

RESEARCH ARTICLE

SNP-SNP Interaction Analysis on Soybean Oil Content under Multi-Environments

Qingshan Chen¹*, Xinrui Mao¹*, Zhanguo Zhang¹*, Rongsheng Zhu¹*, Zhengong Yin^{1,3}, Yue Leng¹, Hongxiao Yu¹, Huiying Jia¹, Shanshan Jiang¹, Zhongqiu Ni¹, Hongwei Jiang², Xue Han², Chunyan Liu², Zhenbang Hu¹, Xiaoxia Wu¹, Guohua Hu², Dawei Xin^{1*}, Zhaoming Qi^{1*}

1 College of Agriculture, Soybean biology Key Laboratory of the Ministry of Education, Northeast Agricultural University, Harbin, 150030, Heilongjiang, People's Republic of China, **2** The Crop Research and Breeding Center of Land-Reclamation of Heilongjiang Province, Harbin, 150090, Heilongjiang, People's Republic of China, **3** Crop Breeding Institute, Heilongjiang Academy of Agricultural Sciences, Harbin, 150086, Heilongjiang, People's Republic of China

* These authors contributed equally to this work.

* xindw@126.com (DX); qizhaoming1860@126.com (ZQ)



CrossMark
click for updates

OPEN ACCESS

Citation: Chen Q, Mao X, Zhang Z, Zhu R, Yin Z, Leng Y, et al. (2016) SNP-SNP Interaction Analysis on Soybean Oil Content under Multi-Environments. PLoS ONE 11(9): e0163692. doi:10.1371/journal.pone.0163692

Editor: Yongshuai Jiang, Harbin Medical University, CHINA

Received: April 8, 2016

Accepted: September 13, 2016

Published: September 26, 2016

Copyright: © 2016 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All available data are attached as a supplement file.

Funding: Funded by the national natural science foundation of China <http://www.nsf.gov.cn/>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Soybean oil content is one of main quality traits. In this study, we used the multifactor dimensionality reduction (MDR) method and a soybean high-density genetic map including 5,308 markers to identify stable single nucleotide polymorphism (SNP)—SNP interactions controlling oil content in soybean across 23 environments. In total, 36,442,756 SNP-SNP interaction pairs were detected, 1865 of all interaction pairs associated with soybean oil content were identified under multiple environments by the Bonferroni correction with $p < 3.55 \times 10^{-11}$. Two and 1863 SNP-SNP interaction pairs detected stable across 12 and 11 environments, respectively, which account around 50% of total environments. Epistasis values and contribution rates of stable interaction (the SNP interaction pairs were detected in more than 2 environments) pairs were detected by the two way ANOVA test, the available interaction pairs were ranged 0.01 to 0.89 and from 0.01 to 0.85, respectively. Some of one side of the interaction pairs were identified with previously research as a major QTL without epistasis effects. The results of this study provide insights into the genetic architecture of soybean oil content and can serve as a basis for marker-assisted selection breeding.

Introduction

Soybean originates from China which is not only grain crop but also commercial crop. The content of protein accounts for 40% of soybean seeds, and the content of oil occupy 20%. Soybean is the largest oil seed crop in the world, which accounts for 57% of world oilseed production in 2015 (<http://www.soystats.com>) [1]. Oil content of soybean seeds is a complex quantitative trait governed by a number of genes mostly with small effects and was affected by the environment [2]. Till now, 188 QTL mapping research on soybean oil content were listed

in the USDA Soybean Genome Database (<http://www.soybase.org>). Epistasis interaction is common and can cause cryptic genetic variation for quantitative traits [3]. Deriving genetic interaction networks from epistatic interactions between loci will enhance our understanding of biological systems that give rise to variation in quantitative traits [4]. Many recent studies have been performed regarding to epistatic interaction analysis, including investigations of agronomic and nitrogen acquisition traits in rice [5–9], and protein content, seed and pot traits, grain weight per plant in soybean [10–12], and proteins, oils, starches and fatty acids in maize [13–14], and grain protein, yield, and plant heights traits in wheat [15–17]. These studies, however, were based on the interactions of QTLs, which encompass more than a single marker. However, quantitative variation in phenotypes must result in multiple factor, therefore, interaction study designs have concentrated on estimating additive effects of single loci [18]. Because analyses of single loci can only partially investigate significant loci and may miss interactions between them, interaction analysis of single nucleotide polymorphisms (SNPs) is a better approach to elucidate genetic mechanisms [19], Goodman et al. [20] has explored the association between colon cancer and 94 SNPs, by the method, polymorphism interaction analysis in 2006. Dinu et al. [21] had discovered the Crohn's disease genetics' SNP-SNP interactions by logic regression. Lin et al. [22] studied five SNP-SNP interactions in three gene pairs associated with aggressive prostate cancer by SNP-SNP interaction network. Farzan et al. [23] found that RNASEL and MIR146A SNP—SNP interaction is a susceptibility factor for skin cancer. Li et al. [24] analyzed SNP—SNP interactions affecting abdominal fat weight by a linear mixed model.

Multifactor Dimensionality Reduction (MDR), inspired by the combinatorial partitioning method, can effectively reduce genotype predictors from n dimensions to one [25–29]. 2001, Ritchie et al. invented Multifactor dimensionality reduction which examines all possible SNP combination from a set of given SNPs and choose the combination that the best predicts risk by minimizing the classification error of cases and controls [26]. In addition, MDR analysis incorporates a cross-validation/permutation procedure to minimize the rate of false positive findings that may otherwise result from tests involving multiple variables or comparisons [25–27]. Although some recent studies on SNP interactions have exploited the MDR method [25–31], most of them has been focused on human and animal complex diseases and related traits. To date, no investigations of SNP interactions involving soybean-related traits have been conducted. This study is on the strength of a high-density genetic map for soybean which is based on specific length amplified fragment sequencing. We set the experiment in 23 environments in order to explore the stable locus across the multiple environments. MDR method was employed to analyze epistatic interactions between SNPs for oil content in soybean, the results on SNP interactions underlying genetic mechanisms of quantitative traits and may help improve soybean quality traits.

Materials and Method

Plant Materials

An $F_{2:10}$ – $F_{2:20}$ population of 147 RILs was advanced by single-seed descent from crosses between two soybean cultivars: 'Charleston' (♀), (American cultivator) [32], and 'Dongnong594' (♂), (Northeast Agricultural University, Harbin, Heilongjiang, China). The 147 RILs were planted in three locations: Harbin (HRB; longitude 126°38'E, latitude 45°45'N) from 2002 to 2012; Hongxinglong (HXL; longitude 129°55'–134°35'E, latitude 45°35'–47°17'N) from 2007 to 2012; and Jiamusi (JMS; longitude 130°21'E, latitude 46°49'N) from 2007 to 2012.

The plants were arranged in a randomized complete block experimental design involving single-row plots (1-m length and 0.5-m width), with two and three replicates in 2002–2007

and 2008–2012, respectively. The timing and frequency of cultural management procedures used for the trials were in accordance with normal production practices for the respective environments.

Measurement of Soybean Oil Content

Oil was analyzed from seeds obtained from five randomly harvested plants of each line per plot. The seed oil phenotypic data was analyzed using the FOSS Infratec TM1241 Grain Analyzer (FOSS Tecator AB, Hoeganaes, Sweden). The oil content of each RIL line was averaged from the data of two or three plot replicates. Two hundreds grams seeds of each replicate were used to analysis at a 13% moisture basis.

Genotyping and Genetic Map Construction

SNP genotype data for the RIL population was detected based on the specific-locus amplified fragment sequencing (SLAF-seq) method following Qi et al. [33] and used to construct a high-density genetic map of 5,308 markers assigned to 20 linkage groups. The map spanned 2294.433 cM, with an average inter-marker distance of 0.43 cM. The percentage of gaps in which the distance between adjacent markers and markers is smaller than 5 cM ($\text{gap} \leq 5$) is 99.84%.

Interaction Analysis

To identify SNP \times SNP effects in this study, we used the MDR method [25–31], The Pearson chi-squared module was used to assess significance ($p < 0.001$), with the selected optimization model based on the maximum Pearson chi-squared statistic combined with cross-validation [30].

The chi-square statistic measures the association between genotype (high-risk and low-risk group) and affection status (case and control group) in a two-way table. It is calculated as sum of the square of the difference between the observed and expected frequency in each combination, divided by the expected value, across all the combinations:

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

The calculation was based on the value of chi-square. Average Pearson chi-square of 10 time 10-fold cross-validation were used to select optimization model [30]. The p value was calculated referred the Bonferroni correction [34–35]. The epistatic interaction values and their contribution to genetic values and variance were calculated following the method proposed by Cheverud et al. [36]. Using contribution rate represent the phenotypic variation of the SNP—SNP interaction pair.

Result

Phenotypic Variation in the RIL Population under Multiple Environments

Oil content datas of the RIL population in 23 environments are shown in Table 1. The differences between the two parents were significant in parts of the 23 environments and the oil content of Charleston and Dongnong 594 ranged from 19.03 to 21.89% and from 18.96 to 21.83%, respectively. Oil content displayed a wide range of variation across the 23 environments, but the range of variation was narrow between different locations in the same year. The mean values of the RIL population ranged from 19.07 to 20.99%. The coefficient of variation ranged

Table 1. Oil content of the population and parents in 23 environments across 11 year.

Year and Environment	Charleston (%)	DongNong594 (%)	RIL lines					
			Mean (%)	Min (%)	Max (%)	SD	Kurtosis	Skewness
2002HRB	19.63	21.83	21.48	16.68	24.33	1.39	0.25	-0.36
2003HRB	19.27	19.68	20.46	17.90	22.61	0.91	-0.02	-0.08
2004HRB	19.03	19.26	19.28	17.87	20.92	0.60	-0.20	0.17
2005HRB	19.11	19.44	19.33	16.99	22.17	1.08	-0.30	0.37
2006HRB	20.30	21.20	20.86	19.05	21.90	0.50	0.46	-0.52
2007HRB	20.97	21.26	21.46	19.49	23.35	0.81	-0.10	0.07
2007HXL	21.09	21.63	22.40	18.99	25.88	1.36	-0.12	-0.01
2007JMS	20.65	20.98	20.99	18.64	22.76	0.77	0.44	-0.45
2008HRB	21.89	20.31	21.67	18.64	23.57	0.89	0.28	-0.43
2008HXL	21.80	19.89	21.02	18.64	23.59	0.82	0.83	0.58
2008JMS	21.77	20.73	22.19	19.38	24.13	0.90	0.67	-0.59
2009HRB	21.36	20.26	20.68	18.32	21.79	0.61	1.96	-1.15
2009HXL	21.24	18.96	20.42	17.85	22.28	0.67	1.37	-0.92
2009JMS	20.93	20.07	20.92	19.01	21.98	0.50	1.57	-0.59
2010HRB	21.21	21.42	21.37	19.73	22.03	0.41	2.63	1.27
2010HXL	21.73	21.42	21.56	20.14	22.36	0.39	1.27	0.69
2010JMS	21.51	21.49	21.49	19.09	23.20	0.48	4.64	0.80
2011HRB	20.56	21.14	21.33	20.01	22.51	0.40	0.55	-0.26
2011HXL	19.98	21.40	21.21	20.17	22.34	0.43	-0.22	-0.26
2011JMS	19.78	20.99	21.41	20.57	22.55	0.36	1.12	0.63
2012HRB	21.29	21.49	20.56	18.02	22.74	0.86	2.09	-0.42
2012HXL	20.73	21.16	19.07	15.51	20.86	0.43	1.85	-0.86
2012JMS	19.58	19.70	21.02	18.58	22.45	0.36	2.51	-0.66

doi:10.1371/journal.pone.0163692.t001

from 0.01 to 0.06, with a standard deviation of 0.36 to 1.39. Overall, the oil content of the RIL population was continuously distributed and consistent in all environments (Table 1).

Pairwise Interaction Analysis

Analyze the oil content and genotype of 23 environments by the method of MDR. The selection criteria of SNP interaction pairs was the $p < 3.55 \times 10^{-11}$ which was calculated based on the Bonferroni correction (Table 2). In total, the available interaction pairs have been detected in 12 of all environments, 36,442,756 interaction pairs were detected in all of the 12 environments. There were 5,612,873 pairs detected in 2006HRB, which were the most SNP interaction pairs. And the minimum number 24,689 pairs have been found in 2007HXL. 2005HRB, 2006HRB, 2007HRB, 2007JMS, 2008HRB, 2008HXL, 2009HXL, 2011JMS and 2012HXL were identified more than 1,000,000 SNP interaction pairs, 2004HRB, 2007HXL and 2008JMS were identified fewer than 100,000, and there were no available interaction pairs detected in the other 11 environments (Table 2).

There were two and 1863 SNP interaction pairs were identified under 12 and 11 environments which have been detected the available interaction pairs (Fig 1). Among the 20 linkage groups, most of stable interaction pairs were from Gm01 to others, including Gm01, Gm05, Gm06, Gm07, Gm08, Gm09, Gm10, Gm13, Gm18 and Gm20. Two SNP interaction pairs detected across 12 environments showed the interaction effect between Gm01 (Mark673223) and Gm05 (Mark310931, Mark312588), furthermore, 1863 SNP interaction pairs detected across 11 environments showed the strong interaction region at Gm01, Gm05, Gm06, Gm07,

Table 2. Interaction pairs numbers selected by MDR method in all environment ($P < 3.55 \times 10^{-11}$).

Environment	The number of interaction pairs
2002HRB	NA ^a
2003HRB	NA ^a
2004HRB	87027
2005HRB	1098849
2006HRB	5612873
2007HRB	3391676
2007HXL	24689
2007JMS	4059216
2008HRB	2133998
2008HXL	2731592
2008JMS	60238
2009HRB	NA ^a
2009HXL	7432164
2009JMS	NA ^a
2010HRB	NA ^a
2010HXL	NA ^a
2010JMS	NA ^a
2011HRB	NA ^a
2011HXL	NA ^a
2011JMS	4113031
2012HRB	NA ^a
2012HXL	5697403
2012JMS	NA ^a
Total	36442756

^aNA means no available interaction pairs in that environment.

doi:10.1371/journal.pone.0163692.t002

Gm09, Gm10, Gm13 and Gm18 (Fig 1). These 1865 SNP interaction pairs were detected stable across environments which almost 50% of all, so we saw them as the stable interaction pairs and used for next step calculation.

Epistatic Value and Contribution Rate Analysis

The epistatic interaction values and their contribution to genetic values and variance were screened by the two way ANOVA test for each environment at significant level of $p < 0.01$. In total, there were 560 of 1865 interaction pairs were screened, 4, 23, 99 and 344 SNP interaction pairs were detected as stable and effective interaction pairs across 5, 4, 3 and 2 environments respectively (Fig 2), which similar as the environment-universal QTLs, defined as QTLs associated with traits of interest in at least one macro-environment (i.e., year-location combination) [37, 38], here, it could be the environment-universal SNP interaction pairs, and there were no significance for the other SNP interaction pairs. Epistatic value and contribution rate data of stable interaction pairs across all environments are shown in S1 Table. Epistasis value was ranged from 0.01 to 0.89 at the $p < 0.01$ level. Contribution rates of epistatic pairs were ranged from 0.01 to 0.85 at the $p < 0.01$ level (S1 Table).

A comparison of the environmentally stable SNP interaction pairs uncovered in this study with those detected in previous research revealed none that matched completely. In a few cases, however, a member of an interaction pair matched a main effect QTL not previously

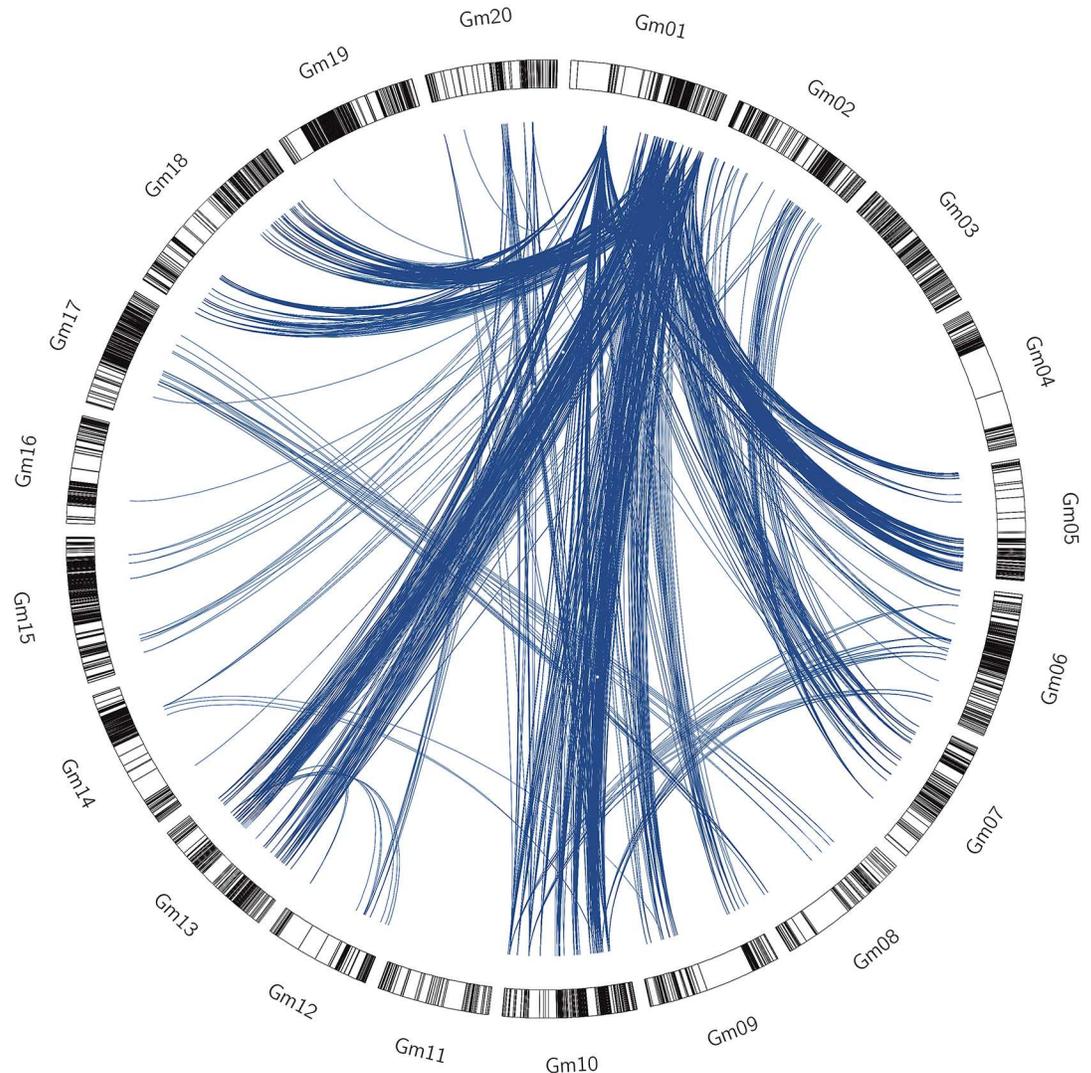


Fig 1. The circle represented the whole soybean genome. Blue lines represented epistatic interactions between two markers across eleven environment, the black lines represented epistatic interactions between two markers across twelve environment.

doi:10.1371/journal.pone.0163692.g001

known to participate in an interaction. Some regions on Gm01 15.0Mb-15.3Mb, 28.8Mb-38.0Mb have been mapped as major QTL fragments by Hyten et al. [39], from 40.7Mb to 41.0Mb has been detected by Specht et al. [40], 42.0Mb-43.3Mb has also been found by Eskandari et al. [41]. On Gm02 (34.6–34.8Mb) has been mapped as major QTL fragments by Qi et al. [42]. Some regions on Gm05, 1.3–1.6Mb, 27.6–27.8Mb, 31.2–32.0Mb, 33.2–34.8Mb and 37.6–39.5Mb all have been detected by Wang et al. [43], Mansur et al. [44], Lark et al. [45], Brummer et al. [46] and Liang et al. [47] respectively. Some regions on Gm06 from 1.2Mb to 1.5Mb, 14.4Mb, 21.2–29.0Mb, 24.7Mb, 37.8Mb and 46.7Mb have been mapped as major QTL fragments by Wang et al. [43], Hyten et al. [39], Chen et al. [48], Reinprecht et al. [49], Kim et al. [50], Qi et al. [42] respectively. On Gm09, 1.0Mb, 3.9Mb and 5.7Mb have been detected by Mansur et al. [44]. Wang et al. [43] have found from 8.2 to 9.1Mb and 41.1–42.8Mb as major QTL fragments. From 31.0Mb to 33.4Mb has been detected by Eskandari et al. [41]. Some regions on Gm10 from 10.2Mb to 13.7Mb and 46.2Mb have been detected by Panthee et al.

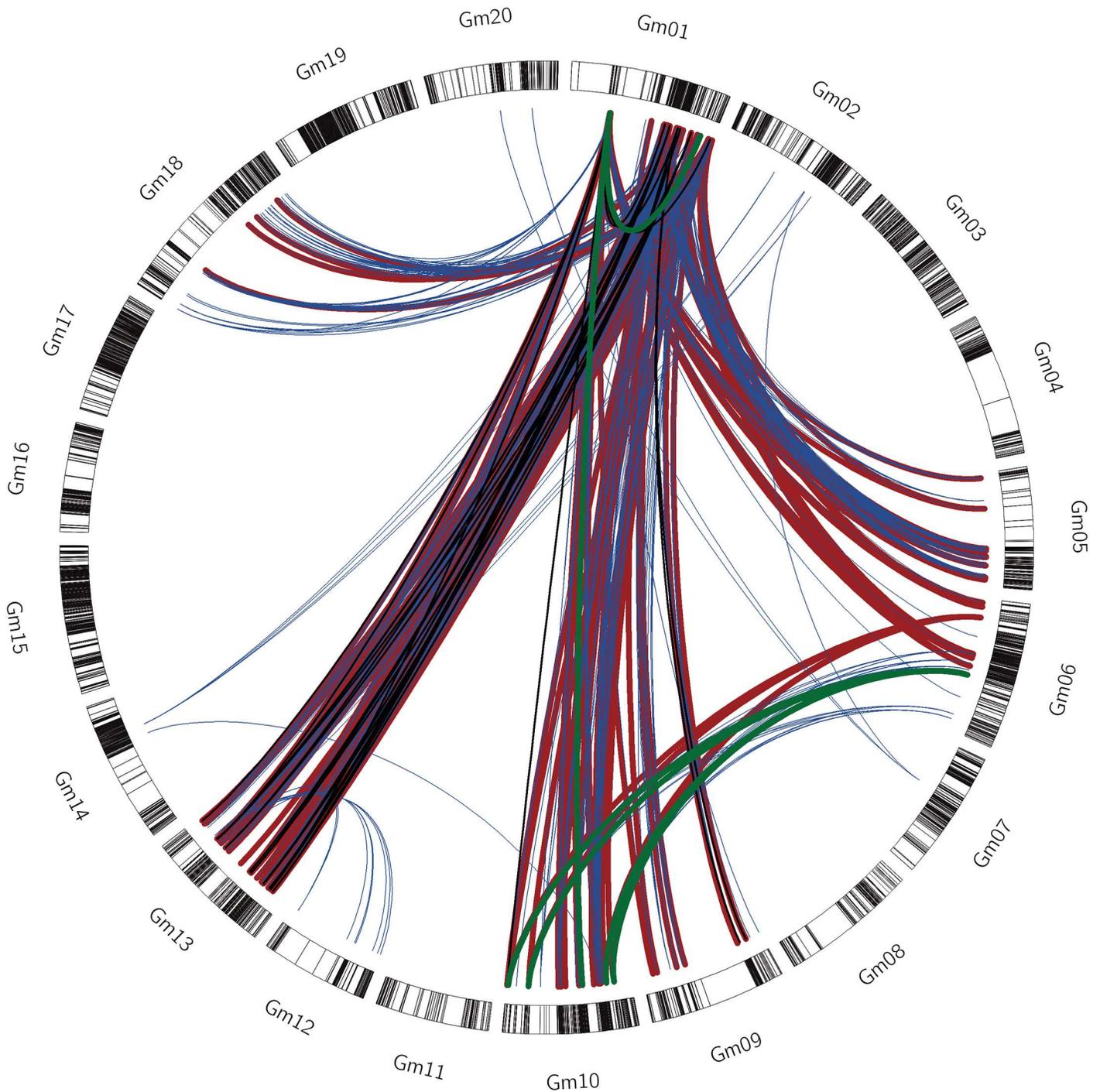


Fig 2. The circle represented the whole soybean genome. Blue lines indicate epistatic interactions across 2 environments, red lines indicate epistatic interactions across 3 environments, black lines indicate epistatic interactions across 4 environments, green lines indicate epistatic interactions across 5 environments.

doi:10.1371/journal.pone.0163692.g002

[51] and Li et al. [52] respectively. The region on Gm12 10.8–14.0Mb and 35.0Mb have been found by Shibata et al. [53] and Eskandari et al. [41] respectively. On Gm13, 6.1–6.7Mb has been mapped as major QTL fragments by Qi et al. [42]; 27.1–28.3Mb has been detected by Specht et al. [40], 29.0–30.7Mb and 32.2–33.6Mb have been found by Eskandari et al. [41] and

31.3–31.5Mb has been found by Rossi et al. [54]. On Gm14 Eskandari et al. [41] and Liang et al. [47] have detected 32.2Mb and 35.9Mb. Some regions on Gm18, 16.7–18.3Mb, 41.66Mb, 46.2–57.6Mb, 50.0–52.7Mb and have been mapped as major QTL fragments by Reinprecht et al. [49], Wang et al. [43], Lee et al. [55] and Brummer et al. [46]. On Gm20, 24.6Mb and 36.5Mb have been detected by Shibata et al. [53] and Reinprecht et al. [49] respectively.

Discussion

Soybean oil content is an important quantitative trait under complex genetic control. Understanding the oil content locus interaction mechanisms could help us effectively increase the oil content in soybean. In this study, MDR method was employed to identify the stable loci controlling oil content in soybean that is based on a high-density genetic map, which has been constructed by specific-locus amplified fragment sequencing (SLAF-seq) with 5308 markers on 20 linkage groups and the map was 2294.43 cM in length, with an average distance of 0.43 cM between adjacent markers [33].

Epistasis is a biochemically plausible feature of the genetic architecture of quantitative traits. So, epistasis causes hidden quantitative genetic variation in natural populations and could be responsible for the small additive effects [4]. In quantitative genetics, epistasis refers to any statistical interaction between genotypes at two or more loci [56–57]. The role of epistasis in the genetic architecture of quantitative traits has been controversial since early formulations of quantitative genetic theory [58–59]. Numerous methods to detect epistatic interactions between SNPs have been published, such as exhaustive algorithms [60] and MDR [25–33], regression [61–63], heuristic [64–65] and mutual information [66], among others [60, 67–70].

Currently, many studies are based on multifactor dimensionality reduction (MDR) focus on the complex human diseases and animal-related traits. Guia et al. [71] reveals interaction of important gene variants involved in allergy by MDR. Su et al. [72] applied the multifactor dimension reduction method to analyze gene-gene and gene-environmental interactions of childhood asthma. Gui et al. [73] detected gene-gene interactions with application to the genetic analysis of bladder cancer susceptibility with multifactor dimensionality reduction method. Cho et al. [74] using MDR method find out that a two-locus interaction genes among 23 loci in the candidate genes of Type 2 diabetes. Julia et al. [75] suggest that 13 genes is associated with synovial fibroblast response to rheumatoid arthritis proinflammatory stimulus by MDR. Many studies using the MDR method have focused on complex human and animal disease-related traits to identify the epistatic interactions [25–31, 71–75], to our knowledge, no study have applied MDR to analyze soybean quantitative traits. In this study, we identified 1865 stable SNP interactions pairs associations with soybean oil content under multi-environments, successfully applied MDR to the plant research. Among 1865 SNP interactions pairs, 2 pairs were detected stable across 12 environments, 1863 pairs were detected stable across 11 environments. MDR method can improve the analysis efficiency as an analysis method based on haplotype, which is a good choice for genome-wide interaction analysis. Nevertheless, this method also has some restrictions. First, it does not have a way to adjust for covariate effects such as age, gender and smoking status, an often necessary step to obtain an unconfounded SNP interaction outcome association [76]. Second, if there are too many covariates, this approach may over fit the data and sometimes fail due to limited sample size [77].

The present study has opened the door for further research on soybean quantitative trait integration analysis. On the one hand, the SNP-SNP interaction of hot region which were identical with some published major QTLs, however, those major QTLs were not shown the interaction each other previously. On the other hand, the candidate major region could be found as only one site of the pair was aligned with the published QTL. Based on this research, the

markers of those hot region could be developed, and used for breeding selection process, furthermore, the candidate genes of hot region could be identified and to analysis the interaction of genes in next step. Future efforts based on the uncovered epistatic interaction values and their contribution to genetic values can be applied for the identification of key loci with higher phenotypic contributions, the prediction of phenotype and genotype design modules with epistatic values, and to provide direction for high-quality soybean breeding.

Supporting Information

S1 Table. Epistatic value and contribution rate data of stable interaction pairs across 11 environments. NS means no signification value in that environments, epistatic value indicates the epistatic effects value of the SNP—SNP pairs, contribution rates indicates the phenotypic variation of the SNP—SNP interaction pairs. (XLSX)

Acknowledgments

This study was conducted in the laboratory of the Crop Research and Breeding Center of Land-Reclamation in Heilongjiang Province and the Soybean Key Laboratory of the Ministry of Education, Northeast Agricultural University. We are also thankful for support provided by the Postdoctoral Research Institute of Northeast Agricultural University.

Author Contributions

Conceptualization: RZ QC.

Data curation: XM DX XH YL HY HJ SJ ZN.

Formal analysis: ZZ.

Funding acquisition: QC ZQ DX HJ YL SJ ZN CL.

Investigation: XM ZY.

Methodology: RZ ZQ.

Project administration: QC ZQ GH.

Software: RZ ZZ.

Supervision: QC ZQ XW.

Validation: QC ZQ.

Visualization: HJ CL ZH.

Writing – original draft: XM.

Writing – review & editing: ZQ.

References

1. Wilson RF. Soybean: market driven research needs. In: Genetics and genomics of soybean. New York: Springer; 2008. pp. 3–15.
2. Burton JW. Quantitative genetics: results relevant to soybean breeding. In: Wilcox JR (ed) Soybeans: improvement, production, and uses, 2nd edn. ASA, CSSA, and SSSA, Madison, 1987. pp. 211–242.
3. Mackay TF. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet.* 2014; 15: 22–23. doi: [10.1038/nrg3627](https://doi.org/10.1038/nrg3627) PMID: [24296533](https://pubmed.ncbi.nlm.nih.gov/24296533/)

4. Phillips PC. Epistasis-the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet.* 2008; 9: 855–867. doi: [10.1038/nrg2452](https://doi.org/10.1038/nrg2452) PMID: [18852697](https://pubmed.ncbi.nlm.nih.gov/18852697/)
5. Liu GF, Yang J, Xu HM, Zhu J. Influence of Epistasis and QTL × Environment Interaction on Heading Date of Rice (*Oryza sativa* L.). *J Genet Genomics.* 2007; 34: 608–615. PMID: [17643946](https://pubmed.ncbi.nlm.nih.gov/17643946/)
6. Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, et al. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA.* 1997; 94: 9226–9231. PMID: [11038567](https://pubmed.ncbi.nlm.nih.gov/11038567/)
7. Xing Z, Tan F, Hua P, Sun L, Xu G, Zhang Q. Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. *Theor Appl Genet.* 2002; 105: 248–257. PMID: [12582526](https://pubmed.ncbi.nlm.nih.gov/12582526/)
8. Zhuang JY, Fan YY, Rao ZM, Wu JL, Xia YW, Zheng KL. Analysis on additive effects and additive-by-additive epistatic effects of QTLs for yield traits in a recombinant inbred line population of rice. *Theor Appl Genet.* 2002; 105: 1137–1145. PMID: [12582891](https://pubmed.ncbi.nlm.nih.gov/12582891/)
9. Senthilvel S, Vinod KK, Malarvizhi P, Maheswaran M. QTL and QTL x environment effects on agronomic and nitrogen acquisition traits in rice. *J Integr Plant Biol.* 2008; 50: 1108–1117. doi: [10.1111/j.1744-7909.2008.00713.x](https://doi.org/10.1111/j.1744-7909.2008.00713.x) PMID: [18844779](https://pubmed.ncbi.nlm.nih.gov/18844779/)
10. Qi ZM, Hou M, Han X, Liu CY, Jiang HW, Xin DW, et al. Identification of quantitative trait loci (QTLs) for seed protein concentration in soybean and analysis for additive effects and epistatic effects of QTLs under multiple environments. *Plant Breeding.* 2014; 133: 499–507.
11. Zhang HJ, Li JW, Liu YJ, Jiang WZ, Du XL, Li L, et al. Quantitative trait loci analysis of individual and total isoflavone contents in soybean seeds. *J Genet.* 2014; 93: 331–338. PMID: [25189227](https://pubmed.ncbi.nlm.nih.gov/25189227/)
12. Han YP, Teng WL, Sun DS, Du YP, Qiu LJ. Impact of epistasis and QTL x environment interaction on the accumulation of seed mass of soybean (*Glycine max* L. Merr.). *Genet Res.* 2008; 90: 481–491.
13. Zhang H, Jin T, Huang Y, Chen J, Zhu L, Zhao Y, et al. Identification of quantitative trait loci underlying the protein, oil and starch contents of maize in multiple environments. *Euphytica.* 2015; 205: 169–183.
14. Yang XH, Guo YQ, Yan JB, Zhang J, Song TM, Rocheford T, et al. Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. *Theor Appl Genet.* 2010; 120: 665–678. doi: [10.1007/s00122-009-1184-1](https://doi.org/10.1007/s00122-009-1184-1) PMID: [19856173](https://pubmed.ncbi.nlm.nih.gov/19856173/)
15. Conti V, Roncallo PF, Beaufort V, Cervigni GL, Miranda R, Jensen CA, et al. Mapping of main and epistatic effect QTLs associated to grain protein and gluten strength using a RIL population of durum wheat. *J Appl Genet.* 2011; 52: 287–398. doi: [10.1007/s13353-011-0045-1](https://doi.org/10.1007/s13353-011-0045-1) PMID: [21523429](https://pubmed.ncbi.nlm.nih.gov/21523429/)
16. Li ZK, Jiang XL, Peng T, Shi CL, Han SX, Tian B, et al. Mapping quantitative trait loci with additive effects and additive x additive epistatic interactions for biomass yield, grain yield, and straw yield using a doubled haploid population of wheat (*Triticum aestivum* L.). *Genet Mol Res.* 2014; 13: 1412–1424. doi: [10.4238/2014.February.28.14](https://doi.org/10.4238/2014.February.28.14) PMID: [24634240](https://pubmed.ncbi.nlm.nih.gov/24634240/)
17. Zhang K, Tian J, Zhao L, Wang S. Mapping QTLs with epistatic effects and QTL x environment interactions for plant height using a doubled haploid population in cultivated wheat. *J Genet Genomics.* 2008; 35: 119–127. doi: [10.1016/S1673-8527\(08\)60017-X](https://doi.org/10.1016/S1673-8527(08)60017-X) PMID: [18407059](https://pubmed.ncbi.nlm.nih.gov/18407059/)
18. Kauffman SA. *The origins of order.* Oxford University Press; 1993.
19. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet.* 2009; 10: 392–404. doi: [10.1038/nrg2579](https://doi.org/10.1038/nrg2579) PMID: [19434077](https://pubmed.ncbi.nlm.nih.gov/19434077/)
20. Goodman JE, Mechanic LE, Luke BT, Ambs S, Chanock S, Harris CC. Exploring SNP-SNP interactions and colon cancer risk using polymorphism interaction analysis. *Int. J. Cancer.* 2006; 118: 1790–1797. PMID: [16217767](https://pubmed.ncbi.nlm.nih.gov/16217767/)
21. Dinu I, Mahasirimongkol S, Liu Q, Yanai H, Sharaf Eldin N, Kreiter E, et al. SNP-SNP Interactions Discovered by Logic Regression Explain Crohn's Disease Genetics. *PLoS ONE.* 2012; 7: e43035. doi: [10.1371/journal.pone.0043035](https://doi.org/10.1371/journal.pone.0043035) PMID: [23071489](https://pubmed.ncbi.nlm.nih.gov/23071489/)
22. Lin HY, Amankwah EK, Tseng TS, Qu X, Chen DT, Park JY. SNP-SNP Interaction Network in Angiogenesis Genes Associated with Prostate Cancer Aggressiveness. *PLoS ONE.* 2013; 8: e59688. doi: [10.1371/journal.pone.0059688](https://doi.org/10.1371/journal.pone.0059688) PMID: [23593148](https://pubmed.ncbi.nlm.nih.gov/23593148/)
23. Farzan SF; Karagas MR; Christensen BC; Li Z; Kuriger JK, Nelson HH. RNASEL and MIR146A SNP-SNP interaction as a susceptibility factor for non-melanoma skin cancer. *PLoS ONE.* 2014; 9: e93602. doi: [10.1371/journal.pone.0093602](https://doi.org/10.1371/journal.pone.0093602) PMID: [24699816](https://pubmed.ncbi.nlm.nih.gov/24699816/)
24. Li FG, Hu G, Zhang H, Wang SZ, Wang ZP, Li H, et al. Epistatic Effects on Abdominal Fat Content in Chickens: Results from a Genome-Wide SNP-SNP Interaction Analysis. *PLoS ONE.* 2013; 8: e81520. doi: [10.1371/journal.pone.0081520](https://doi.org/10.1371/journal.pone.0081520) PMID: [24339942](https://pubmed.ncbi.nlm.nih.gov/24339942/)
25. Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics.* 2003; 19: 376–382. PMID: [12584123](https://pubmed.ncbi.nlm.nih.gov/12584123/)

26. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet.* 2001; 69: 138–147. PMID: [11404819](#)
27. Moore JH. Computational analysis of gene-gene interactions using multifactor dimensionality reduction. *Expert Rev Mol Diagn.* 2004; 4: 795–803. PMID: [15525222](#)
28. Coffey CS, Hebert PR, Ritchie MD, Krumholz HM, Gaziano JM, Ridker PM, et al. An application of conditional logistic regression and multifactor dimensionality reduction for detecting gene-gene interactions on risk of myocardial infarction: the importance of model validation. *BMC Bioinformatics.* 2004; 5: 49. PMID: [15119966](#)
29. Chung Y, Lee SY, Elston RC, Park T. Odds ratio based multifactor-dimensionality reduction method for detecting gene-gene interactions. *Bioinformatics.* 2007; 23: 71–76. PMID: [17092990](#)
30. Jiang Y, Zhang R, Liu G, Wang Z, Chen P, et al. Multifactor dimensionality reduction for detecting haplotype-haplotype interaction. *Fuzzy Systems and Knowledge Discovery, 2009. FSKD'09. Sixth International Conference on. IEEE.* 2009; 1: 241–245.
31. Kuo HC, Chang CC, Guo MM, Hsieh KS, Yeter D, Li SC, et al. Gene-Gene Associations with the Susceptibility of Kawasaki Disease and Coronary Artery Lesions. *PLoS ONE.* 2015; 10: e0143056. doi: [10.1371/journal.pone.0143056](#) PMID: [26619243](#)
32. Cooper RL, Martin RJ, St Martin SK, Calip-DuBois A, Fioritto RJ, Schmitthenner AF (1995). Registration of 'Charleston' soybean. *Crop Sci* 35: 593–593.
33. Qi ZM, Huang L, Zhu RS, Xi DW, Liu CY, Han X, et al. A High-Density Genetic Map for Soybean Based on Specific Length Amplified Fragment Sequencing. *PLoS ONE.* 2014; 9: e104871. doi: [10.1371/journal.pone.0104871](#) PMID: [25118194](#)
34. Audic S, Claverie JM. The significance of digital gene expression profiles. *Genome Res.* 1997; 7: 986–995. PMID: [9331369](#)
35. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* 2001; 29: 1165–1188.
36. Cheverud JM, Routman EJ. Epistasis and its contribution to genetic variance components. *Genetics.* 1995; 139: 1455–1461. PMID: [7768453](#)
37. Palomeque L, Liu LJ, Li W, Hedges BR, Cober ER, Smid MP, et al. Validation of mega-environment universal and specific QTL associated with seed yield and agronomic traits in soybeans. *Theor Appl Genet.* 2010; 120: 997–1003. doi: [10.1007/s00122-009-1227-7](#) PMID: [20012262](#)
38. Mao T, Jiang Z, Han Y, Teng W, Zhao X, Li W, et al. Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments. *Plant Breeding.* 2013; 132: 630–641.
39. Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, et al. Seed quality QTL in a prominent soybean population. *Theor Appl Genet.* 2004; 109: 552–561. PMID: [15221142](#)
40. Deshmukh R, Sonah H, Patil G, Chen W, Prince S, Mutava R, et al. Soybean Response to Water: A QTL Analysis of Drought Tolerance. *Crop Sci.* 2001; 41: 493–509.
41. Eskandari M, Cober ER, Rajcan I. Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents. *Theor Appl Genet.* 2013; 126: 483–495. doi: [10.1007/s00122-012-1995-3](#) PMID: [23192670](#)
42. Qi ZM, Wu Q, Han X, Sun YN, Du XY, Liu CY, et al. Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. *Euphytica.* 2011; 179: 499–514.
43. Wang X, Jiang G, Green M, Scott R, Hyten D, Cregan P. Quantitative trait locus analysis of saturated fatty acids in a population of recombinant inbred lines of soybean. *Mol Breeding.* 2012; 30: 1163–1179.
44. Mansur LM, Orf JH, Chase K, Jarvik T, Cregan PB, Lark KG. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Sci.* 1996; 36: 1327–1336.
45. Lark KG, Orf J, Mansur LM. Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] determined by QTL association with RFLP alleles. *Theor Appl Genet.* 1994; 88: 486–489. doi: [10.1007/BF00223665](#) PMID: [24186039](#)
46. Brummer EC, Graef GL, Orf J, Wilcox JR, Shoemaker RC. Mapping QTL for seed protein and oil content in eight soybean populations. *Crop Sci.* 1997; 37: 370–378.
47. Liang HZ, Yu YL, Wang SF, Lian Y, Wang TF, Wei YL. QTL Mapping of Isoflavone, Oil and Protein Contents in Soybean (*Glycine max* L. Merr.). *Ag. Sci. in China.* 2010; 9: 1108–1116.
48. Chen QS, Zhang ZC, Liu CY, Xin DW, Qiu HM, Shan DP, et al. QTL Analysis of Major Agronomic Traits in Soybean. *Ag. Sci. in China.* 2007; 6: 399–405.

49. Reinprecht Y, Poysa VW, Yu K, Rajcan I, Ablett GR, Pauls KP. Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. *Genome*. 2006; 49: 1510–1527. PMID: [17426766](#)
50. Kim H, Kim Y, Kim S, Son B, Choi Y, Kang J. Analysis of Quantitative Trait Loci (QTLs) for Seed Size and Fatty Acid Composition Using Recombinant Inbred Lines in Soybean. *J. Life Sci.* 2010; 20: 1186–1192.
51. Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE. Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean. *Crop Sci.* 2005; 45:2015–2022.
52. Li HW, Zhao TJ, Wang YF, Yu DY, Chen SY, Zhou RB, et al. Genetic structure composed of additive QTL, epistatic QTL pairs and collective unmapped minor QTL conferring oil content and fatty acid components of soybeans. *Euphytica*. 2011; 182:117–132.
53. Shibata M, Takayama K, Ujiie A, Yamada T, Abe J, Kitamura K. Genetic relationship between lipid content and linolenic acid concentration in soybean seeds. *Breed Sci.* 2008; 58: 361–366.
54. Rossi ME, Orf JH, Liu LJ, Dong Z, Rajcan I. Genetic basis of soybean adaptation to North American vs. Asian mega-environments in two independent populations from Canadian × Chinese crosses. *Theor Appl Genet.* 2013; 126: 1809–23. doi: [10.1007/s00122-013-2094-9](#) PMID: [23595202](#)
55. Lee SH, Bailey MA, Mian MA, Carter TE Jr, Shipe ER, Ashley DA, et al. RFLP loci associated with soybean seed protein and oil content across populations and locations. *Theor Appl Genet.* 1996; 3: 649–657.
56. Cheverud JM, Routman EJ. Epistasis and its contribution to genetic variance components. *Genetics*. 1995; 139: 1455–1466. PMID: [7768453](#)
57. Falconer DS, Mackay TFC. Introduction to quantitative genetics. *Trends Genet.* 1996; 12: 280.
58. Fisher RA. *The Genetical theory of natural selection*. Clarendon: Oxford University Press; 1930.
59. Wright S. Evolution in mendelian populations. *Genetics*. 1931; 16: 97–159. PMID: [17246615](#)
60. Nelson MR, Kardia SL, Ferrell RE, Sing CF. A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. *Genome Res.* 2001; 11: 458–470. PMID: [11230170](#)
61. Wu TT, Chen YF, Hastie T, Sobel E, Lange K. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*. 2009; 25: 714–721. doi: [10.1093/bioinformatics/btp041](#) PMID: [19176549](#)
62. Moore JH, Asselbergs FW, Williams SM. Bioinformatics challenges for genome-wide association studies. *Bioinformatics*. 2010; 26: 445–455. doi: [10.1093/bioinformatics/btp713](#) PMID: [20053841](#)
63. Heidema AG, Boer JM, Nagelkerke N, Mariman EC, Feskens EJ. The challenge for genetic epidemiologists: how to analyze large numbers of SNPs in relation to complex diseases. *BMC Genet.* 2006; 7: 1.
64. Carlborg O, Andersson L, Kinghorn B. The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics*. 2000; 155: 2003–2010. PMID: [10924492](#)
65. Yang C, He Z, Wan X, Yang Q, Xue H. SNPHarvester: a filtering-based approach for detecting epistatic interactions in genome-wide association studies. *Bioinformatics*. 2009; 25: 504–511. doi: [10.1093/bioinformatics/btn652](#) PMID: [19098029](#)
66. Curk T, Rot G, Zupan B. SNPsyn: detection and exploration of SNP-SNP interactions. *Nucleic Acids Res.* 2011; 39: 444–449.
67. Lin HY, Amankwah EK, Tseng TS, Qu X, Chen DT. SNP-SNP interaction network in angiogenesis genes associated with prostate cancer aggressiveness. *PLoS ONE*. 2013; 8: e59688. doi: [10.1371/journal.pone.0059688](#) PMID: [23593148](#)
68. Winham SJ, Colby CL, Freimuth RR, Wang X, Andrade M. SNP interaction detection with Random Forests in high-dimensional genetic data. *BMC Bioinformatics*. 2012; 13: 1.
69. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for wholegenome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81: 559–575. PMID: [17701901](#)
70. Wan X, Yang C, Yang Q, Xue H, Fan X. BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am J Hum Genet.* 2010; 87: 325–340. doi: [10.1016/j.ajhg.2010.07.021](#) PMID: [20817139](#)
71. Guia RM, Echavez MD, Gaw EL, Gomez MR, Lopez KA, Mendoza RC, et al. Multiafactor-dimensionality reduction reveals interaction of important gene variants in allergy. *Int J Immunogenet.* 2015; 42: 182–189. doi: [10.1111/iji.12200](#) PMID: [25876437](#)

72. Su L, Liu G, Wang H, Tian Y, Zhou Z, Han L, et al. Research on Single Nucleotide Polymorphisms Interaction Detection from Network Perspective. *PLoS ONE*. 2015; 10: e0119146. doi: [10.1371/journal.pone.0119146](https://doi.org/10.1371/journal.pone.0119146) PMID: [25763929](https://pubmed.ncbi.nlm.nih.gov/25763929/)
73. Gui J, Andrew AS, Andrews P, Nelson HM, Kelsey KT, Karagas MR, et al. A robust multifactor dimensionality reduction method for detecting gene-gene interactions with application to the genetic analysis of bladder cancer susceptibility. *Ann Hum Genet*. 2011; 75: 20–28. doi: [10.1111/j.1469-1809.2010.00624.x](https://doi.org/10.1111/j.1469-1809.2010.00624.x) PMID: [21091664](https://pubmed.ncbi.nlm.nih.gov/21091664/)
74. Cho YM, Ritchie MD, Moore JH et al. Multifactor-dimensionality reduction shows a two-locus interaction associated with Type 2 diabetes mellitus. *Diabetologia*. 2004; 47: 549. PMID: [14730379](https://pubmed.ncbi.nlm.nih.gov/14730379/)
75. Julia A, Morre J, Miquel L, Alegre C, Barcelo P, Ritchie M, et al. Identification of a two-loci epistatic interaction associated with susceptibility to rheumatoid arthritis through reverse engineering and multifactor dimensionality reduction. *Genomics*. 2007; 90: 6–13. PMID: [17482423](https://pubmed.ncbi.nlm.nih.gov/17482423/)
76. Schulze TG, Zhang K, Chen YS, Akula N, Sun F, McMahon FJ, et al. Defining haplotype blocks and tag single-nucleotide polymorphisms in the human genome. *Hum Mol Genet*. 2004; 13: 335–342. PMID: [14681300](https://pubmed.ncbi.nlm.nih.gov/14681300/)
77. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol*. 1995; 12: 921–927. PMID: [7476138](https://pubmed.ncbi.nlm.nih.gov/7476138/)