QTL mapping of seed oil and seed protein content in two recombinant inbred line soybean populations

Brant McNeece

9/16/2021

# Abstract

The value of soybean (Glycine max L. Merr) is dependent upon its seed oil and protein content. Much research has been performed to pinpoint genomic regions for manipulation of these quantitative and inversely correlated traits through marker assisted selection, yet very few markers have shown to be effective. Therefore, the purpose of this study was to map stable quantitative trait loci (QTL) of two recombinant inbred line (RIL) populations derived from ‘Clermont × PI205805 and ‘Clermont × PI253666A, Pop33 and Pop34, respectively. Populations were genotyped using the BARCSoySNP6K assay. Seven QTL for seed oil content and six QTL for seed protein content were identified. Three of the oil QTLs were collocated with protein QTLs. Two QTLs, qProt-2-1 and qOil-5-1, are first reported here as novel QTLs for their respective traits. This study provides two new QTLs for the study of protein and oil content in soybean, as well as validates other previous findings to aid in advancement of these traits in soybean.

# Introduction

Soybean (Glycine max L. Merr) is one of the most valuable crops in the world. This value is driven by the seed compositional properties of soybean, mainly the seed protein and oil content (Warrington et al. 2015). **MORE ABOUT ECONOMICS** Because of this, many previous studies aimed at identifying quantitative trait loci (QTL) for seed oil and protein have been performed to increase the content of these traits (Patil et al. 2017). However, advancement in production of elite material with higher protein and oil content via conventional breeding techniques has been hindered by the negative relationship between the two compositional traits (Wilcox, 1998). The need for more studies in the area are due to this negative relationship and the difficulty in identifying QTLs that are consistent across broad environments (Patil et al. 2017). The use of genetic markers can allow breeders to obtain a better understanding of the underlying control of the traits by conducting QTL mapping experiments to find markers linked to QTL which consistently influence these traits. These stable QTL which have effects that can be consistently observed across diverse germplasm and environments are valuable to breeders as they present an opportunity for use in marker-assisted selection (MAS) schemes. Effective use of MAS by a breeder requires this stability because unstable QTL are unlikely to produce reliable improvements over conventional breeding techniques if used in a MAS strategy. There have been an exceptional number of QTL reported for seed oil and seed protein content in soybean, to date there have been 315 seed oil and 241 seed protein QTL reported on the soybean data repository SoyBase. These QTL have been mapped over many years using a diverse range of both mapping populations, marker systems, and mapping techniques. The availability of high quality genetic resources such as the soybean consensus linkage map and genome sequence have further helped to compare the findings of these many studies. Despite this progress, few of these QTL have been confirmed which requires that a QTL be detected from a separate set of meiotic events and environments from the environments of the original mapping population, and be detected with an experiment-wise error rate of 0.01 or lower. Among the most often reported of these past confirmed QTL are large effect QTL which may be found on chromosome 20 (LG I) and chromosome 15 (LG E). These QTL have however been found to have variable effects on yield drag where studies which used the Danbaekkong cultivar reported little to no yield drag with improved protein content while other studies reported sizable yield drag. Relatively few QTL exist for seed oil and protein that have been reliably observed across both environments and genetic backgrounds. As such, there is a need to conduct additional mapping studies for these traits which evaluate mapping populations of diverse genetic backgrounds across multiple environments. The results of these studies can then also be compared with the findings of past studies to identify novel QTL, and validate the importance of existing QTL. To accomplish these objectives, we grew two soybean recombinant inbred line (RIL) in multiple environments to assess the magnitude and stability of QTL across both genetic backgrounds and growing environments.

# Materials and Methods

## Plant Material

Material used in this study consisted of two recombinant inbred line (RIL) populations, Pop33 and Pop34. Both populations had the same female parent. Pop33 and Pop34 were derived from ‘Clermont × PI205805 and ‘Clermont × PI253666A, respectively. Populations were advanced by single seed descent (Brim 1966). The number of RILs for each population were 175 and 110 for Pop33 and Pop34, respectively. Plots were grown in Wooster, OH and Clayton, NC with two replications. Pop33 and Pop34 were grown from 2017-2018 and 2018-2019, respectively.

## Phenotypic Data Collection

Plots were rouged for flower color and pubescence prior to harvest. Plots were harvested at maturity with a plot combine. Seed from individual plots were then visually screened and cleaned to remove off types and diseased or split seed. An 80 g sub sample of the cleaned whole seed was analyzed via Perten DA 7250 NIR spectrometer to determine seed protein and oil content on a zero-moisture basis. All protein and oil values are presented as a percent of total seed content.

## Genetic and QTL Mapping

Leaf tissue from F7 plants of the populations was used for genomic DNA extraction with DNAEasy Plant Mini Kit following protocol (Qiagen®). gDNA was diluted to 100 ng mL-1 with nuclease-free water and run on a 2% agarose gel as well as quantified with NanoDrop (Thermo Fisher Scientific, MA) for quality analysis before genotyping. Genotyping was performed via BARCSoySNP6K assay(Song et al. 2020). AsMap package in R was used to filter genotyping data and construction of the linkage map (Taylor and Butler 2017). SNPs were discarded if monomorphic or containing greater than 20% missing data. RILs were discarded if they contained greater than 25% missing data or a duplicated genotype. QTL mapping was performed using the inclusive composite interval mapping (ICIM) method (Meng et al. 2015). LOD (log-likelihood) threshold at P < 0.05 and LOD score ≥ 2.5 were the parameters used to determine QTL presence. QTL results were compared with SoyBase based on the consensus marker interval for comparison to previously identified QTLs (<http://www.soybase.org>). QTLs nomenclature followed that described in (McCouch 1997) Only QTL that were identified in multiple environments were considered stable and reported here.

## Statistical analysis

Values used for seed protein and oil content were determined using SAS PROC GLM where genotype was treated as fixed and environment and replication were treated as random (SAS Institute, 2012). Values are reported for each individual environment (state and year) and all environments within a population (combined). Pearson correlation coefficients of the LSMeans were acquired using the correlation package in R [makowski2020methods]. Summary statistics were produced by QTL IciMapping (Meng et al. 2015).

# Results

## Seed oil and protein content

Seed protein and oil content was determined for 175 and 110 RILs and parents for Pop33 and Pop34, respectively. These values were collected over five environments for each population with two environments in North Carolina, two in Ohio, and a combined analysis spanning two years for each population (Pheno Table). The unique parent in each population had considerably higher protein and lower oil content than the common female parent for both populations (Pheno Table). Pop33 had a protein content range of 37.4 – 53.2% and oil content range of 15.4 – 23.7%. Pop34 protein content ranged from 40.2 – 52.4% and oil ranged from 16.4-23.7%. Coefficients of variation (CV) were higher for oil content compared to protein content in Pop33, while CV for oil content in Pop34 was slightly less than protein content (Pheno Table). Pop33 CV ranges for oil and protein content were 5.72-6.68% and 4.85-5.64%, respectively. Pop34 CV ranges for oil and protein content were 3.80-5.63% and 2.88-4.78%, respectively. Transgressive segregation was observed in both populations for both protein and oil content, but more consistently observed for protein content (Distribution Figure). Pearson correlation coefficients were calculated for each trait and environment for both populations (Corr Table). Protein and oil content values within environments were opposing and significant at P < 0.01 for all environments in both populations. The same negative correlation between protein and oil was significant (p < 0.01) between all environments in Pop33. The correlations of protein and oil were significant and negative in Pop34 between environments occurring in the same state (i.e. NC18 vs. NC19), but were not found significant between environments occurring in different states (NC vs. OH), with few exceptions (Corr Table). This is not surprising given the nature of protein and oil content to be influenced by average daytime temperature (Piper and Boote 1999). The strong agreement of correlations between NC and OH environments in Pop33 suggest that the QTLs mapped for that population would be exceptionally stable.

## QTL Mapping

RILs of both populations Pop33 and Pop34 were genotyped at 5,403 SNPs with the BARCSoySNP6K assay (Song et al. 2020). Marker count used in the final genetic maps were 1134 and 849 for Pop33 and Pop34, respectively. Genetic map of Pop33 had an average marker spacing of 1.5 cM with an average of 56.7 markers per linkage group. Pop34 had an average marker spacing of 3.3 cM with an average of 42.5 markers per linkage group. Pop33 average marker spacing ranged from 2.1 – 14.3 cM on chromosome 18 and 6, respectively. Pop34 average marker spacing ranged from 8.8 – 38.5 cM on chromosome 10 and 6, respectively. QTLs were identified from the phenotypic data and the linkage map using inclusive composite interval mapping (ICIM) method (Meng et al. 2015). Only QTLs identified in multiple environments were considered stable and reported here. QTLs nomenclature followed that proposed in McCouch et al. 1997(McCouch 1997). Seven QTLs were identified for seed oil content (QTL Table). All seven oil QTLs were found on separate chromosomes, qOil-5-1, qOil-6-1, qOil-11-1, qOil-13-1, qOil-15-1, qOil-17-1, and qOil-20-1, on chromosome 5, 6, 11, 13, 15, 17, and 20, respectively. Six QTLs were identified for seed protein content, qProt-2-1, qProt-6-1, qProt-6-2, qProt-11-1, qProt-15-1, and qProt-20-1, on 5 chromosomes 2, 6, 11, 15, and 20 (QTL Table). Three of the oil QTLs (qOil-11-1, qOil-15-1, and qOil-20-1) were collocated with protein QTLs (qProt-11-1, qProt-15-1, and qProt-20-1). qProt-2-1 and qOil-5-1 first identified here have no other previously identified QTLs for their respective trait (<http://www.soybase.org>). The region of chromosome 2, where qProt-2-1 was identified, has one previously reported seed oil QTL (Kim et al., 2010). The region on chromosome 5, where qOil-5-1 was identified has no previously reported QTLs for this trait to date (<http://www.soybase.org>). Strong QTLs (LOD > 10) for protein, qProt-15-1 and qProt-20-1, and oil, qOil-15-1 and qOil-20-1, detected here are known genomic hotspots for protein and oil content ().

# Conclusions

Brim, Charles A. 1966. “A Modified Pedigree Method of Selection in Soybeans 1.” *Crop Science* 6 (2): 220–20.

McCouch, SR. 1997. “Report on QTL Nomenclature.” *Rice Genet. Newsl.* 14: 11–13.

Meng, Lei, Huihui Li, Luyan Zhang, and Jiankang Wang. 2015. “QTL IciMapping: Integrated Software for Genetic Linkage Map Construction and Quantitative Trait Locus Mapping in Biparental Populations.” *The Crop Journal* 3 (3): 269–83.

Patil, Gunvant, Rouf Mian, Tri Vuong, Vince Pantalone, Qijian Song, Pengyin Chen, Grover J Shannon, Tommy C Carter, and Henry T Nguyen. 2017. “Molecular Mapping and Genomics of Soybean Seed Protein: A Review and Perspective for the Future.” *Theoretical and Applied Genetics* 130 (10): 1975–91.

Piper, Ernest L, and Kenneth I Boote. 1999. “Temperature and Cultivar Effects on Soybean Seed Oil and Protein Concentrations.” *Journal of the American Oil Chemists’ Society* 76 (10): 1233–41.

Song, Qijian, Long Yan, Charles Quigley, Edward Fickus, He Wei, Linfeng Chen, Faming Dong, et al. 2020. “Soybean BARCSoySNP6K: An Assay for Soybean Genetics and Breeding Research.” *The Plant Journal* 104 (3): 800–811.

Taylor, Julian, and David Butler. 2017. “R Package ASMap: Efficient Genetic Linkage Map Construction and Diagnosis.” *arXiv Preprint arXiv:1705.06916*.

Warrington, C Vꎬ, H Abdel-Haleem, DL Hyten, PB Cregan, JH Orf, AS Killam, N Bajjalieh, Z Li, and HR Boerma. 2015. “QTL for Seed Protein and Amino Acids in the Benning Danbaekkong Soybean Population.” *Theoretical and Applied Genetics* 128 (5): 839–50.