes 19 and 8, respectively (Table 1).

With all the RILs, all the 11,846 markers covered 2475.86 cM across 20 soybean chromosomes with an average marker-interval of 0.209 cM; the number of markers for each chromosome ranged from 152 (chr5) to 1327 (chr15); total map length for each chromosome ranged from 67.07 cM (chr14) to 176.26 cM (chr2), and chromosomes 15 and 5 showed the minimum (0.11 cM) and maximum (0.71 cM) average marker-intervals, respectively; the maximum (17.03 cM) and minimum (2.71 cM) marker-gaps were found on chromosomes 19 and 8, respectively (Table 1; Fig. 2).

Effect of marker segregation distortion on linkage map construction

The software DistortedMap was used to detect marker segregation distortion. As a result, 2306, 2431 and 1999 markers showed the distortion from Mendelian segregation in the OC, RC and all the RILs, respectively (Table 2). Among these distorted markers, 750, 609 and 486 in the OC, RC and all the RILs, respectively, were significant ($0.01 < P\text{-value} \leq 0.05$), and 1556, 1822 and 1513 were very significant ($P\text{-value} \leq 0.01$). As shown in Table 2, more than 100 distorted markers with $P\text{-values} \leq 0.05$ were observed on chromosomes 2 (131), 15 (128), 17 (1130), and 18 (394) in the OC RILs, on chromosomes 2 (914), 6 (116), 11 (265), 13 (461) and 19 (222) in the RC RILs, and on chromosomes 3 (107), 11 (255), 17 (1097) and 18 (175) in all the RILs. More than 100 distorted markers with $P\text{-values} \leq 0.01$ were observed on chromosomes 17 (1107, 94.78%) and 18 (240) in the OC RILs, on chromosomes 2 (857, 76.79%), 6 (103), 11 (228) and 13 (405) in the RC RILs, and on chromosomes 11 (187) and 17 (1070, 91.61%) in all the RILs. Among these distorted markers, 63 and 18 were common between the orthogonal and reciprocal crosses at the 0.05 and 0.01 probability levels, respectively; 53 and 15 were common across the above three populations at the 0.05 and 0.01 probability levels, respectively.

The program MSTMap (Wu et al. 2008) was used to construct linkage maps using marker datasets that include or exclude the distorted markers in the OC, RC and all the RILs. With the marker datasets that include distorted markers, all the 11846 markers were grouped into 20 linkage groups and each linkage group across different populations has the same number of markers (Table 1). With the marker datasets that exclude all the very significantly distorted markers with $P\text{-values} \leq 0.01$, 10,290 markers in OC RILs were grouped onto 22 linkage groups, because 2 and 4 markers on chromosomes 15 and 17 were clustered onto

Table 1 Description of basic characteristics for 20 linkage groups in orthogonal cross, reciprocal cross and all the BRILs

Chr.	No. of markers	All the RILs					Orthogonal cross RILs					Reciprocal cross RILs				
		Total genetic distance (cM)	Average genetic distance (cM)	Maximum gap (cM)	Collinearity ^a	Total genetic distance (cM)	Average genetic distance (cM)	Maximum gap (cM)	Collinearity ^a	Total genetic distance (cM)	Average genetic distance (cM)	Maximum gap (cM)	Collinearity ^a			
1	507	150.48	0.30	7.14	0.9962	147.36	0.29	5.72	0.9831	158.33	0.31	8.41	0.9857			
2	1116	176.26	0.16	4.84	0.8705	206.21	0.18	5.39	0.8737	177.20	0.16	4.98	0.8848			
3	917	137.01	0.15	6.9	0.9372	146.49	0.16	5.72	0.9271	149.49	0.16	7.99	0.8773			
4	277	80.78	0.29	3.34	0.9406	79.27	0.29	4.14	0.9387	86.63	0.31	4.18	0.9392			
5	152	107.95	0.71	4.24	0.9339	107.96	0.71	4.75	0.9356	107.92	0.71	4.67	0.9338			
6	466	158.82	0.34	4.05	0.9998	164.10	0.35	4.10	0.9995	169.70	0.36	4.14	0.9987			
7	637	153.64	0.24	4.47	0.9645	157.59	0.25	5.72	0.9545	177.66	0.28	5.7	0.9534			
8	229	100.04	0.44	2.71	0.9991	101.29	0.44	3.20	0.9991	104.72	0.46	3.12	0.9990			
9	917	158.93	0.17	5.16	0.9850	159.14	0.17	5.02	0.9551	196.18	0.21	5.29	0.8997			
10	669	149.87	0.22	4.21	0.8214	155.36	0.23	4.34	0.6557	154.13	0.23	4.1	0.7813			
11	354	80.02	0.23	3.41	0.8414	79.70	0.23	3.45	0.8208	85.53	0.24	3.38	0.8157			
12	163	82.72	0.51	6.22	0.9992	74.00	0.45	5.86	0.9988	85.45	0.52	5.24	0.9992			
13	646	159.15	0.25	13.32	0.9999	185.81	0.29	15.88	0.9897	163.71	0.25	11.22	0.9889			
14	499	67.07	0.13	2.78	0.9883	68.13	0.14	2.49	0.9858	79.26	0.16	3.72	0.9879			
15	1327	147.50	0.11	5.96	0.9431	183.80	0.14	6.51	0.8918	159.18	0.12	5.51	0.8488			
16	670	119.53	0.18	3.97	0.8950	138.46	0.21	3.48	0.8837	118.09	0.18	4.41	0.8669			
17	1168	150.68	0.13	4.78	0.9550	183.79	0.16	4.47	0.8889	154.71	0.13	5.39	0.8095			
18	610	136.60	0.22	7.75	0.9926	139.83	0.23	8.23	0.9901	154.69	0.25	6.93	0.9933			
19	236	77.81	0.33	17.03	0.9961	77.81	0.33	18.68	0.9941	77.27	0.33	15.75	0.9950			
20	286	81.01	0.28	4.24	0.9968	75.81	0.27	4.61	0.9964	83.40	0.29	4.18	0.9953			
Total	11846	2475.86	0.21	17.03		2631.89	0.22	18.68		2643.24	0.22	15.75				

^aThe collinearity is measured by correlation coefficient

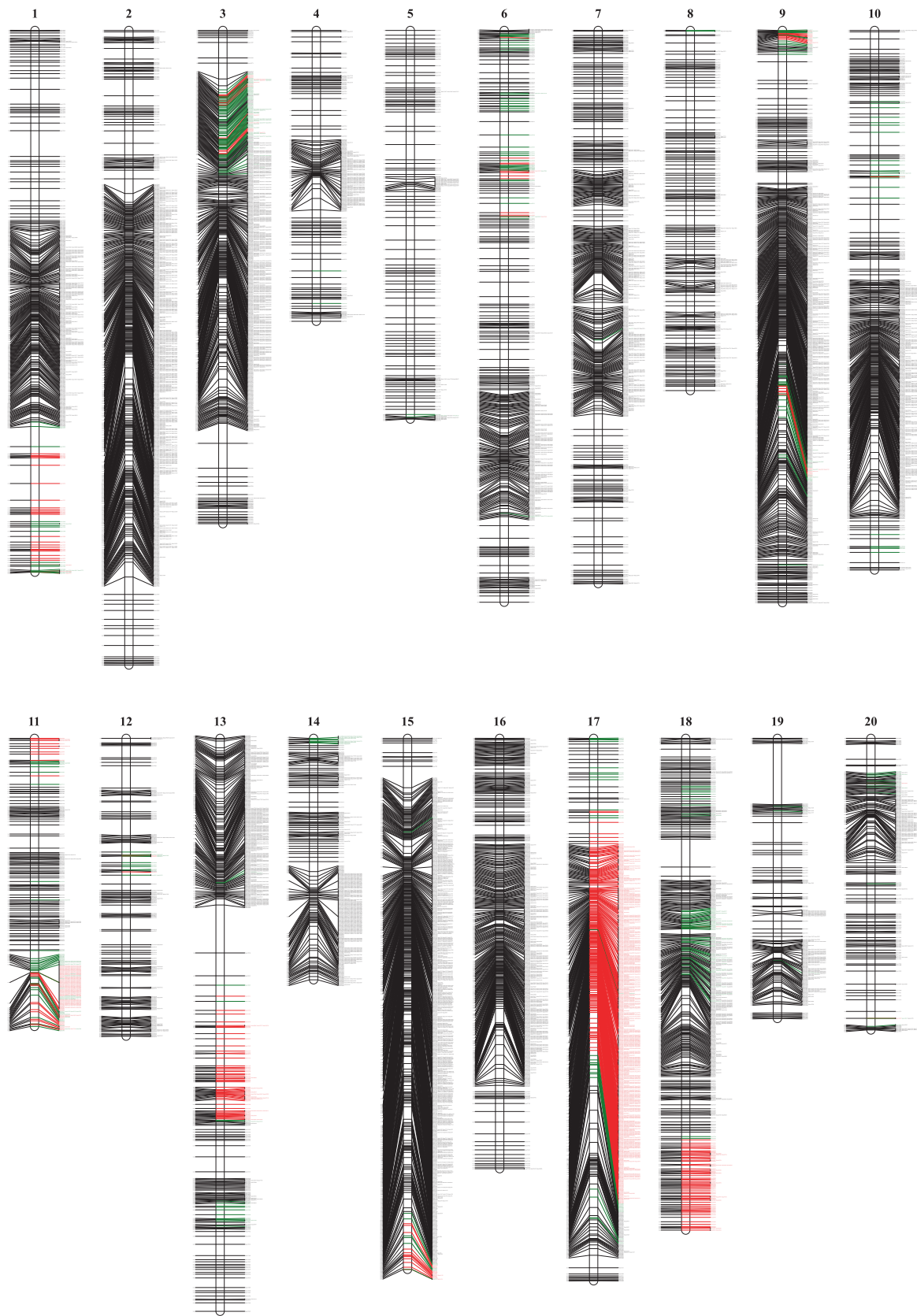


Fig. 2 High-density genetic linkage maps constructed by 11846 markers in all the RILs from the orthogonal and reciprocal crosses between LSZZH and NN493-1 in soybean. The significantly distorted markers at the 0.05 and 0.01 probability levels are marked by green and red colors, respectively

Table 2 Information on segregation distorted markers in orthogonal cross, reciprocal cross and all the RILs

Chr.	No. of markers	All the RILs			Orthogonal cross RILs			Reciprocal cross RILs		
		0.01 < <i>P</i> < 0.05	<i>P</i> ≤ 0.01	Total	0.01 < <i>P</i> < 0.05	<i>P</i> ≤ 0.01	Total	0.01 < <i>P</i> < 0.05	<i>P</i> ≤ 0.01	Total
1	507	20	28	48	18	29	47	2	0	2
2	1116	0	0	0	122	9	131	57	857	914
3	917	89	18	107	6	1	7	90	2	92
4	277	2	0	2	0	0	0	7	0	7
5	152	3	0	3	28	9	37	9	0	9
6	466	44	13	57	1	0	1	13	103	116
7	637	1	0	1	24	3	27	9	10	19
8	229	2	0	2	45	27	72	13	1	14
9	917	33	22	55	70	20	90	0	0	0
10	669	22	1	23	60	6	66	14	10	24
11	354	68	187	255	0	0	0	37	228	265
12	163	13	3	16	19	14	33	24	9	33
13	646	20	76	96	27	9	36	56	405	461
14	499	24	0	24	22	9	31	2	0	2
15	1327	7	13	20	62	66	128	20	9	29
16	670	0	0	0	10	0	10	50	0	50
17	1168	27	1070	1097	23	1107	1130	15	0	15
18	610	95	80	175	154	240	394	14	80	94
19	236	4	0	4	38	1	39	130	92	222
20	286	12	2	14	21	6	27	47	16	63
Total	11846	486	1513	1999	750	1556	2306	609	1822	2431

two additional linkage groups; 10,024 markers in RC RILs were grouped onto 23 linkage groups, because 34, 4 and 9 markers on chromosomes 2, 6 and 11 were clustered onto three additional linkage groups; 10,333 markers in all the RILs were grouped onto 23 linkage groups, because 93 markers on chromosome 13, and 2 and 24 markers on the chromosome 17 were clustered onto three additional linkage groups (Table S1, Fig S2). We compared the above two kinds of linkage maps (including or excluding the very significantly distorted markers with P -value ≤ 0.01). First, in most cases the marker orders of linkage maps excluding the very significantly distorted markers were consistent with those including distorted markers. However, we found three inconsistent regions on chromosome 2 in OC RILs and on chromosome 10 in RC RILs (Fig S2). On chromosome 2, there were two inconsistent regions. The first region included 217 markers, which were between Marker1198674 (56.48 and 54.98 cM on the linkage maps including and excluding the very significantly distorted markers, respectively) and Marker1293705 (79.59 and 78.51 cM). The second region included 585 markers, which were between Marker1187344 (83.57 and 82.49 cM) and Marker1290774 (129.61 and 132.51 cM). On chromosome 10, there was one inconsistent region. This region included 444

markers, which were between Marker1053787 (84.00 and 83.82 cM) and Marker1019033 (112.14 and 111.22 cM).

If all the very significantly distorted markers with P -value ≤ 0.01 are excluded, then the total number of markers changed from 11,846 to 10,333, 10,290 and 10,024 in all, OC and RC RILs, respectively. The total genetic distances changed from 2475.86 to 2336.74 cM in all the RILs, from 2631.89 to 2464.72 cM in the OC RILs, and from 2643.24 to 2495.10 cM in the RC RILs. The average marker-interval changed from 0.21 to 0.23 cM in all the RILs, from 0.22 to 0.24 cM in the OC RILs, and from 0.22 to 0.25 cM in the RC RILs (Table S1). The above changes are mainly derived from a few chromosomes, e.g., the numbers of markers changed from 1168 to 98 and 61 on chr17 in the all and OC RILs, respectively, and from 1116 to 259 on chr2 in the RC RILs. The total genetic distances reduced from 150.68 and 183.79 to 84.61 and 58.37 cM on chr17 in the all and OC RILs, respectively, and from 177.20 to 137.83 cM on chr2 in the RC RILs. The average marker-interval reduced from 0.13 and 0.16 to 0.86 and 0.96 cM on chr17 in the all and OC RILs, respectively, and from 0.16 to 0.53 cM on chr2 in the RC RILs. This means that more markers increase total genetic distance and marker density on the chromosome.

In addition, we found one interesting phenomenon on chr2. Although one hundred and thirty-one (9) and 914

(857) markers had (highly) significant segregation distortion from the Mendelian ratio in the OC and RC RILs, respectively, no marker segregation distortion was observed in all the RILs (Table 2).

The evaluation of linkage map constructed from all the RILs

The collinearity of linkage maps with soybean reference genome

To evaluate the collinearity between linkage maps and soybean reference genome, the Spearman correlation coefficient for each chromosome was calculated. The results are listed in Table 1. Among all the twenty coefficients in the OC RILs, 14 were larger than 0.90, 6 were larger than 0.99, and their range was from 0.6557 (chr10) to 0.9995 (chr6). Among these coefficients in the RC RILs, 12 were larger than 0.90, 6 were larger than 0.99, and their range was from 0.7813 (chr10) to 0.9992 (chr12). Among these coefficients in all the RILs, 16 were larger than 0.90, 8 were larger than 0.99, and their range was from 0.8214 (chr10) to 0.9999 (chr13). In addition, the consecutive curves between soybean genome and linkage groups are found in Fig S3. The high collinearity suggests that the genetic linkage maps we constructed are congruent with the soybean genome for marker orders.

If all the very significantly distorted markers with P -values ≤ 0.01 are excluded, in most cases the Spearman correlation coefficients between linkage maps and soybean reference genome were slightly reduced, and special situations were found for chromosomes 2, 11 and 17 in the RC, all and OC RILs, respectively (Table S2). The three relatively high collinearities were derived from very small (259, 167 and 61) markers, because 857, 187 and 1107 markers were excluded in the above three cases.

Heat map

To assess the quality of the linkage map, heat maps were generated using pair-wise recombination fractions (r) for the 11,846 SLAF markers (Fig S4). The size for r was indicated by different colors ranging from yellow (lower) to purple (higher). As shown in Fig S4, the colors on and near diagonal lines for all the chromosomes are yellow, indicating their lower recombination or high linkage disequilibrium, and the squares of different size along the diagonal lines indicates the existence of LD blocks of different size.

The recombination pattern in soybean genome

Each chromosome was divided into 20 intervals based on their physical positions. In each interval, the recombinant

fraction between adjacent markers was obtained, and the sum of all the recombinant fractions in this interval was viewed as the recombinant fraction of this interval. The results are shown in Fig.S5. In Fig. S5, the recombination fractions were high in the intervals near the two ends of each chromosome and low in the intervals of middle region of each chromosome. This result is consistent with those in previous studies.

Multi-QTL mapping for linoleic acid content in soybean

Mapping QTL for seed linoleic acid content in soybean

The linoleic acid content of LSZZH was larger than NN493-1 in all the environments, and there exists large variation among all the RILs in all the environments, e.g., the linoleic acid contents ranged from 40.64 to 57.70 (%) in WH2014. Similar phenomena were also observed in the other three phenotypic datasets (Table S3). The frequency distributions for the four phenotypic datasets are shown in Fig S6.

If the linkage maps including very significantly distorted markers (case I) were used to conduct QTL mapping, the LOD score thresholds at the 0.05 probability level were 3.77, 3.66, 3.60 and 3.67, respectively, for the datasets I to IV. A total of 13 significant QTLs were detected. Among these 13 QTLs, 3, 2, 8 and 7 were found to be associated with linoleic acid content, respectively, for the datasets I to IV; 10 were previously reported; 9 were around the lipid metabolism related genes in Zhang et al. (2016) (Fig. 3, Table 3 and S4). In addition, a total of 11 suggestive QTLs were identified. Among these 11 suggestive QTLs, 5, 2, 8 and 7 were found to be associated with linoleic acid content, respectively, for the datasets I to IV; 8 were previously reported; 7 were around the lipid metabolism related genes in Zhang et al. (2016) (Fig. 3, Table 3 and S4). In summary, a total of 24 QTLs were identified, 18 were previously reported QTLs and 16 were around the lipid metabolism related genes (Fig. 3 and Table S4).

Effect of marker segregation distortion on QTL mapping

If the linkage maps excluding very significantly distorted markers (case II) were used to conduct QTL mapping, the LOD score thresholds at the 0.05 probability level were 3.65, 3.70, 3.67, and 3.61, respectively, for the datasets I to IV. A total of 15 significant QTLs were detected. Among these 15 QTLs, 2, 4, 6 and 9 were found to be associated with linoleic acid content, respectively, for the datasets I to IV; 11 were previously reported; 13 were around the lipid metabolism related genes in Zhang et al. (2016) (Fig S7, Tables S4). In addition, a total of 5 suggestive QTLs were identified, 2, 3, 4 and 5 were found to be associated with

Table 3 The linoleic acid content QTLs detected by GCIM and confirmed by both previously reported QTLs in soybean and/or lipid metabolism related genes

QTL	Chr	QTLs detected	Previously reported QTLs				Lipid metabolism related genes around QTLs					
			Position (cM)	Marker intervals	LOD	Effect	r^2 (%) ^a	Environment ^b	Type	QTL	Reference	Gene ^c
<i>qLA2-1</i>	2	111.81	Marker1183683	2.85	-0.54	2.11	WH2014	Suggestive	<i>qFA2</i>	Li et al. 2017	<i>Glyma02g13700</i>	Salas and Ohlroge 2002;
<i>qLA4-1</i>	4	22.26	Marker2230222	3.28	0.31	1.99	BLUP	Suggestive			<i>Glyma04g37420</i> (<i>FATB</i>)	Ozseyhan et al. 2018
<i>qLA5-1</i>	5	7.97	Marker2204980	2.74	0.41	1.75	NI2015	Suggestive	<i>qOA5_3</i>	Li et al. 2017	<i>Glyma05g07880</i> (<i>OBC</i>)	Wang et al. 2008
<i>qLA5-2</i>	5	107.75; 107.95	Marker2104599; Marker2161702	6.47; 4.24; 10.72	0.43; 1.07; 0.52	3.84; 8.25; 2.77	BLUP; H2014; NI2015	Significant	<i>seed linoleic 1-1, 1-2, 1-3</i>	Diers and Shoemaker (1992); Li et al. (2017); Cao et al. (2017)	<i>Glyma05g33790</i>	Chen et al. 2018
<i>qLA7-1</i>	7	12.21	Marker401588	5.01	0.40	3.28	BLUP	Significant	<i>qFA7_6</i>	Li et al. 2017	<i>Glyma07g03490</i> (<i>PLD</i>)	Kuppusamy et al. 2014
<i>qLA7-2</i>	7	40.33	Marker297491	8.47; 3.56	-0.77; -0.33	6.06; 2.29;	NI2015; BLUP	Significant	<i>qLA7_2</i>	Li et al. 2017	<i>Glyma07g08740</i> (<i>PLD</i>)	Kuppusamy et al. 2014
<i>qLA7-4</i>	7	79.68; 83.71	Marker274074; Marker347920	4.00; 4.20	0.53; 0.69	2.88; 3.41	NI2015; WH2014	Significant			<i>Glyma07g17720</i> (<i>GPA76</i>)	Misra et al. (2017)
<i>qLA8-2</i>	8	39.09	Marker791165~Marker718573	3.37	-0.57	2.08	EZ2015	Suggestive	<i>qSA8_1</i>	Li et al. (2017)		
<i>qLA9-1</i>	9	151.54	Marker468763	3.42	-0.32	2.11	BLUP	Suggestive	<i>qPA9_6</i>	Li et al. 2017	<i>Glyma09g39911</i> (<i>DGD</i>)	Kelly et al. 2003
<i>qLA11-1</i>	11	47.19	Marker584683	2.59	-0.50	1.79	WH2014	Suggestive	<i>seed linoleic 1-5</i>	Diers and Shoemaker (1992)		
<i>qLA11-2</i>	11	79.33	Marker575184	3.26	0.31	1.94	BLUP	Suggestive			<i>Glyma11g36460</i>	Salminen et al. 2016
<i>qLA12-1</i>	12	59.46	Marker2734590	7.48	-0.71	5.12	NI2015	Significant	<i>qFA12_2</i>	Li et al. 2017	<i>Glyma12g08010</i> (<i>KCS</i>)	Todd et al. 1999; Lee et al. 2010
<i>qLA12-2</i>	12	65.93	Marker2675339	5.75	-0.41	3.48	BLUP	Significant			<i>Glyma12g07230</i>	
<i>qLA13-1</i>	13	65.95	Marker2850221	4.28	-0.54	3.00	NI2015	Significant	<i>seed linoleic 6-3</i>	Bachlava et al. (2009)	<i>Glyma13g16790</i> (<i>PDAT</i>)	Dahlqvist et al. (2000); Eskandari et al. 2013
<i>qLA13-2</i>	13	90.96	Marker2775509	14.39; 9.69	-1.29; -0.55	10.70; 6.24	EZ2015; BLUP	Significant	<i>qOI-13-1</i>	Cao et al. (2017)		
<i>qLA13-3</i>	13	98.49	Marker2761097	4.03	-0.53	2.85	NI2015	Significant	<i>qFA13_3</i>	Li et al. 2017		
<i>qLA13-5</i>	13	156.00	Marker2850173	3.02; 4.53	0.56; 0.37;	2.03; 2.74	EZ2015; BLUP	Significant	<i>seed linoleic 7-3</i>	Kim et al. (2010)	<i>Glyma13g42310</i> (<i>LOX</i>); <i>Glyma13g42320</i> (<i>LOX</i>)	Radmark et al. 2007
<i>qLA14-1</i>	14	26.12; 26.81; 28.59;	Marker1690474; Marker1626134	5.07 9.27 11.86 4.98	-0.73 -0.52 -0.90 -0.73	3.84 5.62 8.26 3.42	WH2014; EZ2015; NI2015; BLUP	Significant	<i>qFA14_4</i>	Li et al. 2017	<i>Glyma14g08400</i> (<i>LPAT</i>)	Misra et al. 2014.
<i>qLA16-1</i>	16	115.60	Marker2543407	2.69	0.52	2.00	WH2014	Suggestive	<i>qPA16_2</i>	Li et al. 2017	<i>Glyma16g01070</i> (<i>LOX</i>)	Radmark et al. 2007

Table 3 (continued)

QTL	Chr	QTLs detected	Previously reported QTLs				Lipid metabolism related genes around QTLs															
			Position (cM)	Marker intervals	LOD	Effect	r^2 (%) ^a	Environment ^b	Type	QTL	Reference	Gene ^c	Reference									
qLA19-2	19	43.08; 44.28	Marker1495349; Marker1590191	3.27; 4.33	-0.46 -0.36	2.21 2.65	NJ2015; BLUP	Significant	qLA19_4	Li et al. 2017												
													21.77	Marker1429985	2.51	-0.49	1.76	WH2014	Suggestive	seed linoleic 6-8	Bachlava et al. (2009); Fan et al. (2015)	
qLA20-2	20	37.30	Marker1439858	2.86	-0.29	1.77	BLUP	Suggestive	seed linoleic 6-8	Bachlava et al. (2009); Fan et al. (2015)	Glyma20g25640 (PAP); Glyma20g25650 (PAP)	Millar et al. 2000										

^a Proportion of total phenotypic variance explained by single QTL

^b Environment for the detected QTL

^c Genes around the detected QTLs in this study. The same QTLs using the linkage maps of including and excluding very significantly distorted markers are marked by bold type

RILs from orthogonal and reciprocal crosses are used to construct the current linkage maps. In this situation, the number of distorted markers in all the RILs (1999) is much less than those in OC (2306) or RC (2431) RILs, especially, on chromosomes 2, 13 and 19, and the quality of linkage maps can be improved. Therefore, all the markers in the biparental segregation population from orthogonal and reciprocal crosses should be used to construct linkage maps in soybean.

Marker segregation distortion may be due to various reasons (Lyttle 1991). In this study, the numbers of distorted markers on chromosomes 2, 6, 13, 18, and 19 in all the RILs are significantly less than the total number of distorted markers in OC and RC RILs. In the two subpopulations, the distorted markers are biased towards female parental genotypes on chromosomes 2, 18 and 19 and towards male parental genotypes on chromosome 6. On chromosome 13, these distorted markers in OC RILs are biased towards female parental genotypes, while 72 (15.6%) and 389 (84.4%) distorted markers in RC RILs are biased towards male and female parental genotypes, respectively. Obviously, the distortion may be caused by gametic selection. This hypothesis is partly supported by the presence of male sterile genes on chromosomes 2 and 13 (Jin et al. 1998; Yang et al. 2014). Based on the above results, we doubt that there are female and male sterile genes, respectively, in chromosomes 6 and 19. On chromosome 17, however, there are 1130, 15 and 1097 distorted markers, respectively, in OC, RC and all the RILs. These distorted markers are biased toward female parental genotypes in OC RILs and toward male parental genotypes in RC RILs, and the reason for marker segregation distortion may be different from those in the above situations.

The current linkage maps have been validated in two aspects. On one hand, all the 11,846 markers are exactly clustered into 20 chromosomes in OC, RC and all the RILs, and the number of markers for each chromosome in the above three populations is exactly the same. The collinearities of the soybean reference genome with each linkage map are relatively high. The recombination fractions in the terminals for each chromosome were significantly higher than those in the middle intervals, which is consistent with results from previous studies. On the other hand, the results of mapping QTL for linoleic acid content in soybean in this study also validate the accuracy of the linkage maps. This is because, among 24 QTLs identified in this study, 18 are previously reported QTLs, and 16 are around lipid metabolism related genes (Zhang et al. 2016) (Table S4).

As compared with the results of excluding distorted markers, inclusion of distorted markers increases the accuracy of grouping all the markers on chromosomes. The reasons are as follows. First, the numbers of distorted markers in chromosomes 2, 6, 13, 18, and 19 in all the RILs

Table 4 Comparison of the soybean linoleic acid content QTLs using linkage maps generated by either including or excluding very significantly distorted markers

Case	QTLs from linkage maps that include very significantly distorted markers		QTLs from linkage maps that exclude very significantly distorted markers	
	No. of detected QTLs	No. of previously reported QTLs	No. of detected QTLs	No. of previously reported QTLs
Wuhan 2014	8 (3 ^a)	6 (3)	4 (3)	3 (3)
Ezhou 2015	4 (2)	2 (2)	7 (2)	4 (2)
Nanjing 2015	10 (7)	7 (6)	10 (7)	8 (5)
BLUP	12 (10)	9 (8)	14 (10)	12 (8)
Total	24 (14)	16 (12)	20 (14)	13 (11)

^aThe number of the same QTLs identified with the linkage maps of both including and excluding very significantly distorted markers were listed in the bracket, and the same is true for the number of previously reported QTLs and the number of lipid metabolism related genes around QTLs in this study. The QTLs include significant and suggestive ones

are significantly fewer than those in OC or RC RILs. Use of fewer distorted markers in all the RILs decreases the impact of distorted marker on linkage map construction (Zhu et al. 2007; Xie et al. 2014), and the quality of the linkage maps increases. Second, most distorted markers are clustered. Once these distorted markers are excluded from the construction of linkage maps, a marker-gap in the linkage map exists, for example chromosome 17. In addition, distorted markers increase the consistency of linkage maps with genome and genome coverage. As described in Table S2 in RC RILs, the Spearman correlation coefficient from chromosome 2 by including distorted markers (0.8737) was higher than that for excluding distorted markers (0.8365).

Conclusion

A total of 11,846 SLAF markers in all the RILs were mapped onto 20 chromosomes with an average marker-interval of 0.21 cM. A total of 1513 (12.77%) very significantly distorted markers increased the accuracy of grouping markers on their corresponding chromosomes, the consistency of linkage maps with genome and genome coverage. Less very significantly distorted markers in all the RILs than in OC and RC RILs improved the quality of linkage maps. The relatively high collinearity of linkage maps with the reference genome, and previously reported QTLs and lipid metabolism related genes around 24 seed linoleic acid content QTLs detected in this study validated the quality of the linkage maps.

Acknowledgements We thank Prof. Jim M. Dunwell at University of Reading for help with improvements to the English text. This work was supported by the National Natural Science Foundation of China (31571268, U1602261, 31871242, 31601209), Huazhong Agricultural University Scientific & Technological Self-Innovation Foundation (2014RC020), and State Key Laboratory of Cotton Biology Open Fund (CB2019B01).

Author contributions YMZ conceived and designed the study; JFZ, YN, SFH, PC performed the experiments; JFZ, PC, JYF, YHZ performed the analyses and wrote the draft; YMZ, YW, GS revised the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as 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. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- 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
- Cao Y, Li S, Wang Z, Chang F, Kong J, Gai J, Zhao T (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
- Chen W, Salari H, Taylor MC, Jost R, Berkowitz O, Barrow R et al. (2018) NMT1 and NMT3 N-methyltransferase activity is critical to lipid homeostasis, morphogenesis, and reproduction. *Plant Physiol* 177(4):1605–1628
- Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV et al. (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. *Genetics* 176:685–696
- Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos Trans R Soc Lond B Biol Sci* 363:557–572
- Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL et al. (1999) An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464–1490
- Dahlqvist A, Ståhl U, Lenman M, Banas A, Lee M, Sandager L et al. (2000) Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc Natl Acad Sci USA* 97:6487–6492
- Diers BW, Shoemaker RC (1992) Restriction fragment length polymorphism analysis of soybean fatty acid content. *J Am Oil Chem Soc* 69:1242–1244
- Doyle J (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Eskandari M, Cober ER, Rajcan I (2013) Using the candidate gene approach for detecting genes underlying seed oil concentration and yield in soybean. *Theor Appl Genet* 126(7):1839–1850
- Fan S, Li B, Yu F, Han F, Yan S, Wang L 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
- Gardner KA, Wittern LM, Mackay IJ (2016) A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. *Plant Biotechnol J* 14(6):1406–1417
- Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* 90:33–38
- Hyten DL, Choi IY, Song QJ, Specht JE, Carter TE, Shoemaker RC 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
- Jin W, Palmer RG, Horner HT, Shoemaker RC (1998) Molecular mapping of a male-sterile gene in soybean. *Crop Sci* 38(6):1681–1685
- Keim P, Diers BW, Olson TC, Shoemaker RC (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. *Genetics* 126:735–742
- Keim P, Schupp JM, Travis SE, Clayton K, Zhu T, Shi L et al. (1997) A high-density soybean genetic map based on AFLP markers. *Crop Sci* 37:537–543
- Kelly AA, Froehlich JE, Dörmann P (2003) Disruption of the two digalactosyldiacylglycerol synthase genes *DGDI* and *DGD2* in Arabidopsis reveals the existence of an additional enzyme of galactolipid synthesis. *Plant Cell* 15(11):2694–2706
- Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12:656–664
- Kim HK, Kim YC, Kim ST, Son BG, Choi YW, Kang JS 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
- Kuppusamy T, Giavalisco P, Arvidsson S, Sulpice R, Stitt M, Finnegan PM et al. (2014) Lipid biosynthesis and protein concentration respond uniquely to phosphate supply during leaf development in highly phosphorus-efficient *Hakea prostrata*. *Plant Physiol* 166(4):1891–1911
- Lander ES, Kruglyak L (1995) Genetic dissection of complex traits guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241–247
- Lee SB, Jung SJ, Go YS, Kim HU, Kim JK, Cho HJ et al. (2010) Two Arabidopsis 3-ketoacyl CoA synthase genes, *KCS20* and *KCS2/DAISY*, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. *Plant J* 60:462–475
- Li B, Fan S, Yu F, Chen Y, Zhang S, Han F 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
- Li B, Tian L, Zhang J, Huang L, Han F, Yan S et al. (2014) Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in *Glycine max*. *BMC Genom* 15:1086
- Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics* 24:713–714
- Lyttle TW (1991) Segregation distortion. *Ann Rev Genet* 25:511–557
- McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG et al. (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Mol Biol* 35:89–99
- McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q et al. (2009) Genetic properties of the maize nested association mapping population. *Science* 325(5941):737–740
- Millar AA, Smith MA, Kunst L (2000) All fatty acids are not equal: discrimination in plant membrane lipids. *Trends Plant Sci* 5(3):95–101
- Misra A, Khan K, Niranjana A, Kumar V, Sane VA (2017) Heterologous expression of two GPATs from *Jatropha curcas* alters seed oil levels in transgenic *Arabidopsis thaliana*. *Plant Sci* 263:79–88
- Misra N, Panda PK, Parida BK (2014) Genome-wide identification and evolutionary analysis of algal LPAT genes involved in TAG biosynthesis using bioinformatic approaches. *Mol Biol Rep* 41:8319–8332
- Ozseyhan ME, Li P, Na G, Li Z, Wang C, Lu C (2018) Improved fatty acid profiles in seeds of *Camelina sativa* by artificial microRNA mediated *FATB* gene suppression. *Biochem Biophys Res Commun* 503:621–624
- Palmer RG, Kilen TC (1987) Qualitative genetics and cytogenetics. *Soybeans Improv Prod Uses* 16:135–209

- Qi Z, Huang L, Zhu R, Xin D, Liu C, Han X et al. (2014) A high-density genetic map for soybean based on specific length amplified fragment sequencing. *PLoS ONE* 9:e104871
- Radmark O, Werz O, Steinhilber D, Samuelsson B (2007) 5-Lipoxygenase: regulation of expression and enzyme activity. *Trends Biochem Sci* 32:332–341
- Salas JJ, Ohlrogge JB (2002) Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. *Arch Biochem Biophys* 403:25–34
- Salminen TA, Blomqvist K, Edqvist J (2016) Lipid transfer proteins: classification, nomenclature, structure, and function. *Planta* 244 (5):971–997
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X et al. (2004) A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 109:122–128
- Song Q, Jenkins J, Jia G, Hyten DL, Pantalone V, Jackson SA et al. (2016) Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. *BMC Genom* 17:33
- Sun X, Liu D, Zhang X, Li W, Liu H, Hong W et al. (2013) SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE* 8: e58700
- Todd J, Post-Beittenmiller D, Jaworski JG (1999) *KCS1* encodes a fatty acid elongase 3-ketoacyl-CoA synthase affecting wax biosynthesis in *Arabidopsis thaliana*. *Plant J* 17:119–130
- Wang CM, Zhu CS, Zhai HQ, Wan JM (2005) Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (*Oryza sativa* L.). *Genet Res* 86:97–106
- Wang S, Basten CJ, and Zeng ZB (2012) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC (<http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>).
- Wang SB, Wen YJ, Ren WL, Ni YL, Zhang J, Feng JY et al. (2016) Mapping small-effect and linked quantitative trait loci for complex traits in backcross or DH populations via a multi-locus GWAS methodology. *Sci Rep* 6:29951
- Wang Y, Wu H, Yang M (2008) Microscopy and bioinformatic analyses of lipid metabolism implicate a sporophytic signaling network supporting pollen development in *Arabidopsis*. *Mol Plant* 1(4):667–674
- Wu XL, He CY, Wang YJ, Zhang ZY, Dongfang Y, Zhang JS et al. (2001) Construction and analysis of a genetic linkage map of soybean. *J Genet Genom* 28:1051–1061
- Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet* 4:e1000212
- Xia Z, Tsubokura Y, Hoshi M, Hanawa M, Yano C, Okamura K et al. (2007) An integrated high-density linkage map of soybean with RFLP, SSR, STS, and AFLP markers using a single F₂ population. *DNA Res* 14:257–269
- Xie SQ, Feng JY, Zhang YM (2014) Linkage group correction using epistatic distorted markers in F₂ and backcross populations. *Heredity* 112:479–488
- Yamanaka N, Ninomiya S, Hoshi M, Tsubokura Y, Yano M, Nagamura Y et al. (2001) An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. *DNA Res* 8:61–72
- Yang Y, Speth BD, Boonyoo N, Baumert E, Atkinson TR, Palmer RG et al. (2014) Molecular mapping of three male-sterile, female-fertile mutants and generation of a comprehensive map of all known male sterility genes in soybean. *Genome* 57(3):155–160
- Zhang Y, Wang L, Xin H, Li D, Ma C, Ding X et al. (2013) Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. *BMC Plant Biol* 13:141
- Zhang L, Wang SB, Li QG, Song J, Hao YQ, Zhou L et al. (2016) An integrated Bioinformatics analysis reveals divergent evolutionary pattern of oil biosynthesis in high- and low-oil plants. *PLoS ONE* 11(5):e0154882
- Zhou L, Luo L, Zuo JF, Yang L, Zhang L, Guang X et al. (2016) Identification and validation of candidate genes associated with domesticated and improved traits in soybean. *The Plant Genome* 9(2) <https://doi.org/10.3835/plantgenome2015.09.0090>.
- Zhu C, Wang C, Zhang YM (2007) Modeling segregation distortion for viability selection I. Reconstruction of linkage maps with distorted markers. *Theor Appl Genet* 114:295–305