ABSTRACT

GILLENWATER, JAY HAUGE. QTL Mapping of Seed Composition Traits and Assessment of Yield in Two Soybean Populations. (Under the direction of Rouf Mian and Jeffrey Dunne.)

Protein and oil content in the soybean seed are two of the primary traits which, along with yield, dictate the value of this crop in the current market. Simultaneously improving all three traits through conventional breeding techniques is difficult due to inverse correlations between protein and yield, and protein and oil. It is possible that this difficulty can be overcome following a better understanding of the genetic architectures which control each trait, and their relationship with one another which may be done using QTL mapping studies. To meet these objectives, two recombinant inbred line (RIL) populations of 180 and 170 individuals were utilized to identify QTL for protein, oil, and seed weight and assess the stability of these QTL across genetic and environmental backgrounds. Linkage maps consisting of 421 and 416 SNP markers were created for each population and used for QTL mapping. QTL for protein and oil were found across genetic backgrounds and populations on chromosomes 14 and 20 at similar positions. An additional QTL for oil was found on chromosome 15 across genetic and environmental backgrounds. Three QTL for seed weight, three for oil, and two for protein were found in multiple environments but a single population. Three QTL for seed weight, seven for oil, and three for protein were found in a single population and environment.

A selection of lines from these populations were evaluated for yield to identify genotypes with improved protein and oil contents without significantly reduced yield. These lines will be used in an ongoing program to identify breeding lines with high yield and protein content. A total of 80 lines were evaluated for yield and protein and a selection of 30 lines was found with increased yield and protein relative to yield checks. These lines will be grown in two yield trials in 2020 to assess the stability of findings from the 2019 season.

QTL Mapping of Seed Composition Traits and Assessment of Yield in Two Soybean Populations

by
Jay Hauge Gillenwater

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Crop Science

Raleigh, North Carolina

2020

APPROVED BY:

Rouf Mian Co-chair of Advisory Committee	Jeffrey Dunne Co-chair of Advisory Committee

Ryan Martin

DEDICATION

To my parents.

BIOGRAPHY

Jay Gillenwater was born in Chapel Hill, North Carolina, but lived for most of his life in Charlottesville, Virginia. He earned his B.S. in Environmental Science and B.A. in Chemistry from the Virginia Polytechnic Institute and State University in May 2018.

ACKNOWLEDGEMENTS

I would like to thank my advisor for his help and guidance through this entire project. I would like to thank my committee members for their help and input to my research and studies. I would like to thank all the members of the soybean nitrogen fixation unit for their companionship and tireless help with all my planting, harvesting, processing and data analysis. I am grateful for the department of Crop and Soil Sciences and NC State for providing me with a good and stimulating place to study. Finally, I am very thankful for my friends and family for supporting and encouraging me all through my studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

1.1 A Brief History of Soybean Use and Breeding Objectives

Common soybean, [*Glycine max* (L) Merr.], is believed to have been domesticated from it's wild progenitor *Glycine soja* somewhere near the Huang-Huai Valley in central China around the 11th century BC, although the exact time and location of it's origin are difficult to determine for certain due to it's antiquity[Hym05]. Soybean was traditionally grown as a food crop, and processed into products such as tofu, edamamé, soy milk, and soy sauce. Soybean was then spread from China to Japan and Korea, and eventually to Europe in the mid 1700's. Soybean was introduced to North America by Samuel Bowen who began growing the crop in Georgia in 1765. It is difficult to determine the exact date at which soybean was introduced into South America, although the first known record is an article written about

the crop by Gustavo D'Utra in Brazil in 1882 (http://www.soyinfocenter.com/).

Following the disruption of oil imports after the onset of world war II, the United States rapidly expanded domestic production to accommodate for the shortage, and doubled domestic oil production between 1941 and 1942. Following world war 2, soybean remained a valuable crop for American growers as consumers worldwide gradually shifted preference from animal to vegetable fats. Additionally, more efficient means of producing protein-rich soybean meal for animal feed were developed in 1949-1950 which further added to the value of the crop. The United States share of worldwide production of soybean continued to grow to a peak of 76.9% in 1969 and then began to decline as exports from Brazil and Argentina increased(http://www.soyinfocenter.com/).

Breeding efforts up until this point were largely undertaken by public institutions but the private sector began to play a much larger role in soybean breeding in America following the passage of the Plant Variety Protection Act in 1970[Sil17]. The development and increased availability of molecular markers, and published research utilizing these technologies allowed for these traits to be analysed at the molecular level, and understood in ways that were impossible before[Sta08].

Seed oil, protein, and yield have continued to be among the most important traits in soybean production in western cultures due to the continued use of derived oil and protein products. Soybean accounted for approximately 59% of global production in 2019, followed by rapeseed which comprised 12%. Soybean is also by far the largest supplier of protein meal worldwide. In 2019, soybean meal comprised nearly 70% or total worldwide consumption, again followed by rapeseed which comprised 12%. Brazil, the United States, and Argentina were the top three producers of soybean in 2019. Brazil was the leader with 4,630 million bushels produced, followed by the United States with 3,557 million bushels, and Argentina with 1,984 million bushels (soystats.com).

1.2 Seed Composition Traits

Seed oil, protein, and yield are three of the most important agronomic traits for current soybean production. All three of these traits are believed to be controlled by many genes, and are influenced by the environmental conditions in which they are grown[Gup17; Ren09]. Furthermore, these traits are not expressed independently of one another, but rather show consistent correlation patterns which make simultaneous improvement of all three difficult[Bur87; Pat17; Cha15]. Before moving to a discussion of how these traits are mapped, a review of the value of these components, their physiology, and their relationship with one another will be worthwhile.

1.2.1 Seed Oil and Protein Content

Yield is important to farmers independent of seed composition. The two major uses for soybean in the United States are for its oil, and then as meal for poultry and swine after extraction of the vegetable oil[Pet13]. It is estimated that 60% of the value of a soybean cultivar comes from its protein content, and the remaining 40% come from it's oil content[Pat17]. Because of the importance of both oil and meal, a soybean cultivar is often evaluated on the basis of it's total oil + protein content, with a "rule of thumb" being that protein + oil on a dry basis should be at least 62.5% [Pat17].

Oil and protein are accumulated in soybean seeds to provide the plant with the energy and nutrients, respectively, which it requires to grow. These components are stored in oil bodies and protein bodies in the cotyledons of the seed embryo[Med14; Sil17]. This competition for shared storage space in the seed is partially responsible for the negative correlation between seed protein and oil content[Med14]. On a dry basis, soybean seed typically contain about 20% oil, and 40% protein content by mass, although the USDA germplasm collection contains lines with a protein content upwards of 57.9%, but with the cost of a severe loss in seed yield and pod count[Per08b; Per08a; Med14]. This negative

correlation between seed protein and yield is well known[Bur87; Cha15] and is a common challenge in producing new, more valuable varieties as all three traits are desirable. In spite of these negative correlations, some varieties developed through conventional breeding techniques have been released fairly recently with high protein content, and only moderately reduced yield[Bur99; Che11; Pan06d].

These traits are also affected by the environment in which the cultivar is grown. Temperature and moisture are two of the most influential factors for seed composition traits[Med14; Cha15]. Seed oil content is positively correlated with growing temperature, but growing temperature tends to have little effect on protein content[How58; How53; Nae08; Kan97; Ren09], although some studies have also noted an increase in protein content with rising temperatures as well[Wol82; Dor92]. The correlation of these traits with temperature is important to consider when comparing seed traits for population grown in different latitudes, or which are planted at a date that exposes them to different temperature conditions. Moisture stress tends to decrease seed size and count by reducing the duration of the seed filling stage, but has uncertain and sometimes contradictory effects on seed protein and oil contents[Mec84; Rot10; Dor92; Kum06; Rot09; Car09; Spe01; Ros88].

Little is currently known about the specific nature of the metabolic control of these traits. However, a substantial amount of effort has been dedicated to unraveling a genetic explanation for the correlations among them[Nic06; Pat13; Seb00]. These studies, among many others, have identified QTL for multiple seed composition traits which co-localize to similar regions in the soybean genome. These findings indicate that these traits are controlled at least partially by the action of pleiotropic genes, or genes which control each trait separately which are in tight linkage with one another. For many of these loci, it is difficult to determine which genetic phenomena is correct although some studies have identifyied QTL which are only associated with a single trait. A more in depth overview of these mapping studies will be presented at the end of this chapter.

The sequencing of the soybean genome[Sch10] and the availability of genomics and

genetics resources like SoyBase (http://www.soybase.org) have been instrumental in facilitating studies of the soybean at a functional genomics level. The advent of next generation sequencing (NGS) technologies have allowed for new approaches to evaluate each trait on the level of the transcriptome in soybean and other oilseed crops, and potentially identify significant hubs in their metabolic framework[Gup17].

1.2.2 Conclusions

Seed oil, protein and yield are controlled by an intricate molecular framework. Pathways controlling yield and seed composition do not act independently from one another, but have some degree of overlap between genes, how the expression is controlled, and how genes respond to environmental factors. The task of uncovering the exact nature of this molecular control is challenging, and far from complete. Historically, it has been difficult to simultaneously increase all three traits using conventional breeding techniques, but a better understanding of how these traits are controlled may allow for continued improvement.

1.3 QTL Studies

QTL is an acronym for *quantitative trait locus*. A *quantitative trait* is some measurable quantity of an organism which displays a quantitative distribution of possible values. In other words, these are traits which may take on any value in a continuous range of possible values, where the probability of observing any value is associated with some underlying statistical distribution. These contrast with *qualitative* traits which may only take on values from a finite set of possibilities. Examples of qualitative traits are the pea pods and flower colors observed by Gregor Mendel in his classical study, while traits such as yield and seed protein and oil content are quantitative in nature. A *locus* is a position on an organism's genome. Together, a QTL is a position on some organism's genome which is associated with the expression of some quantitative trait. It follows that the purpose of a QTL mapping

study is to detect positions on the soybean genome which are associated with the observed variability of these traits.

1.3.1 Mapping Principles

The methods used to find these regions require an understanding of the genetic concepts, and experimental techniques which make them possible. The focus of this next section will be to describe the process of QTL mapping from first principles so that the terminology, methodologies, and findings of QTL mapping studies may be better understood.

1.3.1.1 Genetic Variability

Before mapping is done, there should be some evidence that the traits of interest are controlled at least in part by their underlying genetic material, as opposed to the influence of the environment in which they reside. The degree of this genetic control is often expressed as the trait heritability. Heritability is calculated as *broad sense* or *narrow sense* with the formulations below[Feh91].

$$Broad\ sense:\ H^2 = \frac{\sigma_A^2 + \sigma_D^2 + \sigma_I^2}{\sigma_{ph}^2}$$

$$Narrow\ sense:\ h^2 = \frac{\sigma_A^2}{\sigma_{ph}^2}$$

Where σ_A^2 is the additive genetic variance, σ_D^2 is the dominance variance, σ_I^2 is the epistatic genetic variance, and σ_{ph}^2 is the total phenotypic variance. The formulation with the most applicability to breeding efforts is the narrow sense heritability of a trait[Feh91]. This accounts for only the proportion of the phenotypic variation that is explained through additive genetic variance. This is relevant to QTL mapping and marker assisted selection schemes because the utility of a marker for MAS is often determined by the proportion of phenotypic variation it is able to explain, relative to the overall heritability of the trait[Lan90].

Once it is established that there is sufficient evidence of genetic control of a trait, it is

necessary to chose the initial parents for the mapping population such that you have the best chance of finding regions which control it. In soybean, this process typically involves searching for inbred lines which differ significantly in terms of the trait (or traits) of interest for the mapping study. Preferably these lines will also not be seriously compromised in terms of other valuable agronomic qualities as well[Sil17; Sin15]. There is often an incentive to use distantly related parents to maximize the allelic diversity in the resulting progeny. In soybean, efforts have been made to use wild (*Glycine soja*) or exotic plant introductions to increase this genetic diversity, and potentially detect new alleles which may be introduced to breeding programs[Pri15; Men02; Ase16]. The identification of genetically diverse individuals is important, but challenging for these populations due to the low diversity of North American soybean cultivars from a long history of domestication and genetic bottlenecks[Hyt06].

1.3.1.2 Recombination

Diploid species like soybean possess two copies of each chromosome, where one copy is inherited from each parent. In highly inbred individuals, each copy will be very nearly identical in terms of the nucleotide bases which make them up, and subsequently the allelic identity of the genes which these bases encode. An individual created from crossing two inbred parents will then contain a set of chromosomes that contain the same genes at the positions in it's genome where both parents possessed the same alleles, and have different sequences at all the positions where it's two parents had different sequences. Such a site at which different alleles exist when the whole population is considered are referred to as polymorphic. These sites are important in mapping populations as they are the positions at which the individuals can be compared, and the differences in their phenotypic performance associated with their differences in genetic material.

At this stage, the organism may have different copies of an allele across chromosomes with respect to one parent or the other, but the sequence of alleles along each copy will be

identical to the identity of the alleles in the parent from which it came. This configuration of alleles is referred to as the parental or nonrecombinant genotype. During meiosis, sexually reproducing organisms undergo a process in which chromosomes are first duplicated, and then aligned with one another to form homologous pairs of chromosomes. Following this, the pairs may undergo a process by which each chromosome may exchange a portion of its DNA with the DNA of it's aligned pair. After this exchange, the pairs are divided twice to produce haploid gametes. The exchange of genetic material between aligned chromosomes during meiosis allows for the formation of sequences of alleles along chromosomes of a progeny which differ from the sequences which were present in either of their parents. The configuration of alleles which is not present in either parent is referred to as the recombinant genotype, as it arose from the mechanism of recombination. It follows from this that further generations of progeny may then undergo additional rounds of recombination following further mating or self fertilization to induce a greater divergence from the parental genotypes. The mechanism of recombination is exploited by breeders to systematically induce this deviation from the initial sequence of alleles to create a vast mosaic of new genotypes. In plant breeding programs, the structure of mating is controlled so that the frequency of genotypic classes expected to be observed at a given generation for a population can be known in advance. These controlled populations which are created for the purpose of identifying loci controlling traits of interest are known as mapping populations[Feh91; Sin15; Bal08].

1.3.1.3 Mapping Populations

The breeding scheme used to cross two variable parents, and the progeny of this initial cross are then used has important implications on the implementation of mapping studies. There are many possible schemes which may be used in these studies, ultimately the choice of which type of population to use must be catered to the specific needs of the mapping task at hand, and the organism under study[Sin15; Fer06]. Soybean easily under-

goes self-fertilization which allows for a rapid reduction in heterozygosity in a segregating population[Sta08]. One often used mapping population type in soybean mapping studies are populations of recombinant inbred lines (RILs). By utilizing the single seed descent technique[Bri66; Fun13], breeders can quickly produce large populations of recombinant inbred lines, so named because they are comprised of highly inbred individuals which consist of "reshuffled" versions of their parents genotypes due to the recombination that occurred during each reproductive cycle. These populations are beneficial for mapping studies because the inbred lines which make them up will nearly always self-fertilize to produce genetically identical progeny. Because of this, nearly identical individuals can be grown in different years, and in different locations to assess for the effect of these factors on quantitative traits[Sin15; Sil17; Bal08].

1.3.1.4 Genetic Markers

It is necessary to have some way to track alleles through a segregating population to determine what set of alleles any given individual in a population has in some stage of the population formation. In the early days of genetics, it was only known that there was *some* material which controlled the expression of traits, but the identity of this material was a mystery. Because of this, the first markers did not rely on molecular material at all, and instead used observable physical properties of an organism to track the un-observable genetic variation which controlled their expression. These so-called morphological markers were utilized to follow the transmission of alleles between parents and progeny, and eventually in developing the concepts of genetic linkage, and the construction of the first linkage maps[Stu13]. In soybean, morphological markers such as flower color are often used to identify successful crosses between parents of different flower colors, but these markers are rarely used in modern mapping studies due to the time required for their assessment, and the availability of more precise and numerous molecular markers, so named because they represent specific sequences or individual base pairs of an organism's molecular genetic

material[Sin15; Sil17; Sta08].

These molecular markers are often given a "generational" classification following the order in which they were developed, as well as the systems and molecular material which they utilize. Restriction fragment length polymorphism and randomply amplified polymorphic DNAs (RFLP and RAPD) belong to the first generation, simple sequence repeats and amplified fragment length polymorphisms, (SSR and AFLP), belong to the second generation, and single nucleotide polymorphisms belong to the third generation. Generally, these systems became increasingly user friendly relative to the previous generations, and also tended to provide larger sets of markers[Sin15]. SNP genotyping systems are currently widely used in soybean breeding programs, a procedure which has recently been simplified further with the development of high and medium-density Illumina Infinium BeadChips which allow for high-throughput calling of many SNPs which have been specifically developed to optimize information per SNP and cost[Son13; Son14]. The number of markers continues to increase with recent advances in genotyping by sequencing (GBS) approaches which have become feasible due to the decreasing cost of sequencing an individual's genome. As these techniques continue to decrease in cost, practitioners will be able to genotype individuals at a finer resolution and further decrease the amount of missing genetic information for their populations[Sin15]. If the implementation of pure GBS is not possible for a practitioner due to cost constraints, options such as skim-seq have been implemented which offer comparable performance at a reduced cost[Bat15].

1.3.1.5 Linkage

Linkage is a genetic phenomenon which is closely tied to the process of recombination. Briefly, linkage refers to the nonindependent inheritance of genetic material. Classically this phenomenon was manifested by the tendency for some phenotypic traits to be observed together in a much higher frequency than if the two traits had been passed onto a progeny independently. Linkage is tied to the process of recombination because if the genes which

control some traits are located physically close to one another on the genome, there is less of a chance for a recombination event to occur between the two causal sequences, and you would expect to see a higher proportion of individuals in a population who retain both of the genes than you would if the genes were spaced far apart. As the spacing between locations on a chromosome increases, so does the probability of a recombination event occurring in that space. Eventually the interval becomes large enough that the estimated recombination fraction between the two loci reaches a maximum of $\frac{1}{2}$. Markers with a recombination fraction below $\frac{1}{2}$ are considered to be linked, and those with a pairwise RF of $\frac{1}{2}$ are considered to be unlinked. With the use of molecular markers, this property can be exploited to determine the order of markers, and traits can be statistically associated with a marker. The task is then to determine where each marker is located so that the specific locations of the genome which control some trait can be found[Liu97; Bal08; Wu03].

1.3.1.6 Linkage Maps

The method by which this location is determined is known as linkage mapping, because it exploits the phenomenon of linkage between a set of markers to determine their positions along a genome. The process of creating a linkage map is divided into two main steps: forming linkage groups of markers, and ordering the markers within those linkage groups[Bal08]. Markers are typically partitioned into linkage groups using the criteria of recombination frequency, and a LOD (log_{10} odds of linkage) score which is closely related to these recombination frequencies and is a quantification of the odds of linkage existing between a pair of genetic markers. The ultimate goal of marker grouping is to find linkage group subsets of markers which correspond to the physical chromosomes of the organism. Smaller values of RF or larger values of LOD scores will produce a greater number of linkage groups as the criteria for linkage become more stringent[Wu03].

Recombination fractions are converted to genetic distances based on the number of recombinations observed. The functions used to do this are called mapping functions, and

the two most often used are the Haldane and Kosambi functions. The two functions differ in how they handle the spatial probability of recombination events occurring along the genome. Haldane's formulation[Hal19] assumes that the event of a crossover occurring has no impact of a subsequent recombination event occurring in a nearby region, an assumption which is referred to as no-crossover interference. The Kosambi function[Kos16] is an extension of the Haldane mapping function which was developed to account for empirical observations of actual recombination events which deviated from this assumption. The two functions are formulated as follows:

$$Haldane:\ x = \frac{1}{2} ln(1-2r)$$

$$Kosambi:\ x = \frac{1}{4} ln\left(\frac{1+2r}{1-2r}\right)$$

Where x is the genetic distance in units of Morgans, and r is the recombination frequency.

The naive solution to the problem of ordering markers is to simply list all the possible orders and calculate a likelihood score for each possible order. In fact, this was the strategy implemented to a limited extent in early mapping software[Lan87]. However, this technique became problematic as the number of available markers quickly began to increase. The total number of possible paths through the set of points (markers) can be calculated as:

$$n = \frac{m!}{2}$$

Where n is the total number of orders, and m is the number of markers. This means that for even a modest set of 20 markers will have approximately 1.2e18 possible orders to evaluate, a task that is impossible given the computational resources available today and most likely for the foreseeable future as the number of available markers continues to increase. To overcome this, several computational algorithms have been developed to find the optimal orders among even substantial numbers of markers[Mol09; Wu08;

Wu03]. Various software have been developed to implement these steps, some examples are MAPMAKER[Lan87], IciMapping[Men15], JoinMap[VO06], onemap[Mar07], r/qtl[Bro03], MSTMap[Wu08], and ASMap[Tay17].

Efforts have also been made to combine genetic maps constructed using each generation of marker systems into single consensus maps. The current iteration of such a map for soybean, Soybean consensus map 4.0, contains a 3792 SNP markers, 1006 SSR markers, 664 RFLP markers, and 38 classical markers. Such maps are useful for breeders as they allow for results of many mapping studies to be compared with one another, and common QTL sites to be better identified[Hyt10; Gra09]. Detailed information about this map and many of it's features may be found at https://www.soybase.org/sbt/.

1.3.2 Mapping Techniques

Several methods for mapping QTL have been developed, each with their own strengths and weaknesses. Each method employs a distinct statistical interpretation of what a QTL is, and a brief discussion of these interpretations will be helpful for understanding the results produced by each method. The major methods may be grouped into two general categories for the techniques designed for the detection of single QTL, and those for the detection of multiple QTL.

The simplest of the methods for the detection of single QTL is known by a few names such as *single marker analysis* (SMA), *marker regression*, or *single point analysis*. This method was the first developed, and is conceptually the most straightforward of the available techniques. SMA considers one marker at a time, the individuals in the mapping population are then classified into groups based on the allelic identity of the marker under question. The phenotypic means of some trait for each group are then calculated, and compared using a standard t-test. A QTL may then be declared if the phenotypic means of the two groups differ significantly. The method is limited in that the test can only be performed at the site of the marker itself, which may be problematic in the case that few markers are available,

or if markers are not distributed sufficiently across all chromosomes. Additionally, it is not possible to determine how many QTL a marker is associated with, as it is possible that there may be multiple QTL in the region of a single marker. Furthermore, SMA is not able to detect epistasis between QTL. Still, this is a computationally simple, and useful way of finding markers located near a QTL without the use of a linkage map.

The next step in mapping development for single-QTL methods was *simple interval mapping*. The *interval* component of the name is due to the fact that this method was developed to detect QTL in the intervals between markers, not just at the sites of the markers themselves. The technique operates by calculating the conditional probabilities of genotypes in the intervals between markers by using the known genotype of the markers, and the position of the putative QTL in the interval between the two markers. These conditional probabilities can then be used to scan for QTL along the entire genome. This offers a more powerful and comprehensive technique for detecting QTL than was offered with SMA.

Many traits of interest are not controlled through the action of single QTL however. There are serious pitfalls which can arise when using the single QTL mapping methods for analysing traits which are inherently controlled through multiple QTL. Some of the most well known of these shortcomings are that a *ghost QTL* may be detected in the middle of two real, linked QTL of the same sign[Mar92]. Interval mapping is also only able to estimate the average effect of QTL in a genomic region, and is unable to parse the effects of many small QTL which may actually be present. This effect has important implications on the repeatability of experiments if the small individual QTL are effected by the environment in different ways, which would then produce a different average peak in terms of both magnitude and position[Bal08]. Many important quantitative traits including oil, protein, and yield are controlled through many genes so it was beneficial to develop QTL mapping techniques which could control for the influence of multiple QTL when performing a scan. The two most often used techniques for multiple-QTL mapping are composite interval mapping (CIM) and multiple interval mapping (MIM). Both methods seek to find a multi-

variable model to estimate the effects of many qtl on some quantitative trait.

Composite interval mapping was the first of the multiple QTL mapping techniques to be developed. The method operates by isolating a putative QTL within a moving window of marker cofactors. Using a sliding window, the effects of multiple QTL can be more accurately determined by estimating each of them in turn within the isolated moving window. The issue with this technique is that the somewhat subjective choice of which markers to include in the model as cofactors can reduce the power to detect a QTL[Bro09; Bal08].

Multiple QTL Mapping (MQM) methods were developed to overcome the shortcomings of CIM, most notably the choice of which marker cofactors to include in the model[Bro09; Bal08]. MQM treats the search for QTL as a model selection process where the ultimate goal is to scan through the full space of potential QTL to build the most parsimonious model for a given phenotype[Bal08; Bro09]. This statistical interpretation of QTL mapping matches well with the goals of identifying markers for MAS in quantitative traits where we also want to identify multiple QTL which contribute to a trait, and estimate the relative contribution of each[Sin15; Lan90; Col05]. To identify this model, the software which perform MQM employ different search schemes to progressively build up the best model. The development of the techniques which are used to accomplish this is an ongoing area of research but in practice, common search approaches include forward-backward search, simulated annealing, and genetic algorithms coupled with model comparison criteria such as AIC to produce the final model[Bro09; Bal08].

These techniques, and several specialized varieties of them are currently employed by several software. Some commonly used choices are QTL Cartographer[Bas05], IciMapping[Men15], r/qtl[Bro03], and MapQTL[VO09].

1.3.3 Marker Assisted Selection

Beyond a better understanding of the genetic architecture which control these traits, it is important that the results of these mapping studies have practical application to breeding

efforts as well. Measuring a phenotype in the field or greenhouse is a time-consuming and often expensive task. The motivation of marker-assisted selection is that progeny carrying beneficial alleles may be identified through the presence of markers which were found to be associated in QTL mapping studies at an early stage, and thus circumvent the need for an extensive phenotypic evaluation prior to selection. Additional benefits include the ability to select for low-heritability traits, and the reduction of phenotypic measurement errors which may become more likely as the number of progeny to evaluate increases[Lan90; Col05]. In soybean, the most successful applications of MAS thus far have unsurprisingly been for traits which are controlled through the action of few large-effect QTL. Some examples are markers which have been developed for root-knot nematode resistance, rust resistance, tolerance of iron deficiency, and cyst nematode resistance[Cre99; Nar12; Seb09; Yu11; Jen14; Kla13]. The unfortunate irony of the situation is that the high-heritability traits controlled through the action of few-large effect genes are the traits which benefit least from marker assisted selection[Lan90; You99]. There are still benefits to using MAS for these traits however. Traits such as disease resistance may be controlled by only a few genes but are difficult to score in the field, and therefore using closely associated markers can improve on the difficult phenotypic analysis[You99].

Quantitative traits like oil, protein, and yield are controlled through the combined action of many QTL which each contribute a small proportion of the variance for the trait[Med14; Pat17]. Furthermore, there is evidence that the control of these traits may be due at least in part to the action of pleiotropic genes, leading to the well documented correlation structure between them which was discussed previously. Due to these factors, identifying markers which can be implemented in MAS for these correlated quantitative traits is more difficult, and requires an extensive exploration of the allelic diversity for the identification of suitable markers.

1.3.4 Previous QTL Mapping Studies in Selected Traits

Soybean researchers are fortunate to have access to the genetics and genomics database known as SoyBase (http://soybase.org)[Gra09] which catalogues a large body of research and data relevant to the understanding of the genetics of soybean. Among this data is a well documented record of the studies which have reported QTL for many traits, where these QTL may be found on several consensus linkage maps, and various meta-data about the QTL and the studies which found them. At the time of writing, 482 QTL have been found for traits associated with oil content, 367 for protein, and 339 for yield. A tabulation of the QTL found for these general categories may be found in table A.1 in the appendix. The earliest QTL mapping study in a trait related to seed protein, oil, or yield in the database was published in 1992, and the most recent was published in 2018. There have been quite a large number of QTL identified in just the last few years in these three traits alone. A diagram of this trend can be seen in Figure A.1 in the appendix.

QTL for seed oil, protein, and yield have been found on all 20 chromosomes of the soybean genome. Plots summarising the counts of previously found QTL in the oil, protein, and yield categories along each chromosome of the GmConsensus 4.0 map may be seen in figures A.41-A.45 in the appendix. It can be seen that QTL for all three traits have been identified in similar intervals of the genome, but also that some positions tend have higher counts for a specific trait relative to the others. These plots are only meant to illustrate the rough distribution of many detected QTL, but efforts have been taken to systematically scan published QTL in a similar fashion to identify QTL "hot-spots". These so-called *meta-QTL* studies will be covered in more detail later in this chapter. Additional plots identifying the specific QTL which underlie these counts can be found in the appendix in figures A.46-A.65.

It is difficult to increase protein, oil, and yield simultaneously due to the negative correlation between oil and protein, and between protein and yield, although it is possible to increase each trait one at a time using conventional breeding techniques[Bur87]. Addi-

tionally, it is challenging to identify QTL which are stable across both genetic backgrounds and environments both due to the resources required to conduct such studies, and the effect of the environment on these traits. Despite these difficulties, several studies have been successful in identifying stable QTL and associated markers which are positively associated with one or more of these traits, without compromising the value of the others. The identification of these stable QTL which are uniquely associated with one value trait, or positively associated with multiple traits is important due to their potential value in MAS and gene pyramiding breeding schemes[Pal10; Bru97].

QTL mapping studies have been performed for 12 traits related to oil content, 28 for protein, and 20 traits related to yield. The traits most relevant to this study are seed protein and seed oil content, and their relationship with seed yield or traits closely related to yield such as seed weight. Other common categories for each trait focus on specific aspects relevant to protein, oil, or yield. Counts of previously discovered QTL for these classifications may be found in the appendix in tables A.1 and A.2. The seed protein, oil, and yield categories are also by far the most studied aspects of each general trait. Of particular relevance to this dissertation are those studies which jointly considered the relationship between all three traits. There have been a large number of QTL mapping studies published for these three traits, here I will present a selection of these mapping studies which are representative of the goals of this dissertation, as well as other studies which have detected particularly stable QTL.

Despite the abundance of mapping studies for oil, protein, and yield, relatively few QTL have been given a *confirmed* status. The requirements of the Soybean Genetics Committee (http://www.soybase.org) dictate that a confirmed QTL must be detected in a separate set of meiotic events and environments from the original population and be detected with an experiment-wise error rate of 0.01 or lower. These QTL are identified with the *cq* prefix in Soybase. At the time of writing, there are only 16 confirmed QTL for seed oil, 16 for seed protein, 13 for seed weight, and 1 for seed yield. All of these QTL were found in three

mapping studies[Fas04; Pat13; Nic06]. More detailed data on these confirmed QTL may be found in the appendix in Table A.3.

The most recent of these studies[Pat13] identified seven QTL for seed protein on chromosomes 3, 5, 6, 7, 8; six QTL for seed oil on chromosomes 2, 3, 5, 6, and 15; and four QTL associated with seed weight on chromosomes 1, 6, and 19. The experiment was conducted using two populations derived from crosses of two lines (PI 438489B and PI 597516C) to a common parent (Magellan). Both populations were grown in multiple environments in Missouri such that the study could meet both the meiotic events and environmental requirements of the confirmed QTL designation. The QTL detected for oil and protein on chromosome 5 and 6 were found in similar genetic positions.

The study conducted by Nichols et al.[Nic06] was a fine mapping study aimed at discovering if additional recombination in the vicinity of a previously detected QTL on linkage group I (chromosome 20) could reduce the inverse correlation of this QTL with seed oil and yield traits. The specific aims of this study were to validate previous findings[Seb00; Die92b; Chu03; Bol10] of a major QTL controlling protein content on this linkage group, and to determine if the correlation between these three traits was due to pleiotropy, or due to close linkage of QTL controlling each trait independently. The study was successful in confirming the QTL but was unable to conclusively determine if the negative effects associated with the QTL were due to close linkage with nearby alleles. While a large number of studies have identified QTL at this locus, few alleles which are not associated with a decrease in yield or oil have been identified. Of these studies, only the high protein allele from the Danbaekkong cultivar has been shown to be unlinked with yield drag [War15; Mia17]. The difficulty in fine mapping this QTL may be due to its position in a heterochromatic region of the chromosome, which have inherently lower rates of recombination[Sch10; AK19].

The earliest of the confirmed QTL studies[Fas04] was conducted specifically to address the need for QTL to be validated across environments and genetic backgrounds as the authors noted the inconsistencies reported in previous quantitative trait studies. The study

was conducted between between two years and in three environments and evaluated QTL for seed oil, protein, and weight. The study was successful in confirming three QTL for seed oil and two QTL for seed protein. The positions, and markers associated with the QTL found in these three sources are tabulated at the end of the chapter.

Beyond these confirmed QTL, there are still promising results from studies with similar objectives to this dissertation. These studies focused on identifying QTL for oil and protein content, while evaluating the relationship of these traits with seed yield. Several of these studies have been conducted fairly recently and have utilized multiple mapping populations across multiple environments and/or years.

One such recent study[Wan14] identified 16 QTL for seed oil, 11 for seed protein, and 12 for seed yield. The study identified QTL associated positively with yield and oil and negatively with protein, QTL with opposite effects on protein and oil, QTL with positive effects on yield and oil, and QTL associated with each trait individually. The findings suggested pleiotropy or linkage between QTL governing the three traits, but also identified markers specific to each trait which may be suitable for MAS schemes. Findings were consistent within five environments, and the combined data. These QTL were further checked using a separate validation population in which 8 of the QTL for seed oil, 5 for seed protein, and 5 for seed yield were detected in the different genetic background.

Another study[Pat13] obtained similar results. Two RIL mapping populations were created from crosses of two soybean cyst nematode resistant PIs (PI 438489B and PI567516C) to a common, high-yielding parent (Magellan). Each population was grown in three locations. Seven QTL for seed protein were detected although only one was significant across environments and genetic backgrounds. The study identified three other QTL which were detected in both environments, but in only one of the populations. Six QTL for oil were identified, and two were detected across environments and populations. Four QTL were detected for seed weight, of which two were detected across environments and genetic backgrounds. Multiple QTL for protein and oil were detected in similar locations, leading the authors to

predict that the genes underlying these QTL were likely to have pleiotropic effects given the negative correlations between the protein and oil traits which they observed.

Other studies have consistently shown the presence of a large effect QTL for protein content on chromosome 15 (LG E)[Fas04; Die92b; Pat17]. One recent fine mapping study of this qtl[Kim16] was successful in mapping the location of this QTL to a 535 kb interval between makrers SSR BARCSOYSSR_15_0161 and BARCSOYSSR_15_0194. Using this information, the authors were then able to identify potential genes underlying the QTL using the positions of each marker in the Glyma.Wm82.a2 genome assembly.

It is important for a breeder that there be high confidence in the utility of a marker before it is used in a MAS breeding program, however, despite the large number of QTL mapping studies that have been conducted for these traits, there are relatively few markers that have been confirmed to be stable across environments and genetic backgrounds. To overcome this deficiency, efforts have recently been made to take advantage of the large number of published mapping studies to conduct meta-QTL analyses[Qi18; Van17; Qi11c; Qi11a; Qi11b]. The incentive in performing these studies is that the position of QTL can be more accurately and precisely determined by systematically combining the findings of many mapping studies using a likelihood based approach than they could be otherwise attained by considering the results of a single study alone[Gof00]. Performing a systematic review in this fashion also allows a practitioner to identify patterns in the published data which could have been easily overlooked otherwise given the large number of published QTL studies. The general procedure for these meta-analyses are to scour QTL mapping publications for traits of interest to obtain a data set of previously discovered QTL, and the markers closely associated with them. These QTL are then projected onto a genetic map which must contain the associated markers from each QTL's original study. Both of these steps are greatly facilitated by the availability of genetics databases like Soybase. Software such as BioMercator[Sos12] can then be used to refine positions for the meta-qtl. These studies have been used to mine causal genes for a variety of traits in soybean, including seed protein and

oil content[Qi18; Van17]. One particularly promising study[Qi18] was conducted recently. In this study, a meta-QTL analysis was first performed to identify positions associated with increased protein and oil content. These hot spots were then successfully explored on the transcriptome level in a set of chromosome segment substitution lines through a RNA seq analysis to identify regulatory genes which underlie the QTL.

Meta studies such as these offer hopeful examples how past results can be evaluated together to detect influential QTL. Their utility for identifying markers effective for MAS is yet to be tested in soybean, however QTL obtained from meta studies in other crop species have had good results when implemented in MAS[Kho09; Cha04; Zha17]. Beyond this, the findings of[Qi18] show how the results of traditional mapping studies can be used hand in hand with next-generation genomics studies of the same traits to provide additional details about how these traits are controlled. In Soybase, meta-qtl may be identified with the *mq* prefix.

There are many other[Man96; Spe01; Fas04; Kab04; Zha04; Pan06c; Rei06a; Bac09; Cha03; Csa01; Die92a; Du09; Esk13a; Esk13b; Fan15; Fun05; Gai07; Guz07; Ha14; Hac18; Han15; Hne96; Hyt04; Jeo11; Jun08; Kaz10; Kim12; Kur13; Lar94; Lee96; Lee01; Lei16; Li08; Li10; Li11; Lia10a; Lia10b; Lu13; Man93b; Man93a; Mao13; Ngu12; Nin18; Orf99; Pal09a; Pan12; Pan04; Pan06b; Pan06a; Pan07; Pan05; Qi14a; Qi14c; Qi14b; Qi11d; Qiu99; Ram14; Rei06b; Rey01; Rod10; Ros13a; Sal06; Shi08; Spe04; Sto04; Sun06; Taj03; Tis03; Vol02; Wan04; Wan12; Wan15b; Xie12; Yua02] QTL mapping studies which focused on these major seed quality traits. This is unsurprising given the close relationship between the traits, and their importance to breeders and growers alike. A more extensive tabulation of such studies, the positions of the identified QTL on the consensus map, and any markers associated with the QTL may be found in the appendix.

1.3.5 Conclusions

There has been extensive research into QTL mapping for seed oil, protein, and yield traits in soybean, but few genetic markers have been identified which are consistent across genetic and environmental backgrounds. The phenotypic correlations between protein, oil, and yield are often reflected in the genotypic data through QTL localised to similar genetic regions for the traits, although QTL associated with a single trait, or which counteract the inverse correlation between protein and oil, or protein and yield have been found. Tools such as meta-analyses and next-generation genomics technologies have shown recently that the combined results of these mapping studies can be used to obtain a better understanding of the causal genes which underlie these QTL, although additional research will be required to make a definitive statement about their utility in identifying markers for MAS programs. It is necessary to continue QTL mapping in these traits both to validate previously discovered markers in different environments and genetic backgrounds, and to potentially discover new alleles that can escape or reduce undesirable correlations between seed composition and yield.

CHAPTER

2

QTL MAPPING OF SEED QUALITY TRAITS IN TWO RECOMBINANT INBRED LINE SOYBEAN POPULATIONS

2.1 Abstract

Soybean [Glycine max (L.) Merr.], is one of the most widely grown commodity crops worldwide. In large part, the value of the crop depends on the seed composition traits, two of the most important of which are seed protein and oil content. The objective of this study was to identify quantitative trait loci (QTL) for these seed traits and 100 seed weight in two recombinant inbred line (RIL) populations consisting of 180 and 170 genotypes. Each population

was grown in three environments to assess the stability of QTL across both genetic and environmental backgrounds. Each population was genotyped with the SoySNP6k Illumina Infinium BeadChip. Linkage maps covering all 20 chromosomes were created for each population which contained 421 and 416 polymorphic markers for populations 201 and 202, respectively. A total of 13 QTL for oil content, 7 for protein content, and 6 for seed weight were detected through multiple interval mapping. Many of these QTL are co-located with previously detected QTL, and a tentative novel QTL for seed protein and oil was detected on chromosome 15 (LG E). The findings of this study provide an insight into the genetic architecture which controls these three traits, particularly in the given genetic backgrounds. The QTL identified or confirmed in this study hold promise for marker assisted selection (MAS) using the linked SNP markers.

2.2 Introduction

Seed protein and oil content are both valuable traits in soybean due to the widespread use of soybean seed for production of protein meal and oil products. It is also important that lines with improved composition traits do not severely compromise yield as it is the factor by which growers are often compensated.

Oil content is an important trait in soybean cultivars due to the widespread use of soybean for production of vegetable oils. In 2019, soybean made up 59% of the total oilseed production worldwide and 56% of vegetable oil consumption in the United States[ASA19]. Seed protein is used in meal for animal feed[ASA19]. In 2019, soybean comprised 69.3% of the total worldwide production of meal[ASA19]. Cultivars with optimal protein or oil content will be more valuable due to their use in the production of these products.

Historically, improving oil, protein, and yield in soybean by conventional means has been difficult due to the negative correlation between protein and oil, and between protein and yield[Bur87]. These three traits are quantitative in nature and are controlled through

the simultaneous action of many genes[Pat17; Gup17]. The positions of these genes on the soybean genome, and how much of the variation of each trait which they explain can be estimated through QTL mapping techniques. It is desirable to use QTL mapping to identify loci which are either uniquely associated with one of these traits, or jointly associated with an increase in multiple traits, because markers associated with such loci may be used in marker assisted selection (MAS) schemes to rapidly, and efficiently pyramid multiple valuable traits in a cultivar. To be effective in MAS, a QTL must be stable across both environments and genetic backgrounds[Sin15; Bru97; Lan90]. Furthermore, the results of QTL mapping studies can be expanded on in future fine mapping or gene cloning efforts which seek to identify the specific genes which underlie QTL[Sin15]. QTL mapping studies which consistently detect QTL for a given trait to a common region on the genome provide targets for these more focused studies.

Many QTL mapping studies for these three traits in soybean have been published, but relatively few QTL which are stable across both environments and genetic backgrounds have been found[Ros13b; Wan14; Pat13; Li10; Pal09b; Gai07; Che07b]. Due to this, there have been only limited attempts to incorporate these QTL into MAS schemes[Zha15; Wan15a]. It is necessary to continue QTL mapping studies for these traits so that previously discovered QTL may be validated in different genetic backgrounds and environments, and to identify specific alleles uniquely associated with one of these traits which can then be used in the future by breeders for further population development. Stable QTL detected by these mapping studies can then be further refined using additional markers as provided by genotyping by sequencing (GBS), GWAS , and fine mapping approaches to identify additional markers which can then be used for marker assisted selection.

Many studies have identified large effect QTL for protein and oil content on chromosomes 15, (LG E)[Kim16; Pat17], and 20, (LG I)[Mia17; Seb00; Die92b; Chu03; Bol10]. Although, historically it has been difficult to identify alleles on chromosome 15 that are not associated with decreased yield[Pat17].

As of June 2020, there are only 16 confirmed QTL for seed oil, 16 for seed protein, 13 for seed weight, and 1 for seed yield published on the soybean genetics and genomics database, SoyBase (www.soybase.com). To have a confirmed status, a QTL must be detected in separate sets of meiotic events, in multiple environments, and be detected with an error rate of less than 0.001. All confirmed QTL for the seed protein and oil traits were discovered in three publications[Fas04; Pat13; Nic06]. This is in stark contrast to the total of 258 QTL detected for seed oil, 209 for seed protein, and 153 for seed yield at the time of writing. This reflects the difficulty in finding universally stable QTL for these quantitative traits. QTL for protein, oil, and seed yield traits have been found on all 20 chromosomes of the soybean genome, and many are co-located in similar genetic regions which suggests a degree of pleiotropy or linkage among the genes which control these traits[Pat17; Gup17].

It is important to perform additional mapping studies using new populations, and evaluate these populations in many environments so that the stability of QTL can be better understood. This study utilized two recombinant inbred line (RIL) populations which were grown in three environments to test the stability of QTL across both genetic backgrounds and environments. Multiple interval mapping was performed to identify QTL and estimate their effects.

2.3 Materials and methods

2.3.1 Plant Materials and Experimental Design

The two populations used in this study were developed from separate crosses between parents which were found to consistently differ in seed protein and seed oil content. Population 201 was developed from a cross of LMN09-119 x N09-09 in 2015. LMN09-119 was from a cross of N6202[CJ10] x G98SF-114, G98SF-114 from a 'Benning'[Boe97] x 'Danbaekkong'[Kim96] cross; N09-09 was from a cross of N02-70 x G98-1420, N02-70 was from 'Santee'[Shi03] x 'Holladay'[Bur96], G98-1420 was from the cross of 'Boggs'[Boe00] x

'Doles'[Boe94]. Population 202 was developed from a cross of LMN09-19 x Nl3-47 also in 2015. LMN09-19 is a sister line with the same pedigree as LMN09-119; N13-47 was derived from a recurrent selection breeding started with intermating F1 progeny from eight crosses in 1965 at the ARS Soybean Research Unit in Raleigh, NC. These eight crosses ('Arksoy' x 'Lee', Arksoy x 'Ogden', Arksoy x D60-8107, 'Jackson' x Lee, Jackson x Ogden, Jackson x D60-8107, 'Roanoke' x Ogden, and Roanoke x D60-8107) were made in 1964. The male sterile female fertile line with the ms1 gene was used in this recurrent selection breeding scheme over many cycles of selections. In each cycle the top 10% high oil lines were selected from the progeny to start the next cycle. The N13-47 with desirable agronomic attributes and high (>24%) oil content was selected from as a plant row in 2013. The RILs of the two mapping populations were advanced to F5 generation by single seed descent (SSD) method of selection. A single plant was harvested to represent each F5-derived RIL.

In 2019, populations 201 and 202 consisted of 232 and 204 recombinant inbred lines, respectively. Each population was grown in a randomized complete block design with two replications at three locations. The locations used for the experiment were the Central crops research station in Clayton, NC, the Tidewater Research Station in Plymouth, NC, and the Caswell Research Farm in Kinston, NC in 2019. These locations will be referred to as Clayton, Plymouth, and Caswell, respectively.

2.3.2 Phenotypic Data Collection and Statistical Analysis

The traits mapped in this study were seed oil and seed protein content measured on a dry basis as well as hundred seed weight (SDWT). Protein and oil contents were measured using a Perten DA 7250 NIR infrared spectrometer. Seeds were sifted and cleaned of debris, off-color, and cracked seeds prior to measurement with the NIR instrument. SDWT was measured by manually weighing 100 seeds from each plot. Seeds were visually inspected prior to measurement to ensure that the sample contained whole seeds which were representative of the total plot and free of debris.

For each location, data were analyzed using a mixed model where genotype was considered as a fixed effect and replication was considered to be a random effect. The general formulation of the model was then:

$$y_{ij} = \mu + G_i + R_j + \epsilon_{ij}$$

Where y_{ij} is the phenotypic measurement for individual i in replication j, μ is the global mean, G_i is the effect of genotype i, R_j is the random effect of replicate j, and ϵ_{ij} is the random measurement error. Least square means (LSMeans) for each genotype were then obtained for each model using the *emmeans* package in R[Len19]. Phenotype measurements were averaged across locations to make combined datasets for each population. These least-square means and global means for each trait were then used in the subsequent QTL analysis.

2.3.3 DNA Extraction and SNP Analysis

Leaf trifoliates were collected from each RIL (F_5 Plant) and freeze-dried. Dried tissue was stored at -20°C until DNA extraction. DNA was extracted using a Qiagen DNEasy Plant Mini®kit following the manufacturers protocol. Extracted DNA was stored at -80°C. Before genotyping, extraction quality was checked by visualizing extracted DNA with an 2% agarose gel. DNA samples passing the initial quality check were then sent to the USDA ARS Soybean Genomics and Improvement Laboratory in Beltsville, Maryland and were genotyped using 5,403 SNPs from the SoySNP6k Illumina BeadChip. Genotype clusters were visualized and manually adjusted when appropriate using the GenomeStudio[Ill11] software following manufacturer guidelines.

2.3.4 Linkage Mapping

Prior to linkage mapping, SNPs were filtered for each population to remove monomorphic markers. Only markers with less than 20% heterozygosity + missing data were kept for mapping. Genotypes with greater than 25% missing data were excluded from the mapping process. Markers exhibiting segregation distortion at a p-value less than 0.001 were excluded as well. Groups of individuals with 95% or greater identical genotypes were identified, and only one individual was kept from each set. The identification and removal of these highly similar individuals was done with the *genClones* and *fixClones* functions from the ASMap package[Tay17]. Following these criteria, final populations used for QTL mapping consisted of 180 and 170 genotypes for populations 201 and 202, respectively.

Markers were first assigned to linkage groups following the chromosomes they have been aligned to in the Wm82.a2 genome assembly[Son14; Son13]. Markers were ordered within each group using the *mstmap.cross* function of the ASMap[Tay17] package in R using the Kosambi[Kos16] mapping function with default parameters. The procedure was used to create linkage maps for each population.

Linkage maps for populations 201 and 202 contained 421 and 416 SNP markers, respectively. Marker orders were generally in good agreement with their physical positions in the Wm82.a2 genome assembly. The average marker spacing in the population 201 linkage map was 6.27 cM. The number of markers on each linkage group ranged between 6 on chromosome 11 (LG B1), and 37 on chromosome 14 (LG B2). On average, each linkage group had 21.1 markers. The linkage map for population 202 had an average marker spacing of 5.76 cM. The number of markers per linkage group ranged between four markers on chromosome 4 (LG C1) and 40 on chromosome 14 (LG E). The average number of markers per linkage group was 20.8. More detailed information on each map may be found at the end of this chapter in Tables 2.1 and 2.2.

2.3.5 QTL Mapping

Least-square means (LSMeans) for seed protein, seed oil, and SDWT at each location as well as average data across locations were used as the phenotypes for QTL mapping. The r/qtl[Bro03] add-on package to the R statistical language[R C13] was used to preform QTL mapping. A scan for QTL was first performed using the *mqmscan* function with 100 cofactors and a cofactor significance of 0.05. QTL significance thresholds were determined following a LOD threshold obtained from 1000 permutations of the data with the *mqmpermutation* function. QTL statistics were obtained using the *fitqtl* function from r/qtl.

2.4 Results

2.4.1 Distribution of Phenotypes

Seed protein, seed oil, and SDWT were approximately normally distributed in all environments and in both populations. Detailed phenotype summaries for each population and location may be found in the appendix in tables A.28-A.31.

2.4.2 QTL Detected

The primary goals of this study were to identify and characterize QTL in terms of their stability across genetic backgrounds and environments. To facilitate this comparison, QTL have been categorized based on those which were identified in multiple environments and both populations, those which were identified in a single population but multiple environments, and those which were only identified in a single environment and population. Detailed tabulations of these QTL may be found in tables 2.3-2.5 at the end of this chapter.

2.4.2.1 QTL Detected in multiple populations and environments

Three QTL for seed oil content, and two QTL for seed protein were mapped to similar genetic regions in both populations, and each were detected in multiple environments as well (Table 2.3). An oil QTL was found at approximately 14 cM on chromosome 14 (LG B2) in the Plymouth location of mapping population 201, and in the Plymouth and combined data of population 202. This QTL was co-located with a protein QTL which was detected in the Caswell and Plymouth environments in population 201, and the combined and Plymouth environments in population 202. The QTL explained between 6.8% - 17.8% of the variation in oil content, and between 3.7% - 23.2% of the variation in protein content. A second QTL for oil content was found at approximately 74 cM on chromosome 15 (LG E). This QTL was detected in the Caswell and combined data for both populations and explained between 5% - 8.6% of the variation in oil content. An additional QTL was detected for both oil and protein traits on chromosome 20 (LG I). The QTL for protein and oil was detected at approximately 30 cM for both traits in population 201, and approximately 53 cM in population 202. This QTL explained between 7.8% - 31.7% of the variation in seed oil content, and between 6.7% - 40.1% of the variation in seed protein content.

2.4.3 QTL detected in a single population at multiple environments

Three QTL for seed weight, three for oil content, and two for protein content were detected in a single population at multiple environments (Table 2.4). QTL for seed weight were found on chromosomes 13, 15, and 20 at positions of approximately 60 cM, 74 cM, and 30 cM, respectively. These QTL explained between 13.6%-13.8%, 4.5%-6.0%, and 5.0%-11.0% of the phenotypic variance, respectively. QTL for seed oil were found on chromosomes 2, 17, and 18 at positions of approximately 30 cM, 45 cM, and 42 cM, respectively. These QTL explained between 4.5%-9.2%, 4.1%-4.4%, and 6.5%-7.4% of the variation in oil content, respectively. The QTL for protein were found on chromosomes 2 and 15 at positions of

approximately 30 cM and 72 cM. These QTL explained between 6.9%-7.8% and 7.5%-8.2% of the variation in protein content.

2.4.4 QTL detected in a single population and a single environment

Three QTL were detected for seed weight, seven for seed oil, and three for protein in a single population and in a single environment.

Seed weight QTL were found on chromosomes 9, 12, and 14 (LGs K, H, and B2). Oil QTL were found on chromosomes 2, 4, 8, 14, 15, 17, and 18 (LGs D1b, C1, A2, B2, E, D2, and G). Protein QTL were found on chromosomes 2, 4, and 15 (LGs D1b, C1, and E). More detailed information on the positions of these QTL, their effects, and the population and environment in which they were found is given in table 2.5.

2.4.5 Comparison with Past QTL

To allow for comparison with previously detected QTL, the markers used in this study were projected onto the GmConsensus 4.0 map according to the physical positions of each marker in the Glyma.Wm.82.a2 reference genome. The nearest flanking markers present on the consensus map where then found using these physical positions to estimate genetic intervals for each QTL on the consensus map.

The QTL which were found in both populations and at multiple locations have been found in similar genetic regions in previous QTL mapping studies. A QTL was found on chromosome 14 for seed oil content on chromosome 14 (LG B2) between 16.77 - 18.77 cM by Han et al.[Han15]. Many studies have identified the large effect QTL associated with protein and oil content on chromosome 20 (LG I)[Seb00; Die92b; Chu03; Bol10; Mia17]. The oil QTL detected at approximately 74 cM on chromosome 15 is located near QTL previously found to be associated with seed set, and near a single oil QTL. The closest oil QTL to this position is found at approximately 84.62 - 85.15 cM[Mao13]. As such, this QTL may represent a novel QTL for seed oil content which is stable across genetic backgrounds and

environments.

QTL for flooding yield index[Ngu12], seed abortion[Tis03], and seed oil[Wan12] have been found near the seed weight QTL found on chromosome 13 (LG F). QTL for seed set have also been found in a similar region to the seed weight QTL found in this study on chromosome 15[Nin18]. Confirmed QTL for seed weight and yield have been identified in the region of the seed weight QTL identified on chromosome 20 (LG I)[Nic06; Chu03].

Previously, oil QTL have been found on chromosome 2 (LG D1b) between 36 cM - 38 cM[Kab04; Rei06a], compared to the QTL detected in this study in population 202 at 30 cM. Given the proximity, it cannot be conclusively said given the current data that these represent different QTL. The seed oil QTL on chromosome 17 at approximately 45 cM was co-located with previously described QTL for seed protein and yield[Pan06c; Rei06a; Pan06a]. The oil QTL detected on chromosome 18 is co-located with previously described QTL for oil, protein, and seed weight[Rei06b; Mao13; Li11; Ako14; Kab04; Fas04; Rei06a].

The closest QTL listed on soybase to the protein QTL found at 72 cM on chromosome 15 is a oil QTL found between 84.62 - 85.15 cM[Mao13]. This is the only previously described seed oil or protein QTL that has been found in the region of this stable OIL qtl found in this stufy. This QTL is however co-located with the oil and seed weight QTL which was found in this study. Previous studies have mapped a QTL with a large effect on protein concentration to the genetic region between 29-31 cM on chromosome 15. This is a confirmed QTL for protein content, but QTL detected in this genetic region were only found in a single environment and population in the current study. An oil QTL was found at approximately 31 cM in Mapping population 201 in the Plymouth environment, and a protein QTL was found at approximately 30 cM in the same population in the Caswell environment.

QTL which are associated with only seed oil or seed protein content are valuable for their potential application in MAS. These QTL are relatively rare as seed protein and oil tend to be inversely correlated with one another. In this study, many of the QTL for seed protein and oil were co-located with one another. These include the QTL on chromosomes

14 and 20 which were detected in both populations, and in multiple environments. The QTL for seed oil which was found at approximately 74 cM in both populations, and in multiple environments was not co-located with a QTL for protein content which was found in both populations and environments. However, this QTL is co-located with a protein QTL which was detected in population 201 in the Caswell and combined environments. A protein QTL at this position was not found in population 202. These findings would suggest that the utility of markers linked to this QTL for selecting lines with high oil content, and not severely reduced protein content will be subject to the genetic background in which the selection is done. Additionally, given that this is a tentative novel QTL for both protein and oil content, it would be prudent to validate the findings in additional environments with denser sets of markers before a conclusive statement of the utility of this QTL for MAS can be made.

As stated previously, the QTL which was found at approximately 42 cM on chromosome 18 for seed oil content was co-located with previously described QTL for seed palmitic acid, seed stearic acid, seed oil, and seed protein although the density of such QTL in this region is relatively low when compared to other regions in the soybean genome as may be seen in figures A.41 - A.65 in the appendix. This QTL was however only found to be associated with seed oil content in the current study, although it was only detected in population 202. This QTL was detected in the Caswell, Clayton, and combined environments. It explained a moderate (6.4% - 7.4%) portion of the variation in seed oil content, and was not co-located with a protein QTL. Given the scarcity of such QTL, further investigation of this genetic region would be worthwhile to uncover additional linked markers.

2.5 Discussion

To be effective in a MAS breeding program, there should be evidence that the QTL under consideration can have a significant effect on a phenotype under diverse environments,

and across genetic backgrounds. In this study, the large effect QTL on chromosome 20 was detected for both protein and oil traits. This QTL has been found in many mapping studies for oil and protein content and it's validation in the current study is further evidence of it's stability. The QTL found in this study at approximately 72 cM on chromosome 15 is interesting in that it is not found near previously reported QTL for oil content, but was found in both populations, and in multiple environments. Furthermore, QTL were found in this region for both seed weight and protein content in multiple environments, but only a single population. This QTL may represent a novel, and stable QTL for protein and oil content and would merit further investigation.

Many of the QTL found in this study were found in only one of the mapping populations, but were detected in multiple environments. These QTL have been found in similar genetic regions in many previous mapping studies and as such, would likely still be good considerations for inclusion in a marker assisted breeding scheme. One such QTL is the QTL at approximately 42 cM on chromosome 18. This QTL was detected in multiple environments and was found in a genetic region with relatively few previously reported protein QTL and was not found to be associated with protein content in the current study. The QTL for oil content at approximately 45 cM on chromosome 17 is another such candidate as it was detected in both the Clayton and combined environments and was not found to be co-located with a protein QTL in this study. Furthermore, at the time of writing no seed protein QTL have been reported near (± 20 cM) this position, but it is co-located with a previously reported seed linoleic QTL[Ha14]. The QTL near 72 cM on chromosome 15 is another potential candidate as it was found to be stable across both genetic backgrounds and environments. This QTL was however associated with a decrease in protein content in multiple environments in population 201, but not population 202. Given that this QTL was also found in a genetic region with few previously reported QTL, further investigation with additional markers would be wise before it's use for MAS. Such studies would also allow for further exploration of the observed effect of the genetic background on the association of

this QTL with seed protein and seed oil contents.

Many of the QTL found for these three traits were co-located with one another suggests some degree of pleiotropy or linkage among the genes which control each. The evidence for overlap in the genetic control of these traits matches well with the well-documented phenotypic correlation among these traits which was observed in this study as well. A precise definition of whether this correlation is due to pleiotropy or close linkage will only be established through the use of a denser set of markers in a fine mapping strategy and cannot be conclusively determined with the current data.

2.6 Conclusions

Many of the QTL found in this study match well with previously found QTL for the same traits, and a tentative novel QTL on chromosome 15 (LG E) was found for seed oil in multiple environments and populations. QTL for seed oil content which were not co-located with protein QTL were found on chromosomes 17 and 18 in a single mapping population and in multiple environments. These QTL can be extremely useful to break the inverse correlation between seed protein and seed oil contents and should be investigated further to identify additional linked markers. The QTL found here which are located in similar regions as previous findings provide additional evidence for the stability of these QTL across genetic and environmental backgrounds which is important when considering the incorporation of these QTL in marker assisted breeding programs. The detection of a novel QTL on chromosome 15 in both populations and in multiple environments which was associated with all three traits should be followed by a validation of the findings of this study.

Table 2.1 Summary data for the linkage map created for soybean oil mapping population 201 using 421 SNPs genotyped using the Illumina Infinium SoySNP6K BeadChip.

Chromosome[a]	Number of Markers[b]	Chromosome size (cM)[c]	Average marker spacing[d]
1 (D1a)	30	146.36	5.05
2 (D1b)	30	180.45	6.22
3 (N)	20	92.91	4.89
4 (C1)	34	158.68	4.81
5 (A1)	17	86.75	5.42
6 (C2)	23	152.63	6.94
7 (M)	27	95.54	3.67
8 (A2)	14	38.63	2.97
9 (K)	19	104.08	5.78
10 (O)	13	77.8	6.48
11 (B1)	6	88.91	17.78
12 (H)	20	114.98	6.05
13 (F)	14	92.46	7.11
14 (B2)	37	116.05	3.22
15 (E)	27	138.62	5.33
16 (J)	13	89.1	7.43
17 (D2)	9	109.19	13.65
18 (G)	26	131.6	5.26
19 (L)	22	83.72	3.99
20 (I)	20	64.73	3.41
Mean	21.1	108.16	6.27
Total	421	2163.18	

[a] Chromosme number and the corresponding linkage group name for each linkage group in the linkage map.

[b] The number of SNP markers which were positioned on each linkage group.

[c] The size of each linkage group in centimorgans. Genetic distances were calculated using Kosambi's function.

[d] The average spacing betweeb each marker of a linkage group in centimorgans.

Table 2.2 Summary data for the linkage map created for soybean oil mapping population 202 using 416 SNPs genotyped using the Illumina Infinium SoySNP6K BeadChip.

Chromosome[a]	Number of Markers[b]	Chromosome size (cM)[c]	Average marker spacing[d]
1 (D1a)	26	158.41	6.34
2 (D1b)	33	147.03	4.59
3 (N)	26	149.48	5.98
4 (C1)	4	9.12	3.04
5 (A1)	19	111.86	6.21
6 (C2)	21	109.02	5.45
7 (M)	35	166.13	4.89
8 (A2)	29	141.57	5.06
9 (K)	14	76.41	5.88
10 (O)	21	83.46	4.17
11 (B1)	13	89.78	7.48
12 (H)	11	74.52	7.45
13 (F)	10	87.8	9.76
14 (B2)	25	159.5	6.65
15 (E)	40	162.76	4.17
16 (J)	13	73.24	6.1
17 (D2)	18	109.56	6.44
18 (G)	20	91.26	4.8
19 (L)	26	122.89	4.92
20 (I)	12	64.7	5.88
Mean	20.8	109.42	5.76
Total	416	2188.49	

[a] Chromosme number and the corresponding linkage group name for each linkage group in the linkage map.

[b] The number of SNP markers which were positioned on each linkage group.

[c] The size of each linkage group in centimorgans. Genetic distances were calculated using Kosambi's function.

[d] The average spacing betweeb each marker of a linkage group in centimorgans.

Table 2.3 Soybean QTL Found in Multiple Environments of North Carolina and Both Populations in 2019.

QTL Name[a]	Population[b]	Trait[c]	Location[d]	Chromosome[e]	LOD[f]	Variation[g]	QTL Effect[h]	Flanking (consensus)[i]	Position[j]
Oil-14-16.9	Population 201	Oil	Ply	14	12.60	17.8	0.52	BARC-065411-19443 - BARC-065327-19348	13.1 - 15.33
Oil-14-6.5	Population 202		Combined	14	5.19	6.8	0.30	BARC-065411-19443 - BARC-020561-04671	13.1 - 14.85
Oil-14-0.0			Ply	14	6.27	10.4	0.50	BARC-022025-04259 - BARC-051559-11161	11.34 - 14.48
Oil-15-107.2	Population 201		Cas	15	7.08	8.7	0.30	BARC-060905-16966 - BARC-048215-10508	71.42 - 75.71
Oil-15-107.2			Combined	15	6.43	5.0	0.16	BARC-060905-16966 - BARC-048215-10508	71.42 - 75.71
Oil-15-2.7	Population 202		Cas	15	3.76	7.2	0.26	BARC-063067-18226 - BARC-060729-16817	77.47 - 78.9
Oil-15-8.9			Combined	15	6.00	8.0	0.24	BARC-057349-14704 - BARC-030059-06795	75.71 - 76.37
Oil-20-41.7	Population 201		Cas	20	19.24	27.7	0.60	BARC-041129-07912 - BARC-020713-04700	33.21 - 38.61
Oil-20-33.8			Cla	20	26.62	31.7	0.49	BARC-010347-00606 - BARC-055303-13196	31.84 - 30.3
Oil-20-33.8			Combined	20	9.65	7.8	0.54	BARC-010347-00606 - BARC-055303-13196	31.84 - 30.3
Oil-20-0.0	Population 202		Cla	20	6.88	11.5	0.31	BARC-053115-11717 - BARC-041445-07985	49.95 - 57.45
Oil-20-0.0			Combined	20	8.05	11.0	0.32	BARC-053115-11717 - BARC-041445-07985	49.95 - 57.45
Protein-14-16.9	Population 201	Protein	Cas	14	3.86	3.7	0.41	BARC-065411-19443 - BARC-065327-19348	13.1 - 15.33
Protein-14-16.9			Ply	14	11.02	23.2	-1.45	BARC-065411-19443 - BARC-065327-19348	13.1 - 15.33
Protein-14-6.5	Population 202		Combined	14	7.37	9.5	-0.73	BARC-065411-19443 - BARC-020561-04671	13.1 - 14.85
Protein-14-24.2			Ply	14	9.37	17.8	-1.53	BARC-017877-02411 - BARC-050249-09527	15.32 - 38.37
Protein-20-1.8			Cas	20	3.78	8.6	-0.52	BARC-038869-07364 - BARC-025815-05092	55.3 - 87.03
Protein-20-0.0			Cla	20	8.89	16.7	-0.74	BARC-053115-11717 - BARC-041445-07985	49.95 - 57.45
Protein-20-0.0			Combined	20	11.55	15.7	-0.71	BARC-053115-11717 - BARC-041445-07985	49.95 - 57.45
Protein-20-41.7	Population 201		Cas	20	9.44	9.8	-1.05	BARC-041129-07912 - BARC-020713-04700	33.21 - 38.61
Protein-20-33.8			Cla	20	26.91	40.1	-1.06	BARC-010347-00606 - BARC-055303-13196	31.84 - 30.3
Protein-20-33.8			Combined	20	27.25	38.8	-1.04	BARC-010347-00606 - BARC-055303-13196	31.84 - 30.3
Protein-20-28.3			Ply	20	3.52	6.7	-0.79	BARC-040841-07852 - BARC-049639-09085	27.14 - 29.56

[a] QTL name in the format Trait-Chromosome-Linkage map position
[b] The population which the QTL was found in
[c] The QTL phenotype. Protein and oil are measured on a dry basis, SDWT has units of grams.
[d] The environment in which the QTL was found. Cla = Clayton, Cas = Caswell, Ply = Plymouth.
[e] The chromosome on which the QTL was found
[f] Logarithm of odds for the QTL
[g] Proportion of the phenotypic variation explained by the QTL
[h] Additive effect explained by the QTL. A positive value indicates a higher phenotypic mean was observed for the N09-09 and N13-47 parental alleles for populations 201 and 202 respectively. A negative value indicates a higher mean was observed in the alleles donated by the LMN09-119 and LMN09-19 parents.
[i] Flanking markers on the GmConsensus 4.0 map
[j] The position of flanking markers on the GmConsensus 4.0 map in centimorgans

Table 2.4 Soybean QTL Found in Multiple Environments of North Carolina and a Single Population

QTL Name[a]	Population[b]	Trait[c]	Location[d]	Chromosome[e]	LOD[f]	Variation[g]	QTL Effect[h]	Flanking (consensus)[i]	Position)[j]
Seed Weight-13-0.0	Population 202	Seed Weight	Cas	13	7.34	13.6	-0.57	BARC-065495-19507 - BARC-055499-13329	51.9 - 61.35
Seed Weight-13-9.7			Combined	13	8.04	13.8	-0.50	BARC-013257-00462 - BARC-038503-10136	65.12 - 66.91
Seed Weight-15-107.2	Population 201			15	3.63	4.5	-0.39	BARC-060905-16966 - BARC-048215-10508	71.42 - 75.71
Seed Weight-15-8.9	Population 202		Ply	15	3.45	6.0	-0.48	BARC-057349-14704 - BARC-030059-06795	75.71 - 76.37
Seed Weight-20-41.7	Population 201		Cas	20	7.40	11.0	0.66	BARC-041129-07912 - BARC-020713-04700	33.21 - 38.61
Seed Weight-20-33.8			Combined	20	4.49	5.7	0.41	BARC-010347-00606 - BARC-055303-13196	31.84 - 30.3
Seed Weight-20-49.4			Ply	20	3.27	5.0	0.48	BARC-011685-00328 - BARC-041717-08071	35.51 - 39.32
Oil-2-12.2	Population 202	Oil	Combined	2	3.47	4.5	0.22	BARC-016573-02145 - BARC-056237-14178	29.98 - 30.67
Oil-2-12.2			Ply	2	5.60	9.2	0.41	BARC-016573-02145 - BARC-056237-14178	29.98 - 30.67
Oil-17-3.7	Population 201		Cla	17	4.96	4.4	0.22	BARC-048381-10555 - BARC-031145-07005	41.55 - 50.41
Oil-17-0.0			Combined	17	5.40	4.1	0.19	BARC-048381-10555 - BARC-031145-07005	41.55 - 50.41
Oil-18-66.2	Population 202		Cas	18	3.41	6.5	0.29	BARC-039993-07626 - BARC-027814-06681	40.81 - 45.11
Oil-18-66.2			Cla	18	3.97	6.4	0.21	BARC-039993-07626 - BARC-027814-06681	40.81 - 45.11
Oil-18-66.2			Combined	18	5.61	7.4	0.25	BARC-039993-07626 - BARC-027814-06681	40.81 - 45.11
Protein-2-0.0	Population 202	Protein	Cla	2	3.93	6.9	-0.43	BARC-020481-04638 - BARC-018187-02537	25.11 - 27.01
Protein-2-12.2			Combined	2	5.41	6.8	-0.54	BARC-016573-02145 - BARC-056237-14178	29.98 - 30.67
Protein-2-12.2			Ply	2	4.38	7.8	-0.95	BARC-016573-02145 - BARC-056237-14178	29.98 - 30.67
Protein-15-107.2	Population 201		Cas	15	8.04	8.2	-0.79	BARC-060905-16966 - BARC-048215-10508	71.42 - 75.71
Protein-15-107.2			Combined	15	6.98	7.5	-0.45	BARC-060905-16966 - BARC-048215-10508	71.42 - 75.71

[a] QTL name in the format Trait-Chromosome-Linkage map position
[b] The population which the QTL was found in
[c] The QTL phenotype. Protein and oil are measured on a dry basis, SDWT has units of grams.
[d] The environment in which the QTL was found. Cla = Clayton, Cas = Caswell, Ply = Plymouth.
[e] The chromosome on which the QTL was found
[f] Logarithm of odds for the QTL
[g] Proportion of the phenotypic variation explained by the QTL
[h] Additive effect explained by the QTL. A positive value indicates a higher phenotypic mean was observed for the N09-09 and N13-47 parental alleles for populations 201 and 202 respectively. A negative value indicates a higher mean was observed in the alleles donated by the LMN09-119 and LMN09-19 parents.
[i] Flanking markers on the GmConsensus 4.0 map
[j] The position of flanking markers on the GmConsensus 4.0 map in centimorgans

Table 2.5 Soybean QTL Found in a Single Environment of North Carolina and a single Population

QTL Name[a]	Population[b]	Trait[c]	Location[d]	Chromosome[e]	LOD[f]	Variation[g]	QTL Effect[h]	Flanking (consensus)[i]	Position)[j]
Seed Weight-9-9.8	Population 202	Seed Weight	Ply	9	4.23	7.4	0.26	BARC-041991-08155 - BARC-018259-03533	31.57 - 42.64
Seed Weight-12-0.0	Population 201		Cas	12	3.73	5.3	0.44	BARC-047839-10407 - BARC-050539-09728	6.17 - 5.76
Seed Weight-14-16.9			Ply	14	10.53	17.6	-0.64	BARC-065411-19443 - BARC-065327-19348	13.1 - 15.33
Oil-2-122.5	Population 202	Oil	Combined	2	3.71	4.8	0.23	BARC-055839-13759 - BARC-064837-18812	90.55 - 94.41
Oil-4-153.2	Population 201		Cla	4	4.11	3.6	0.16	BARC-054297-12455 - BARC-032045-07244	103.42 - 110.22
Oil-8-0.0			Combined	8	3.10	2.3	0.16	BARC-022013-04250 - BARC-037229-06749	30.83 - 29.75
Oil-14-69.4			Cla	14	3.47	3.0	-0.04	BARC-044549-08718 - BARC-064873-18956	40.03 - 45.46
Oil-15-33.7			Ply	15	4.37	5.5	-0.26	BARC-029599-06237 - BARC-029771-06336	28.33 - 34.43
Oil-17-26.8			Cas	17	3.57	4.2	0.27	BARC-055705-13616 - BARC-054153-12360	22.64 - 34.27
Oil-18-95.8			Combined	18	4.25	3.2	0.11	BARC-041331-07965 - BARC-010491-00654	80.96 - 93
Protein-15-37.1		Protein	Cas	15	3.31	3.2	-0.07	BARC-029599-06237 - BARC-041187-07932	28.33 - 31.26
Protein-4-143.2				4	3.92	3.8	-0.26	BARC-041539-08030 - BARC-024463-04896	100.42 - 102.48
Protein-2-147.0	Population 202		Combined	2	3.80	4.6	-0.41	BARC-021561-04146 - BARC-013983-01297	95.45 - 103.6

[a] QTL name in the format Trait-Chromosome-Linkage map position
[b] The population which the QTL was found in
[c] The QTL phenotype. Protein and oil are measured on a dry basis, SDWT has units of grams.
[d] The environment in which the QTL was found. Cla = Clayton, Cas = Caswell, Ply = Plymouth.
[e] The chromosome on which the QTL was found
[f] Logarithm of odds for the QTL
[g] Proportion of the phenotypic variation explained by the QTL
[h] Additive effect explained by the QTL. A positive value indicates a higher phenotypic mean was observed for the N09-09 and N13-47 parental alleles for populations 201 and 202 respectively. A negative value indicates a higher mean was observed in the alleles donated by the LMN09-119 and LMN09-19 parents.
[i] Flanking markers on the GmConsensus 4.0 map
[j] The position of flanking markers on the GmConsensus 4.0 map in centimorgans

CHAPTER

3

ASSESSMENT OF SEED YIELD IN TWO MAPPING POPULATIONS

3.1 Abstract

Seed yield is one of the most important traits considered when releasing a new soybean, [Glycine max (L.) Merr.] variety. However, high seed protein and oil content is also desirable as meal and oil products produced from these seeds is more valuable. Because of this, it is important to consider these seed composition traits along with yield when selecting lines for further evaluation, and potential release. This is difficult in practice due to the negative correlation between yield and protein. In this study two yield trials were conducted from each of two recombinant inbred line (RIL) populations which were segregating for protein,

oil, and yield traits. Thirty lines were identified with increased, or comparable yield when compared to high yielding check cultivars. Additionally, these lines have a high average protein and oil content when compared to available checks. These lines represent the first stage in an ongoing process to produce high-yielding cultivars with increased protein content. These selections will be evaluated in additional environments in the 2020 growing season to assess the stability of the protein, oil, and yield content of these lines in different environments.

3.2 Introduction

Seed yield, oil, and protein are all valuable traits in a soybean variety, however breeding lines which have both high yield and protein has been difficult to develop due to the negative correlation between the two traits[Bur87; Har72; Cha15]. While considerable efforts have been made to identify loci which control these seed quality traits so that MAS breeding strategies can be utilized for their improvement, to date the applications of such markers have been few. This is largely due to the lack of markers which are uniquely associated with one trait, and are also stable across genetic and environmental backgrounds. While there is still reason to continue this genetic research, it is important that breeders take every opportunity to identify lines with both high yield, and seed composition traits like oil and protein content so that new varieties can be released.

Soybean lines typically contain about 20% oil and 40% protein content on a dry weight basis[Pat17]. The market for soybean meal requires 47.5% protein content in the meal, which corresponds to approximately 41.5% protein content on a dry weight basis[Pat17]. Oil and protein content are two of the most important seed composition traits in soybean so if one is decreased, the other should be correspondingly increased to account for the loss in value. The inverse correlation between protein and oil contents is well known and is suspected to be due at least partially to the action of pleiotropic genes and competing

metabolic pathways which control the expression of each trait[Gup17].

Despite the difficulty in simultaneously breeding for all three of these traits, releases of varieties with elevated protein contents and only moderately reduced yield has shown that it is not impossible. The high protein germplasm lines R05-1415 and R05-1772 were released recently and contain 46.9% and 46.1% protein content with while still producing yields 94% and 91% of that of the high yielding 5002T cultivar[Che11]. Lines TN03-350 and TN04-5321 contain 43.9% and 43.1% protein content while having superior or comparable performance to yield checks[Pan06d]. The Prolina cultivar has a protein content of 46.1% with a yield only 13% reduced from the Centennial check[Bur99]. The Highpro1 cultivar was released in 2016 and has a yield which is 97% of that of the highest yielding check cultivar, IA3023 with a protein content of 40.1%[Mia17]. Cultivars produced through conventional breeding techniques such as these have shown that it is possible to identify lines with both high seed protein and seed yield. Efforts to find these lines should be continued to provide growers and breeders with additional high value cultivars, and germplasm which can be further used to improve protein and yield traits.

To meet this goal, the same mapping populations from the previous chapter were screened for lines which showed promising combinations of yield and seed composition traits. A preliminary selection of lines with high values for yield, protein, and oil traits has been performed between 2018-2019 for these two populations and thirty lines have been identified which have the potential of being a valuable genetic resource for improvement of these three traits simultaneously. Further evaluation of this selection is currently being performed at additional environments to assess the stability of these preliminary results.

3.3 Experimental Design and Methods

3.3.1 Plant Materials

Two recombinant inbred line populations, named populations 201 and 202 were formed from the crosses LMN09-119 x N09-09 and LMN09-19 x N13-47, respectively. These lines were initially grown in un-replicated plant rows in the Central Crops research station in Clayton, NC in 2018.

Population 201 was developed from a cross of LMN09-119 x N09-09 in 2015. LMN09-119 was from a cross of N6202[CJ10] x G98SF-114, G98SF-114 from a 'Benning'[Boe97] x 'Danbaekkong'[Kim96] cross; N09-09 was from a cross of N02-70 x G98-1420, N02-70 was from 'Santee'[Shi03] x 'Holladay'[Bur96], G98-1420 was from the cross of 'Boggs'[Boe00] x 'Doles'[Boe94]. Population 202 was developed from a cross of LMN09-19 x Nl3-47 also in 2015. LMN09-19 is a sister line with the same pedigree as LMN09-119; N13-47 was derived from a recurrent selection breeding started with intermating F1 progeny from eight crosses in 1965 at the ARS Soybean Research Unit in Raleigh, NC. These eight crosses ('Arksoy' x 'Lee', Arksoy x 'Ogden', Arksoy x D60-8107, 'Jackson' x Lee, Jackson x Ogden, Jackson x D60-8107, 'Roanoke' x Ogden, and Roanoke x D60-8107) were made in 1964. The male sterile female fertile line with the ms1 gene was used in this recurrent selection breeding scheme over many cycles of selections. In each cycle the top 10% high oil lines were selected from the progeny to start the next cycle. The N13-47 with desirable agronomic attributes and high (>24%) oil content was selected from from a plant row in 2013. The RILs of the two mapping populations were advanced to F5 generation by single seed descent (SSD) method of selection. A single plant was harvested to represent each F5-derived RIL.

3.3.2 Experimental Design

3.3.2.1 Development of 2019 Yield Tests

In 2018, oil mapping populations 201 and 202 were grown at the Central Crops Research Station in Clayton, NC. These populations consisted of 273 and 237 recombinant inbred lines (RILs) respectively. Several agronomic traits were scored in the field for each population. These included height, lodging, maturity date, and a composite agronomic score. Lodging was scored on a scale of 1-5 where 5 indicates that all plants in a plot are on the ground, and a score of 1 indicates that all plants are erect[Feh87]. The agronomic score aimed to capture other traits of value such as visual estimation of pod load and plot uniformity to provide a general score of a line's agronomic desirability. Agronomic score was recorded on a scale of 1-5 as well, with 1 identifying the best lines of a population, and 5 the worst. Maturity was recorded at the R8 maturity date[Feh77]. Height was measured in centimeters from the soil to the top of the plant. Following harvest, yield, seed weight, protein, and oil content were measured after seed was air dried to approximately 7% moisture content in a greenhouse. Protein and oil contents were measured on a dry basis using a Perten DA 7250 NIR®instrument. Yield and seed weight were measured after seed had been sifted and cleaned of debris and cracked seed.

To select lines for the 2019 growing season, lines with abnormally low bulk weights or extreme maturity dates from 2018 were first removed from consideration. Two yield trials were then developed for each mapping population. The maturity data of RILs were considered when forming tests such that the lines of each test would have a maturity data range approximately half that of the total mapping population from which it was derived. RILs were selected for each test which were also representative of the distribution of seed protein and seed oil traits for each population. Eighty unique lines were selected from each population which satisfied these criteria, and each yield test was comprised of 40 RILs. Three high-yielding check cultivars and the two parents of the respective population were

also included in each test. Yield check cultivars Dunphy, Osage[Che07a], and Roy[Bur05] were used in tests 1 and 2, while Dunphy, Dilday, and N.C. Raleigh[Bur06] were used for tests 3 and 4. These lines were selected to represent the estimated maturities of the RILs in each test. The parents for tests 1 and 2 were cultivars LMN09-119 and N09-09, and the parents for tests 3 and 4 were LMN09-19 and N13-47. A tabulation of phenotypic data for each of these tests from measurements taken in 2018 is shown in table 3.1.

3.3.3 Plot Techniques

Each test was grown at the Tidewater Research Station in Plymouth, NC, and the Caswell Research Farm in Kinston, NC in 2019. These locations will be referred to as Plymoth, (PLY), and Caswell, (CAS). Lines were grown in a randomized complete block design with three replications at each location. The yield plots consisted of three rows on 96.5 cm row spacing with 6.7 m row length. The plots were end trimmed to 4.57 m prior to harvesting at or near maturity. Seeds were planted at a density of approximately 31 seeds meter-1 in these plots. Best management practices were used throughout the growing season.

3.3.4 Phenotypic Data Collection

Each test was scored for the same traits as the plant rows in 2018 using the same scoring schemes. Yield, seed weight, protein, and oil content were measured following seed drying to approximately 7% moisture content in a greenhouse. Seed traits were measured after each sample had been cleared of debris and cracked seeds by manual sifting. Seed protein and seed oil contents were measured on a dry basis using a Perten DA 7250®near infrared spectrometer.

3.3.5 Statistical Methodology

Phenotypes for each line were analyzed using the linear model:

Table 3.1 2018 Soybean plant row data summary statistics for the RILs in the yield tests in North Carolina.

Test	Trait	Average	SD	Range
Yield Test 1	Maturity Date	48.35	1.98	44 - 50
	Height	31.05	3.45	22 - 36
	Lodging	1.64	0.28	1 - 2
	Agronomic Score	2.38	0.40	1.5 - 3
	Hundred Seed Weight	15.66	1.11	13.67 - 18.11
	Bulk Weight	963.08	192.46	643.9 - 1391.2
	Protein (dry basis)	44.84	2.05	40.13 - 48.29
	Oil (dry basis)	22.28	0.93	20.52 - 24.53
Yield Test 2	Maturity Date	51.77	1.14	50 - 54
	Height	32.95	2.56	28 - 40
	Lodging	1.64	0.25	1 - 2
	Agronomic Score	2.39	0.35	2 - 3
	Hundred Seed Weight	16.43	1.17	14.65 - 19.75
	Bulk Weight	1024.86	277.47	663.6 - 1743
	Protein (dry basis)	44.45	2.18	41.23 - 49.37
	Oil (dry basis)	22.25	1.15	20.04 - 24.36
Yield Test 3	Maturity Date	55.58	2.98	49 - 59
	Height	35.35	3.40	28 - 42
	Lodging	1.66	0.24	1.5 - 2
	Agronomic Score	2.54	0.38	2 - 3
	Hundred Seed Weight	17.58	1.86	13.35 - 21.85
	Bulk Weight	1005.25	183.95	741.2 - 1552.3
	Protein (dry basis)	45.61	2.11	41.5 - 49.28
	Oil (dry basis)	22.74	1.21	19.86 - 25.6
Yield Test 4	Maturity Date	62.33	2.57	59 - 70
	Height	37.98	3.70	30 - 45
	Lodging	1.59	0.19	1.5 - 2
	Agronomic Score	2.54	0.31	2 - 3
	Hundred Seed Weight	18.48	2.31	11.32 - 23.11
	Bulk Weight	1081.14	215.14	749.7 - 1609.5
	Protein (dry basis)	45.43	2.53	39.49 - 50.47
	Oil (dry basis)	22.48	1.07	19.92 - 24.62

$$y_{ijk} = \mu + L_i + B(L_i) + G_k + GL_{ik} + \epsilon_{ijk}$$

Where y_{ijk} is the phenotypic measurement for rep j of genotype k in environment i, L_i is the effect of location i, G_k is the effect of genotype G, GL_{ik} is the interaction effect of location L and genotype G, and ϵ_{ijk} is the measurement error.

Models were fit using the lm function in R, and analysis of variance (ANOVA) performed using the *anova* function. Least-square means (LS Means) for each genotype and trait were calculated using the above model using the *emmeans* package in R. Least-significant difference values for each trait were calculated using the *LSD.test* function from the *agricolae* package. Pearson correlation coefficients between each phenotype were calculated with the *cor* function in R.

Pearson correlation is calculated for each pair of traits as:

$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}}$$

Where x and y are measurements of the two phenotypes, m_x and m_y are the means of each phenotype, and r is the correlation coefficient.

3.4 Results and Discussion

3.4.1 Phenotypic Evaluation

An analysis of variance for hundred seed weight, yield, height, protein, oil, and protein + oil traits was performed for each test. The effect of genotype was found to be significant across all traits and tests.

All traits were approximately normally distributed and on average, lines from tests 1 and 2 outperformed those from tests 3 and 4 in terms of both yield and agronomic traits. Plots of trait distributions for each test may be found in the appendix in figures A.2 - A.25. Seed

protein and seed yield were significantly negatively correlated in all yield tests. Pearson correlation coefficients for seed yield and seed protein ranged from -0.44 to -0.59 across tests. Seed oil and seed yield correlation coefficients were positive and significant in all tests except for test 4 which had a non-significant, positive correlation of 0.24. Significant correlation coefficients between seed oil and seed yield ranged between 0.57 and 0.63 for tests 1-3. Trait pairwise correlation coefficients, and ANOVA tables may be found in the appendix in figures A.32-A.36 and tables A.4-A.27, respectively.

Data from the top ten lines from each yield test, and the checks of each test may be found below in tables 3.2-3.5. Dunphy, Osage, and Roy cultivars were included in tests 1 and 2 as checks while Dunphy, Dilday, and N.C. Raleigh were used as checks in tests 3 and 4. The parents of each population were also included in each test. The parents of the RILs in tests one and two LMN09-119 and N09-09, the parents of RILs on tests three and four N13-47 and LMN09-19. It can be seen in tables 3.2 and 3.3 that there were RILs which exceeded the seed yield of some, or all of the checks which were included in tests 1 and 2. This was not the case in tests 3 and 4 where the lines tended to under-perform the checks in terms of yield.

The top ten yielding lines of tests 1 and 2 also still had relatively high seed protein and oil content, 40-46% and 21-23% on a dry basis for seed protein and seed oil, respectively. Transgressive segregation for protein, oil, and yield was observed for lines in tests 1 and 2. Transgressive segregation was observed for seed yield in tests 3 and 4, but little was observed for protein and oil traits. This observation was not unexpected as the parents for population 201 tended to have similar seed protein and seed oil contents when compared with the parents of population 202 which had more extreme seed protein and seed oil contents. The distribution and the parental phenotypic values for each trait may be seen in figures A.12 - A.31 in the appendix.

Several RILs had seed yield measurements comparable to, or exceeding those of the check cultivars for their respective tests. Additionally, several of these lines had superior

seed composition traits, most notably seed protein content given the negative correlation between seed protein and yield. The majority of these lines are from yield tests one and two. These two yield tests are comprised of RILs drawn from mapping population 201. On average, tests one and two were also superior to yield tests three and four in terms of agronomic traits. This may be seen in the phenotypic data presented in table 3.6. The data from 2019 was used to identify lines with yield comparable to, or exceeding that of the check cultivars which also had high protein or oil contents. The procedure which was used to select these lines is described in the following section.

Top 10 Lines: Soybean Yield Test 1, North Carolina, 2019. The 10 RILs presented in the table are the 10 highest yielding lines from the respective yield test, according to the least-square mean estimate of yield for the RIL. Data for the checks and parent cultivars are included at the bottom of the table for comparison.

Table 3.2 Top 10 Lines: Soybean Yield Test 1, North Carolina, 2019. The 10 RILs presented in the table are the 10 highest yielding lines from the respective yield test, according to the least-square mean estimate of yield for the RIL. Data for the checks and parent cultivars are included at the bottom of the table for comparison.

Test[a]	Genotype[b]	Yield[c]	Oil[d]	Protein[e]	Protein + Oil[f]	Check[g]
Yield Test 1	N18-1669	2522.329	22.63	42.23	64.86	RIL
	N18-1580	2511.343	23.01	40.22	63.23	
	N18-1803	2485.999	21.95	43.14	65.08	
	N18-1661	2416.092	23.25	42.02	65.27	
	N18-1643	2409.855	22.60	43.26	65.87	
	N18-1820	2387.729	22.89	42.51	65.40	
	N18-1796	2382.025	21.29	46.27	67.55	
	N18-1769	2324.479	21.09	46.57	67.66	
	N18-1807	2320.751	21.87	44.37	66.24	
	N18-1627	2319.375	22.24	44.58	66.82	
	Dunphy	2397.649	22.99	39.71	62.69	Check
	Osage	2235.153	21.95	43.73	65.68	
	Roy	2224.567	21.90	42.53	64.43	
	N09-09	2108.432	21.38	45.92	67.29	
	LMN09-119	2068.640	20.63	48.42	69.05	
LSD		**216.440**	**0.49**	**0.95**	**0.61**	

[a] 2019 yield test
[b] The genotype identifier
[c] Yield in kg/ha
[d] Oil on a % dry basis
[e] Protein on a % dry basis
[f] Protien + Oil Content
[g] Is the genotype a check or a RIL

Table 3.3 Top 10 Lines: Soybean Yield Test 2, North Carolina, 2019. The 10 RILs presented in the table are the 10 highest yielding lines from the respective yield test, according to the least-square mean estimate of yield for the RIL. Data for the checks and parent cultivars are included at the bottom of the table for comparison.

Test[a]	Genotype[b]	Yield[c]	Oil[d]	Protein[e]	Protein + Oil[f]	Check[g]
Yield Test 2	N18-1620	2728.167	22.31	43.01	65.33	RIL
	N18-1751	2726.103	22.32	43.52	65.84	
	N18-1641	2657.328	22.74	42.95	65.69	
	N18-1665	2497.095	22.63	41.93	64.56	
	N18-1604	2495.742	22.80	43.26	66.06	
	N18-1748	2485.888	21.11	44.75	65.86	
	N18-1632	2483.979	22.46	44.97	67.43	
	N18-1674	2477.566	22.08	44.56	66.64	
	N18-1608	2473.771	21.55	43.22	64.77	
	N18-1595	2445.142	21.33	45.60	66.93	
	Osage	2327.098	21.68	44.13	65.81	Check
	Roy	2298.114	21.77	42.93	64.70	
	N09-09	2297.449	21.03	46.63	67.66	
	Dunphy	2176.120	22.23	41.37	63.60	
	LMN09-119	1873.410	20.14	49.23	69.37	
LSD		**251.840**	**0.54**	**1.10**	**0.72**	

[a] 2019 yield test
[b] The genotype identifier
[c] Yield in kg/ha
[d] Oil on a % dry basis
[e] Protein on a % dry basis
[f] Protien + Oil Content
[g] Is the genotype a check or a RIL

Table 3.4 Top 10 Lines: Soybean Yield Test 3, North Carolina, 2019. The 10 RILs presented in the table are the 10 highest yielding lines from the respective yield test, according to the least-square mean estimate of yield for the RIL. Data for the checks and parent cultivars are included at the bottom of the table for comparison.

Test[a]	Genotype[b]	Yield[c]	Oil[d]	Protein[e]	Protein + Oil[f]	Check[g]
Yield Test 3	N18-1855	2211.895	22.36	46.23	68.60	RIL
	N18-1923	2150.820	23.25	42.92	66.17	
	N18-1978	2138.126	22.44	44.76	67.20	
	N18-1906	2110.429	22.97	44.20	67.17	
	N18-1908	2100.332	23.12	43.18	66.30	
	N18-1844	2079.071	23.32	43.83	67.15	
	N18-1944	2077.695	21.32	45.84	67.16	
	N18-1950	2075.986	23.86	43.38	67.24	
	N18-1962	2074.366	23.08	44.76	67.83	
	N18-1953	2050.664	22.15	44.65	66.81	
	Dilday	2577.655	23.83	39.96	63.80	Check
	Dunphy	2406.859	22.63	40.94	63.57	
	N.C. Raleigh	2358.279	24.55	40.18	64.73	
	LMN09-19	2037.015	19.68	50.26	69.94	
	N13-47	1991.365	25.32	39.43	64.75	
LSD		**370.810**	**0.63**	**1.06**	**0.64**	

[a] 2019 yield test
[b] The genotype identifier
[c] Yield in kg/ha
[d] Oil on a % dry basis
[e] Protein on a % dry basis
[f] Protien + Oil Content
[g] Is the genotype a check or a RIL

Table 3.5 Top 10 Lines: Soybean Yield Test 4, North Carolina, 2019. The 10 RILs presented in the table are the 10 highest yielding lines from the respective yield test, according to the least-square mean estimate of yield for the RIL. Data for the checks and parent cultivars are included at the bottom of the table for comparison.

Test[a]	Genotype[b]	Yield[c]	Oil[d]	Protein[e]	Protein + Oil[f]	Check[g]
Yield Test 4	N18-1841	2343.610	22.69	42.72	65.41	RIL
	N18-1930	2336.952	22.50	41.58	64.08	
	N18-1890	2280.027	21.57	44.89	66.46	
	N18-1904	2263.826	21.89	44.21	66.10	
	N18-1916	2230.759	22.20	43.40	65.60	
	N18-1967	2214.913	23.22	42.38	65.60	
	N18-1956	2201.531	23.22	43.34	66.55	
	N18-1979	2199.711	21.32	46.26	67.58	
	N18-1849	2186.063	22.32	42.42	64.75	
	N18-1921	2177.341	20.75	47.08	67.83	
	N.C. Raleigh	2459.634	24.18	40.08	64.26	Check
	Dunphy	2349.935	22.31	41.02	63.33	
	Dilday	2328.141	23.06	40.71	63.77	
	N13-47	2107.389	25.23	39.02	64.24	
	LMN09-19	2090.389	19.70	49.48	69.17	
LSD		**316.820**	**0.59**	**0.98**	**0.66**	

[a] 2019 yield test
[b] The genotype identifier
[c] Yield in kg/ha
[d] Oil on a % dry basis
[e] Protein on a % dry basis
[f] Protien + Oil Content
[g] Is the genotype a check or a RIL

Table 3.6 The mean, standard deviation, and range of least square means for agronomic and seed composition traits for each soybean yield test. Tests were grown in two environments in North Carolina in 2019.

Test	Trait	Average	Standard Deviation	Range
Yield Test 1	Average R8 Maturity Date	49.84	2.39	47 - 58
	Height (cm)	32.51	2.49	27 - 38.33
	Oil Content (% dry basis)	21.66	0.76	20.12 - 23.25
	Protein + Oil Content (% dry basis)	66.23	1.41	62.69 - 69.05
	Protein Content (% dry basis)	44.57	2.00	39.71 - 48.42
	SDWT	14.88	1.22	12.7 - 17.98
	Yield (kg per ha)	2240.89	124.64	1999.24 - 2522.33
Yield Test 2	Average R8 Maturity Date	55.69	3.41	48 - 62
	Height (cm)	35.05	2.92	26.67 - 43.33
	Oil Content (% dry basis)	21.56	0.90	19.82 - 23.18
	Protein + Oil Content (% dry basis)	66.38	1.17	63.6 - 69.37
	Protein Content (% dry basis)	44.82	1.84	41.37 - 49.23
	SDWT	15.86	1.24	12.67 - 18.52
	Yield (kg per ha)	2335.21	169.99	1873.41 - 2728.17
Yield Test 3	Average R8 Maturity Date	54.51	4.27	47 - 62
	Height (cm)	36.20	2.80	30.17 - 43
	Oil Content (% dry basis)	22.32	1.18	19.66 - 25.32
	Protein + Oil Content (% dry basis)	67.53	1.38	63.57 - 69.94
	Protein Content (% dry basis)	45.21	2.34	39.43 - 50.26
	SDWT	17.11	1.50	13.98 - 20.25
	Yield (kg per ha)	1965.90	212.87	1447.91 - 2577.66
Yield Test 4	Average R8 Maturity Date	61.51	2.78	54 - 67
	Height (cm)	38.41	2.42	28.83 - 43
	Oil Content (% dry basis)	21.95	1.14	18.84 - 25.23
	Protein + Oil Content (% dry basis)	66.82	1.63	63.33 - 69.61
	Protein Content (% dry basis)	44.87	2.54	39.02 - 49.99
	SDWT	17.47	1.89	11.05 - 21.02
	Yield (kg per ha)	2058.39	177.01	1634.37 - 2459.63

3.4.2 Selections for 2020

Using the phenotypic data collected from the 2019 growing season, we identified lines with both high yield as well as high protein plus oil content. This was done by first identifying lines which were within, or above an LSD of the average of the yield for the checks for each test. The lines which met this first requirement were then chosen on the basis of their overall protein + oil content such that only the top performing lines for the seed composition traits were selected.

Thirty lines were selected using these criteria. The new selections consist of 12 lines from yield test 1, 15 from yield test 2, 1 from yield test 3, and 2 from yield test 4. More lines were drawn from yield tests 1 and 2 because these tests had both higher yield, and better agronomic qualities on average than the lines from mapping population 202. More descriptive statistics for each new test may be found in table 3.7.

Table 3.7 The mean, standard deviation, and range of least square means for agronomic and seed composition traits for each soybean yield test for the 2020 growing season. Tests are being grown in two environments in North Carolina in 2020.

Test	Trait	Average	Standard Deviation	Range
2020 - Test 1	Average R8 Maturity Date	58.33	3.31	54 - 66
	Height (cm)	35.80	2.45	32.67 - 42
	Oil Content (% dry basis)	22.30	0.58	21.08 - 23.22
	Protein + Oil Content (% dry basis)	66.75	1.00	65.33 - 68.81
	Protein Content (% dry basis)	44.45	1.37	42.95 - 47.2
	SDWT	16.63	1.21	14.95 - 19.2
	Yield (kg per ha)	2423.28	181.62	2123.35 - 2728.17
2020 - Test 2	Average R8 Maturity Date	49.93	1.79	48 - 53
	Height (cm)	32.63	2.86	28.5 - 39.17
	Oil Content (% dry basis)	21.88	0.85	20.55 - 23.25
	Protein + Oil Content (% dry basis)	67.13	1.29	65.27 - 68.98
	Protein Content (% dry basis)	45.25	2.10	42.02 - 48.32
	SDWT	15.28	1.20	13.85 - 18.05
	Yield (kg per ha)	2300.89	98.17	2140.57 - 2434.67

The selected lines have on average higher yield, and similar protein and oil content to the remaining lines of the four tests. Selected lines have an average yield of 1064g, average oil content of 22.1%, and an average protein of 44.9%. Non-selected lines had an average yield of 953g, an oil content of 21.8%, and a protein content of 44.9%. These 30 lines were grouped into two new yield tests based on their average maturity dates so that each new test covered approximately half the maturity date range of the full set of selected lines. Lines in 2020 - Test 1 have an average maturity date between 54-66 while lines in 2020 - Test 2 are between 48 - 54. Yield and oil content were likely increased simultaneously in these selections due to the correlation between the two traits. The lines with the highest protein content in 2019 were not selected for advancement as they had reduced yield. It is notable that the selections did not have a severely reduced protein content due to the negative correlation between the two traits. A plot showing the distribution of protein, oil, and yield traits in the set of selected lines, compared to the lines which were not selected may be seen in figure 3.1.

These two yield trials will be grown at the central crops research station in Clayton, NC, and the Caswell Research Farm in Kinston, NC. Tests will be grown in a randomized complete block design with four replications in three-row plots with 650 seeds per plot. The same traits will be evaluated for each trial to assess the stability of the lines performance over years, and in a different set of environments.

3.5 Conclusions and Future Research

The results of this study are the first step in an ongoing effort to identify high-yielding lines with exceptional protein and oil contents from two mapping populations. Using the data from the 2019 season, we were able to identify lines which potentially have yield comparable to existing yield checks, and have improved protein and oil contents relative to these checks. It will be necessary to validate these findings by evaluating the lines at

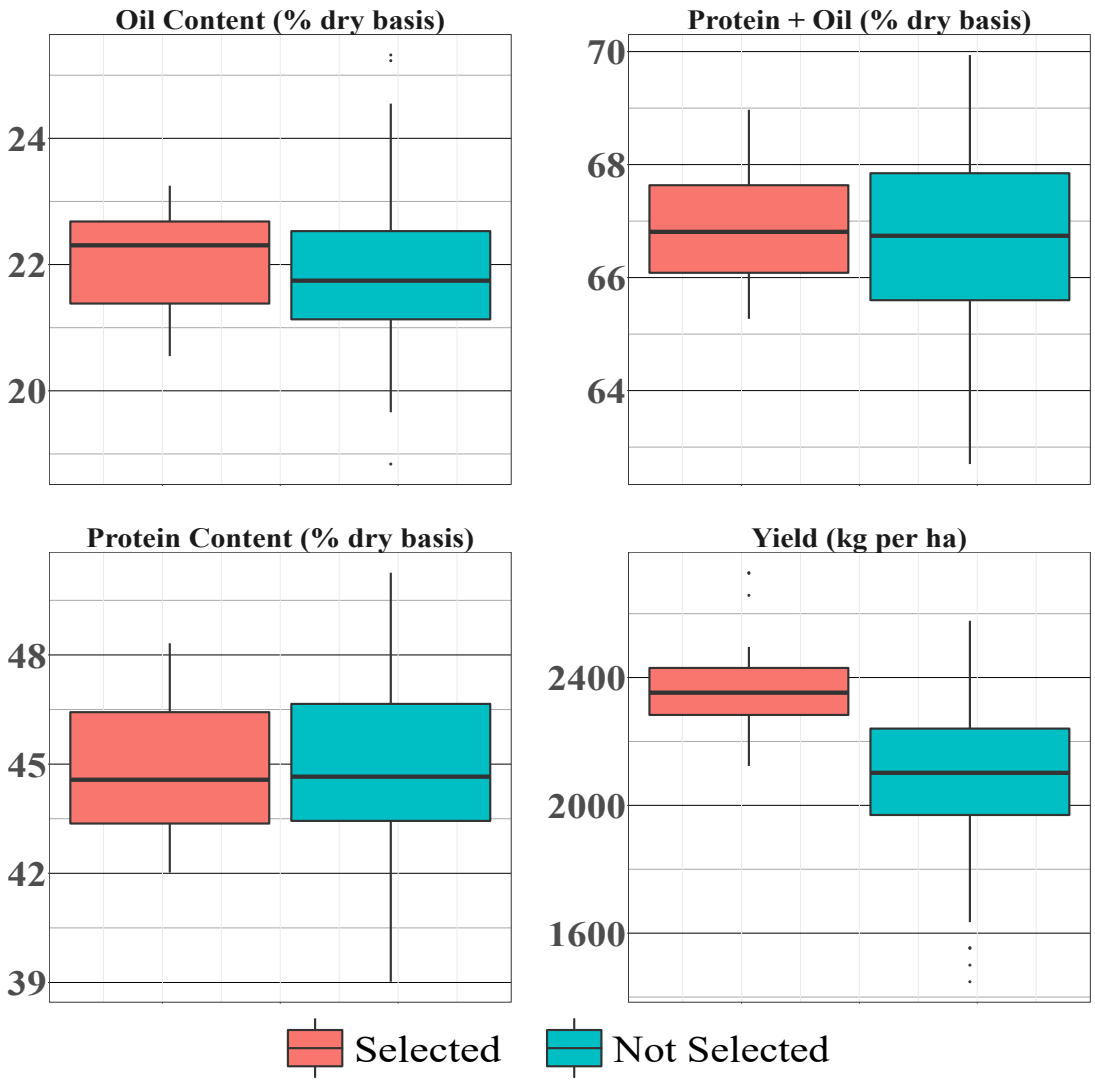

Figure 3.1 Comparison of Trait Distributions for Selected and Non-Selected Lines for the 2020 Growing Season. Protein, Oil, and Protein + Oil are measured on a % dry basis, Yield has units of kg/ha.

additional locations, and over multiple years to determine the effects of these factors on the current findings.

BIBLIOGRAPHY

[Ako14] Akond, M. et al. "Identification of quantitative trait loci (QTL) underlying protein, oil, and five major fatty acids' contents in soybean". *American Journal of Plant Sciences* **2014** (2014).

[AK19] Al-Khayri, J. et al. *Advances in Plant Breeding Strategies: Legumes: Volume 7*. Springer International Publishing, 2019.

[ASA19] ASA. *Soy Stats. A Reference Guide to Important Soybean Facts and Figures.* 2019.

[Ase16] Asekova, S. et al. "Genetic analysis of shoot fresh weight in a cross of wild (G. soja) and cultivated (G. max) soybean". *Molecular Breeding* **36**.7 (2016), p. 103.

[Bac09] Bachlava, E. et al. "Mapping and Comparison of Quantitative Trait Loci for Oleic Acid Seed Content in Two Segregating Soybean Populations". *Crop Science* **49**.2 (2009), pp. 433–442.

[Bal08] Balding, D. J. et al. *Handbook of statistical genetics*. John Wiley & Sons, 2008.

[Bas05] Basten, C. J. et al. "QTL Cartographer, version 1.17". *Department of Statistics, North Carolina State University, Raleigh* **189** (2005).

[Bat15] Batley, J. *Plant genotyping: methods and protocols*. Springer, 2015.

[Boe94] Boerma, H. et al. "Registration of 'Doles' soybean". *Crop science* **34**.5 (1994), pp. 1411–1412.

[Boe97] Boerma, H. et al. "Registration of'Benning'soybean." *Crop science* **37**.6 (1997).

[Boe00] Boerma, H. et al. "Registration of 'Boggs' soybean". *Crop science* **40**.1 (2000), pp. 294–295.

[Bol10] Bolon, Y.-T. et al. "Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean". *BMC Plant Biology* **10**.1 (2010), p. 41.

[Bri66] Brini, C. A. "A modified pedigree method of selection in soybeans". *Crop Sci* **6** (1966), p. 220.

[Bro09] Broman, K. W. & Sen, S. *A Guide to QTL Mapping with R/qtl*. Vol. 46. Springer, 2009.

[Bro03] Broman, K. W. et al. "R/qtl: QTL mapping in experimental crosses". *Bioinformatics* **19**.7 (2003), pp. 889–890.

[Bru97] Brummer, E. C. et al. “Mapping QTL for seed protein and oil content in eight soybean populations”. *Crop Science* **37**.2 (1997), pp. 370–378.

[Bur96] Burton, J. W. et al. “Registration of ‘Holladay’soybean”. *Crop science* **36**.2 (1996), pp. 467–467.

[Bur87] Burton, J. “Quantitative genetics: results relevant to soybean breeding”. *Agronomy (USA)* (1987).

[Bur99] Burton, J., WILSON, R., et al. “Registration ofProlina’Soybean”. *Crop science* **39**.1 (1999), pp. 294–295.

[Bur05] Burton, J. et al. “Registration of’NC-Roy’soybean”. *Crop science* **45**.6 (2005), pp. 2654–2655.

[Bur06] Burton, J. et al. “Registration of’NC-Raleigh’soybean”. *Crop science* **46**.6 (2006), p. 2710.

[Car09] Carrera, C. et al. “Water deficit effect on the relationship between temperature during the seed fill period and soybean seed oil and protein concentrations”. *Crop Science* **49**.3 (2009), pp. 990–998.

[CJ10] Carter Jr, T. et al. “Registration of N6202 soybean germplasm with high protein, favorable yield potential, large seed, and diverse pedigree”. *Journal of plant registrations* **4**.1 (2010), pp. 73–79.

[Cha03] Chapman, A. et al. “Quantitative trait loci for agronomic and seed quality traits in an F(2) and F(4 : 6) soybean population”. *Euphytica* **129**.3 (2003), pp. 387–393.

[Cha04] Chardon, F. et al. “Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome”. *Genetics* **168**.4 (2004), pp. 2169–2185.

[Cha15] Chaudhary, J. et al. “Expanding omics resources for improvement of soybean seed composition traits”. *Frontiers in Plant Science* **6** (2015), p. 1021.

[Che07a] Chen, P et al. “Registration of ‘Osage’soybean”. *Journal of plant registrations* **1**.2 (2007), pp. 89–92.

[Che11] Chen, P et al. “Registration of R05-1415 and R05-1772 high-protein soybean germplasm lines”. *Journal of Plant Registrations* **5**.3 (2011), pp. 410–413.

[Che07b] Chen, Q.-s. et al. “QTL Analysis of Major Agronomic Traits in Soybean”. *Agricultural Sciences in China* **6**.4 (2007), pp. 399–405.

[Chu03] Chung, J. et al. “The seed protein, oil, and yield QTL on soybean linkage group I”. *Crop Science* **43**.3 (2003), pp. 1053–1067.

[Col05] Collard, B. C. et al. “An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts”. *Euphytica* **142**.1-2 (2005), pp. 169–196.

[Cre99] Cregan, P. et al. “Two simple sequence repeat markers to select for soybean cyst nematode resistance coditioned by the rhg1 locus”. *Theoretical and Applied Genetics* **99**.5 (1999), pp. 811–818.

[Csa01] Csanadi, G. et al. “Seed quality QTLs identified in a molecular map of early maturing soybean”. *Theoretical and Applied Genetics* **103**.6-7 (2001), pp. 912–919.

[Die92a] Diers, B. W. & Shoemaker, R. C. “RESTRICTION-FRAGMENT-LENGTH-POLYMORPHISM ANALYSIS OF SOYBEAN FATTY-ACID CONTENT”. *Journal of the American Oil Chemists Society* **69**.12 (1992), pp. 1242–1244.

[Die92b] Diers, B. W. et al. “RFLP ANALYSIS OF SOYBEAN SEED PROTEIN AND OIL CONTENT”. *Theoretical and Applied Genetics* **83**.5 (1992), pp. 608–612.

[Dor92] Dornbos, D. & Mullen, R. “Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature”. *Journal of the American Oil Chemists Society* **69**.3 (1992), pp. 228–231.

[Du09] Du, W. J. et al. “Mapping QTLs for seed yield and drought susceptibility index in soybean (Glycine max L.) across different environments”. *Journal of Genetics and Genomics* **36**.12 (2009), pp. 721–731.

[Esk13a] Eskandari, M. et al. “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”. *Theoretical and Applied Genetics* **126**.2 (2013), pp. 483–495.

[Esk13b] Eskandari, M. et al. “Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield”. *Theoretical and Applied Genetics* **126**.6 (2013), pp. 1677–1687.

[Fan15] Fan, S. X. et al. “Analysis of additive and epistatic quantitative trait loci underlying fatty acid concentrations in soybean seeds across multiple environments”. *Euphytica* **206**.3 (2015), pp. 689–700.

[Fas04] Fasoula, V. A. et al. “Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations”. *Crop Science* **44**.4 (2004), pp. 1218–1225.

[Feh87] Fehr, I. *Soybeans: Improvement, Production and Uses,”*. 1987.

[Feh91] Fehr, W. *Principles of cultivar development: theory and technique*. Macmillian Publishing Company, 1991.

[Feh77] Fehr, W. R. & Caviness, C. E. "Stages of soybean development" (1977).

[Fer06] Ferreira, A. et al. "Estimating the effects of population size and type on the accuracy of genetic maps". *Genetics and Molecular Biology* **29**.1 (2006), pp. 187–192.

[Fun13] Funada, M. et al. "Single-seed descent, single-pod, and bulk sampling methods for soybean". *Euphytica* **192**.2 (2013), pp. 217–226.

[Fun05] Funatsuki, H. et al. "Mapping of QTL associated with chilling tolerance during reproductive growth in soybean". *Theoretical and Applied Genetics* **111**.5 (2005), pp. 851–861.

[Gai07] Gai, J. et al. "A comparative study on segregation analysis and QTL mapping of quantitative traits in plants—with a case in soybean". *Frontiers of Agriculture in China* **1**.1 (2007), pp. 1–7.

[Gof00] Goffinet, B. & Gerber, S. "Quantitative trait loci: A meta-analysis". *Genetics* **155**.1 (2000), pp. 463–473.

[Gra09] Grant, D. et al. "SoyBase, the USDA-ARS soybean genetics and genomics database". *Nucleic Acids Research* **38**.suppl$_1$ (2009), pp. D843–D846.

[Gup17] Gupta, M. et al. "Integration of omics approaches to understand oil/protein content during seed development in oilseed crops". *Plant Cell Reports* **36**.5 (2017), pp. 637–652.

[Guz07] Guzman, P. S. et al. "QTL associated with yield in three backcross-derived populations of soybean". *Crop Science* **47**.1 (2007), pp. 111–122.

[Ha14] Ha, B. K. et al. "Identification of quantitative trait loci controlling linolenic acid concentration in PI483463 (Glycine soja)". *Theoretical and Applied Genetics* **127**.7 (2014), pp. 1501–1512.

[Hac18] Hacisalihoglu, G. et al. "Quantitative trait loci associated with soybean seed weight and composition under different phosphorus levels". *Journal of Integrative Plant Biology* **60**.3 (2018), pp. 232–241.

[Hal19] Haldane, J. "The combination of linkage values and the calculation of distances between the loci of linked factors". *J Genet* **8**.29 (1919), pp. 299–309.

[Han15] Han, Y. P. et al. "Unconditional and conditional QTL underlying the genetic interrelationships between soybean seed isoflavone, and protein or oil contents". *Plant Breeding* **134**.3 (2015), pp. 300–309.

[Har72] Hartwig, E. E. & Hinson, K. “Association between chemical composition of seed and seed yield of soybeans 1”. *Crop science* **12**.6 (1972), pp. 829–830.

[Hne96] Hnetkovsky, N. et al. “Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS)”. *Crop Science* **36**.2 (1996), pp. 393–400.

[How53] Howell, R. W. & Cartter, J. L. “Physiological factors affecting composition of soybeans. I. Correlation of temperatures during certain portions of the pod filling stage with oil percentage in mature beans”. *Agron. J* **45**.11 (1953), pp. 526–528.

[How58] Howell, R. W. & Cartter, J. L. “Physiological factors affecting composition of soybeans: Ii. Response of oil and other constituents of soybeans to temperature under controlled conditions 1”. *Agronomy Journal* **50**.11 (1958), pp. 664–667.

[Hym05] Hymowitz, T & Shurtleff, W. “Debunking soybean myths and legends in the historical and popular literature”. *Crop science* **45**.2 (2005), pp. 473–476.

[Hyt04] Hyten, D. L. et al. “Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci”. *Journal of the American Oil Chemists Society* **81**.12 (2004), pp. 1115–1118.

[Hyt10] Hyten, D. L. et al. “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 Science* **50**.3 (2010), pp. 960–968.

[Hyt06] Hyten, D. L. et al. “Impacts of genetic bottlenecks on soybean genome diversity”. *Proceedings of the National Academy of Sciences* **103**.45 (2006), pp. 16666–16671.

[Ill11] Illumina, I. “GenomeStudio Software 2011.1”. *San Diego: Illumina* (2011).

[Jen14] Jenkinson, J. et al. *Utility of SNP markers associated with major soybean plant maturity and growth habit genomic regions*. US Patent 8,847,006. 2014.

[Jeo11] Jeong, N. et al. “Fine genetic mapping of the genomic region controlling leaflet shape and number of seeds per pod in the soybean”. *Theoretical and Applied Genetics* **122**.5 (2011), pp. 865–874.

[Jun08] Jun, T. H. et al. “Association analysis using SSR markers to find QTL for seed protein content in soybean”. *Euphytica* **162**.2 (2008), pp. 179–191.

[Kab04] Kabelka, E. A. et al. “Putative alleles for increased yield from soybean plant introductions”. *Crop Science* **44**.3 (2004), pp. 784–791.

[Kan97] Kane, M. V. et al. “Early-maturing soybean cropping system: III. Protein and oil contents and oil composition”. *Agronomy Journal* **89**.3 (1997), pp. 464–469.

[Kaz10] Kazi, S. et al. “Iso-lines and inbred-lines confirmed loci that underlie resistance from cultivar ’Hartwig’ to three soybean cyst nematode populations”. *Theoretical and Applied Genetics* **120**.3 (2010), pp. 633–644.

[Kho09] Khowaja, F. S. et al. “Improved resolution in the position of drought-related QTLs in a single mapping population of rice by meta-analysis”. *BMC genomics* **10**.1 (2009), p. 276.

[Kim12] Kim, K. S. et al. “Identification of positive yield QTL alleles from exotic soybean germplasm in two backcross populations”. *Theoretical and Applied Genetics* **125**.6 (2012), pp. 1353–1369.

[Kim16] Kim, M. et al. “Identification and fine mapping of a soybean seed protein QTL from PI 407788A on chromosome 15”. *Crop Science* **56**.1 (2016), pp. 219–225.

[Kim96] Kim, S. et al. *A new high protein and good seed quality soybean variety" Danbaegkong"*. Tech. rep. 1996.

[Kla13] Klaiber, J. A. et al. *Molecular markers associated with soybean root-knot nematode tolerance and methods of their use*. US Patent App. 13/585,956. 2013.

[Kos16] Kosambi, D. D. “The estimation of map distances from recombination values”. *DD Kosambi*. Springer, 2016, pp. 125–130.

[Kum06] Kumar, V. et al. “Influence of growing environment on the biochemical composition and physical characteristics of soybean seed”. *Journal of Food Composition and Analysis* **19**.2-3 (2006), pp. 188–195.

[Kur13] Kuroda, Y. et al. “QTL affecting fitness of hybrids between wild and cultivated soybeans in experimental fields”. *Ecology and Evolution* **3**.7 (2013), pp. 2150–2168.

[Lan90] Lande, R & Thompson, R. “Efficiency of marker-assisted selection in the improvement of quantitative traits.” *Genetics* **124**.3 (1990), pp. 743–756. eprint: `https://www.genetics.org/content/124/3/743.full.pdf`.

[Lan87] Lander, E. S. et al. “MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations”. *Genomics* **1**.2 (1987), pp. 174–181.

[Lar94] Lark, K. G. et al. “EPISTATIC EXPRESSION OF QUANTITATIVE TRAIT LOCI (QTL) IN SOYBEAN GLYCINE-MAX (L) MERR DETERMINED BY QTL ASSOCIATION WITH RFLP ALLELES”. *Theoretical and Applied Genetics* **88**.3-4 (1994), pp. 486–489.

[Lee96] Lee, S. H. et al. “RFLP loci associated with soybean seed protein and oil content across populations and locations”. *Theoretical and Applied Genetics* **93**.5-6 (1996), pp. 649–657.

[Lee01] Lee, S. H. et al. “Genetic mapping of QTLs conditioning soybean sprout yield and quality”. *Theoretical and Applied Genetics* **103**.5 (2001), pp. 702–709.

[Lei16] Leite, D. C. et al. “QTL mapping of soybean oil content for marker-assisted selection in plant breeding program”. *Genet Mol Res* **15**.1 (2016).

[Len19] Lenth, R. *emmeans: Estimated marginal means, aka least-squares means. R package version 1.4. 3.01*. 2019.

[Li08] Li, D. D. et al. “Soybean QTL for yield and yield components associated with Glycine soja alleles”. *Crop Science* **48**.2 (2008), pp. 571–581.

[Li10] Li, D. M. et al. “Identification of QTL underlying soluble pigment content in soybean stems related to resistance to soybean white mold (Sclerotinia sclerotiorum)”. *Euphytica* **172**.1 (2010), pp. 49–57.

[Li11] Li, H. W. 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* **182**.1 (2011), pp. 117–132.

[Lia10a] Liang, H. Z. et al. “QTL Mapping of Isoflavone, Oil and Protein Contents in Soybean (Glycine max L. Merr.)” *Agricultural Sciences in China* **9**.8 (2010), pp. 1108–1116.

[Lia10b] Liang, Q. A. et al. “QTL analysis of root traits as related to phosphorus efficiency in soybean”. *Annals of Botany* **106**.1 (2010), pp. 223–234.

[Liu97] Liu, B. H. *Statistical genomics: linkage, mapping, and QTL analysis*. CRC press, 1997.

[Lu13] Lu, W. G. et al. “Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean”. *Theoretical and Applied Genetics* **126**.2 (2013), pp. 425–433.

[Man93a] Mansur, L. M. et al. “DETERMINING THE LINKAGE OF QUANTITATIVE TRAIT LOCI TO RFLP MARKERS USING EXTREME PHENOTYPES OF RECOMBINANT INBREDS OF SOYBEAN (GLYCINE-MAX L MERR)”. *Theoretical and Applied Genetics* **86**.8 (1993), pp. 914–918.

[Man93b] Mansur, L. M. et al. “INTERVAL MAPPING OF QUANTITATIVE TRAIT LOCI FOR REPRODUCTIVE, MORPHOLOGICAL, AND SEED TRAITS OF SOYBEAN (GLYCINE-MAX L)”. *Theoretical and Applied Genetics* **86**.8 (1993), pp. 907–913.

[Man96] Mansur, L. M. et al. “Genetic mapping of agronomic traits using recombinant inbred lines of soybean”. *Crop Science* **36**.5 (1996), pp. 1327–1336.

[Mao13] Mao, T. T. et al. “Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments”. *Plant Breeding* **132**.6 (2013), pp. 630–641.

[Mar07] Margarido, G. R. et al. “OneMap: software for genetic mapping in outcrossing species”. *Hereditas* **144**.3 (2007), pp. 78–79.

[Mar92] Martinez, O & Curnow, R. “Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers”. *Theoretical and Applied Genetics* **85**.4 (1992), pp. 480–488.

[Mec84] Meckel, L et al. “Effect of Moisture Stress on Seed Growth in Soybeans 1”. *Agronomy Journal* **76**.4 (1984), pp. 647–650.

[Med14] Medic, J. et al. “Current Knowledge in Soybean Composition”. *Journal of the American Oil Chemists Society* **91**.3 (2014), pp. 363–384.

[Men02] Men, A. E. et al. “Fast neutron mutagenesis of soybean (Glycine soja L.) produces a supernodulating mutant containing a large deletion in linkage group H”. *Genome letters* **1**.3 (2002), pp. 147–155.

[Men15] Meng, L. et al. “QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations”. *The Crop Journal* **3**.3 (2015), pp. 269–283.

[Mia17] Mian, M. et al. “Registration of ‘Highpro1’soybean with high protein and high yield developed from a North× South cross”. *Journal of Plant Registrations* **11**.1 (2017), pp. 51–54.

[Mol09] Mollinari, M et al. “Evaluation of algorithms used to order markers on genetic maps”. *Heredity* **103**.6 (2009), pp. 494–502.

[Nae08] Naeve, S. L. & Huerd, S. C. “Year, region, and temperature effects on the quality of Minnesota’s soybean crop”. *Agronomy Journal* **100**.3 (2008), pp. 690–695.

[Nar12] Narvel, J. et al. *Methods and compositions for selecting soybean plants resistant to southern root knot nematode*. US Patent 8,198,509. 2012.

[Ngu12] Nguyen, V. T. et al. “Mapping of Quantitative Trait Loci Associated with Resistance to Phytophthora sojae and Flooding Tolerance in Soybean”. *Crop Science* **52**.6 (2012), pp. 2481–2493.

[Nic06] Nichols, D. M. et al. “Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits”. *Crop Science* **46**.2 (2006), pp. 834–839.

[Nin18] Ning, H. L. et al. “Identification of QTLs related to the vertical distribution and seed-set of pod number in soybean Glycine max (L.) Merri”. *Plos One* **13**.4 (2018), p. 15.

[Orf99] Orf, J. H. et al. “Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations”. *Crop Science* **39**.6 (1999), pp. 1642–1651.

[Pal09a] Palomeque, L. et al. “QTL in mega-environments: I. Universal and specific seed yield QTL detected in a population derived from a cross of high-yielding adapted x high-yielding exotic soybean lines”. *Theoretical and Applied Genetics* **119**.3 (2009), pp. 417–427.

[Pal09b] Palomeque, L. et al. “QTL in mega-environments: II. Agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted x high-yielding exotic soybean lines”. *Theoretical and Applied Genetics* **119**.3 (2009), pp. 429–436.

[Pal10] Palomeque, L. et al. “Validation of mega-environment universal and specific QTL associated with seed yield and agronomic traits in soybeans”. *Theoretical and Applied Genetics* **120**.5 (2010), pp. 997–1003.

[Pan12] Pandurangan, S. et al. “Relationship between asparagine metabolism and protein concentration in soybean seed”. *Journal of Experimental Botany* **63**.8 (2012), pp. 3173–3184.

[Pan04] Panthee, D. R. et al. “Quantitative trait loci for beta-conglycinin (7S) and glycinin (11S) fractions of soybean storage protein”. *Journal of the American Oil Chemists Society* **81**.11 (2004), pp. 1005–1012.

[Pan05] Panthee, D. R. et al. “Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean”. *Crop Science* **45**.5 (2005), p. 2015.

[Pan06a] Panthee, D. R. et al. “Genomic regions associated with amino acid composition in soybean”. *Molecular Breeding* **17**.1 (2006), pp. 79–89.

[Pan06b] Panthee, D. R. et al. “Modifier QTL for fatty acid composition in soybean oil”. *Euphytica* **152**.1 (2006), pp. 67–73.

[Pan06c] Panthee, D. R. et al. “Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds”. *Theoretical and Applied Genetics* **112**.3 (2006), pp. 546–553.

[Pan07] Panthee, D. R. et al. “Quantitative trait loci for agronomic traits in soybean”. *Plant Breeding* **126**.1 (2007), pp. 51–57.

[Pan06d] Panthee, D. & Pantalone, V. “Registration of soybean germplasm lines TN03-350 and TN04-5321 with improved protein concentration and quality”. *Crop science* **46**.5 (2006), p. 2328.

[Pat13] Pathan, S. M. et al. “Genetic Mapping and Confirmation of Quantitative Trait Loci for Seed Protein and Oil Contents and Seed Weight in Soybean”. *Crop Science* **53**.3 (2013), p. 765.

[Pat17] Patil, G. et al. “Molecular mapping and genomics of soybean seed protein: a review and perspective for the future”. *Theoretical and Applied Genetics* **130**.10 (2017), pp. 1975–1991.

[Per08a] Peregrine, E. et al. “Evaluation of the USDA Soybean Germplasm Collection: Maturity Group V (FC 30265–PI 612614) and Maturity Groups VI–VIII (PI 416758–PI 606432B)”. *US Department of Agriculture Technical Bulletin* 1920 (2008), pp. 1–367.

[Per08b] Peregrine, E. et al. “Evaluation of the USDA soybean germplasm collection: maturity groups 000–IV (PI 578371-PI612761)”. *US Department of Agriculture Technical Bulletin* 1919 (2008), pp. 1–155.

[Pet13] Pettersson, D & Pontoppidan, K. “Soybean meal and the potential for upgrading its feeding value by enzyme supplementation”. *Soybean-bio-active compounds* (2013), pp. 288–307.

[Pri15] Prince, S. J. et al. “Genetic variants in root architecture-related genes in a Glycine soja accession, a potential resource to improve cultivated soybean”. *BMC genomics* **16**.1 (2015), p. 132.

[Qi14a] Qi, X. P. et al. “Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing”. *Nature Communications* **5** (2014), p. 11.

[Qi11a] Qi, Z. M. et al. “A meta-analysis of seed protein concentration QTL in soybean”. *Canadian Journal of Plant Science* **91**.1 (2011), pp. 221–230.

[Qi11b] Qi, Z. M. et al. “An Integrated Quantitative Trait Locus Map of Oil Content in Soybean, Glycine max (L.) Merr., Generated Using a Meta-Analysis Method for Mining Genes”. *Agricultural Sciences in China* **10**.11 (2011), pp. 1681–1692.

[Qi11c] Qi, Z. M. et al. “Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes”. *Euphytica* **179**.3 (2011), pp. 499–514.

[Qi14b] Qi, Z. M. 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* **133**.4 (2014), pp. 499–507.

[Qi14c] Qi, Z. M. et al. “QTL analysis of soybean oil content under 17 environments”. *Canadian Journal of Plant Science* **94**.2 (2014), pp. 245–261.

[Qi18] Qi, Z. M. et al. “Meta-analysis and transcriptome profiling reveal hub genes for soybean seed storage composition during seed development”. *Plant Cell and Environment* **41**.9 (2018), pp. 2109–2127.

[Qi11d] Qi, Z.-m. et al. “Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes”. *Euphytica* **179**.3 (2011), pp. 499–514.

[Qiu99] Qiu, B. X. et al. “RFLP markers associated with soybean cyst nematode resistance and seed composition in a ’Peking’ x ’Essex’ population”. *Theoretical and Applied Genetics* **98**.3-4 (1999), pp. 356–364.

[R C13] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2013.

[Ram14] Ramamurthy, R. K. et al. “Identification of new QTLs for seed mineral, cysteine, and methionine concentrations in soybean Glycine max (L.) Merr.” *Molecular Breeding* **34**.2 (2014), pp. 431–445.

[Rei06a] Reinprecht, Y. et al. “Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (Glycine max (L.) Merrill) germplasm”. *Genome* **49**.12 (2006), pp. 1510–1527.

[Rei06b] Reinprecht, Y. et al. “Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (Glycine max (L.) Merrill) germplasm”. *Genome* **49**.12 (2006), pp. 1510–1527.

[Ren09] Ren, C. et al. “Composition, vigor, and proteome of mature soybean seeds developed under high temperature”. *Crop science* **49**.3 (2009), pp. 1010–1022.

[Rey01] Reyna, N. & Sneller, C. H. “Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean”. *Crop Science* **41**.4 (2001), pp. 1317–1321.

[Rod10] Rodrigues, J. I. D. et al. “Mapping QTL for protein and oil content in soybean”. *Pesquisa Agropecuaria Brasileira* **45**.5 (2010), pp. 472–480.

[Ros88] Rose, I. “Effects of moisture stress on the oil and protein components of soybean seeds”. *Australian Journal of Agricultural Research* **39**.2 (1988), pp. 163–170.

[Ros13a] Rossi, M. E. et al. “Genetic basis of soybean adaptation to North American vs. Asian mega-environments in two independent populations from Canadian x Chinese crosses”. *Theoretical and Applied Genetics* **126**.7 (2013), pp. 1809–1823.

[Ros13b] Rossi, M. E. et al. “Genetic basis of soybean adaptation to North American vs. Asian mega-environments in two independent populations from Canadian

x Chinese crosses". *Theoretical and Applied Genetics* **126**.7 (2013), pp. 1809–1823.

[Rot09] Rotundo, J. L. & Westgate, M. E. "Meta-analysis of environmental effects on soybean seed composition". *Field Crops Research* **110**.2 (2009), pp. 147–156.

[Rot10] Rotundo, J. L. & Westgate, M. E. "Rate and duration of seed component accumulation in water-stressed soybean". *Crop science* **50**.2 (2010), pp. 676–684.

[Sal06] Salas, P. et al. "Genetic mapping of seed shape in three populations of recombinant inbred lines of soybean (Glycine max L. Merr.)" *Theoretical and Applied Genetics* **113**.8 (2006), pp. 1459–1466.

[Sch10] Schmutz, J. et al. "Genome sequence of the palaeopolyploid soybean". *nature* **463**.7278 (2010), pp. 178–183.

[Seb09] Sebastian, S. et al. *Genetic loci associated with iron deficiency tolerance in soybean*. US Patent 7,582,806. 2009.

[Seb00] Sebolt, A. M. et al. "Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean". *Crop Science* **40**.5 (2000), pp. 1438–1444.

[Shi08] Shibata, M. et al. "Genetic relationship between lipid content and linolenic acid concentration in soybean seeds". *Breeding Science* **58**.4 (2008), pp. 361–366.

[Shi03] Shipe, E. et al. "Registration of 'Santee'soybean". *Crop science* **43**.6 (2003), pp. 2305–2306.

[Sil17] Silva, F. L. da et al. *Soybean Breeding*. Springer, 2017.

[Sin15] Singh, B. & Singh, A. K. *Marker-assisted plant breeding: principles and practices*. Springer, 2015.

[Son14] Song, Q et al. "Soybean BARCSoySNP6K Beadchip: A tool for soybean genetics research". *Poster presented at: plant and animal genome XXII conference, San Diego, CA, USA*. 2014, pp. 10–15.

[Son13] Song, Q. J. et al. "Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean". *Plos One* **8**.1 (2013).

[Sos12] Sosnowski, O. et al. "BioMercator V3: an upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms". *Bioinformatics* **28**.15 (2012), pp. 2082–2083.

[Spe01] Specht, J. E. et al. "Soybean response to water: A QTL analysis of drought tolerance". *Crop Science* **41**.2 (2001), pp. 493–509.

[Spe04] Spencer, M. M. et al. “Molecular markers associated with linolenic acid content in soybean”. *Journal of the American Oil Chemists Society* **81**.6 (2004), pp. 559–562.

[Sta08] Stacey, G. et al. “Genetics and genomics of soybean” (2008).

[Sto04] Stombaugh, S. K. et al. “Quantitative trait loci associated with cell wall polysaccharides in soybean seed”. *Crop Science* **44**.6 (2004), pp. 2101–2106.

[Stu13] Sturtevant, A. H. “The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association”. *Journal of Experimental Zoology* **14**.1 (1913), pp. 43–59. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/jez.1400140104`.

[Sun06] Sun, D. H. et al. “Quantitative trait loci analysis for the developmental behavior of Soybean (Glycine max L. Merr.)” *Theoretical and Applied Genetics* **112**.4 (2006), pp. 665–673.

[Taj03] Tajuddin, T. et al. “Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines”. *Breeding Science* **53**.2 (2003), pp. 133–140.

[Tay17] Taylor, J. & Butler, D. “R package ASMap: efficient genetic linkage map construction and diagnosis”. *arXiv preprint arXiv:1705.06916* (2017).

[Tis03] Tischner, T. et al. “Genetics of seed abortion and reproductive traits in soybean”. *Crop Science* **43**.2 (2003), pp. 464–473.

[Van17] Van, K. & McHale, L. K. “Meta-Analyses of QTLs Associated with Protein and Oil Contents and Compositions in Soybean Glycine max (L.) Merr. Seed”. *International Journal of Molecular Sciences* **18**.6 (2017).

[VO06] Van Ooijen, J. “JoinMap 4”. *Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands* **33** (2006).

[VO09] Van Ooijen, J. & Kyazma, B. “MapQTL 6”. *Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV: Wageningen, Netherlands* (2009).

[Vol02] Vollmann, J. et al. “The presence or absence of the soybean Kunitz trypsin inhibitor as a quantitative trait locus for seed protein content”. *Plant Breeding* **121**.3 (2002), pp. 272–274.

[Wan04] Wang, D. et al. “Identification of putative QTL that underlie yield in interspecific soybean backcross populations”. *Theoretical and Applied Genetics* **108**.3 (2004), pp. 458–467.

[Wan15a] Wang, J. et al. “Identification and mapping of stable QTL for protein content in soybean seeds”. *Molecular Breeding* **35**.3 (2015), p. 10.

[Wan12] Wang, X. Z. et al. “Quantitative trait locus analysis of saturated fatty acids in a population of recombinant inbred lines of soybean”. *Molecular Breeding* **30**.2 (2012), pp. 1163–1179.

[Wan14] Wang, X. Z. et al. “Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean”. *Molecular Genetics and Genomics* **289**.5 (2014), pp. 935–949.

[Wan15b] Wang, X. Z. et al. “Quantitative trait locus analysis of seed sulfur-containing amino acids in two recombinant inbred line populations of soybean”. *Euphytica* **201**.2 (2015), pp. 293–305.

[War15] Warrington, C. V. et al. “QTL for seed protein and amino acids in the Benning x Danbaekkong soybean population”. *Theoretical and Applied Genetics* **128**.5 (2015), pp. 839–850.

[Wol82] Wolf, R. et al. “Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty acids, amino acids and sugars”. *Journal of the American Oil Chemists’ Society* **59**.5 (1982), pp. 230–232.

[Wu03] Wu, J. et al. “Monte Carlo simulations on marker grouping and ordering”. *Theoretical and Applied Genetics* **107**.3 (2003), pp. 568–573.

[Wu08] Wu, Y. et al. “Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph”. *PLoS genetics* **4**.10 (2008).

[Xie12] Xie, D. W. et al. “SSR- and SNP-related QTL underlying linolenic acid and other fatty acid contents in soybean seeds across multiple environments”. *Molecular Breeding* **30**.1 (2012), pp. 169–179.

[You99] Young, N. D. “A cautiously optimistic vision for marker-assisted breeding”. *Molecular breeding* **5**.6 (1999), pp. 505–510.

[Yu11] Yu, J.-K. et al. *Markers associated with soybean rust resistance and methods of use therefor*. US Patent App. 13/054,760. 2011.

[Yua02] Yuan, J. et al. “Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance”. *Crop Science* **42**.1 (2002), pp. 271–277.

[Zha04] Zhang, W. K. et al. “QTL mapping of ten agronomic traits on the soybean (Glycine max L. Merr.) genetic map and their association with EST markers”. *Theoretical and Applied Genetics* **108**.6 (2004), pp. 1131–1139.

[Zha17] Zhang, X. et al. “Meta-analysis of major QTL for abiotic stress tolerance in barley and implications for barley breeding”. *Planta* **245**.2 (2017), pp. 283–295.

[Zha15] Zhang, Y. H. et al. “Marker-assisted breeding for transgressive seed protein content in soybean Glycine max (L.) Merr.” *Theoretical and Applied Genetics* **128**.6 (2015), pp. 1061–1072.

APPENDIX

A

APPENDIX

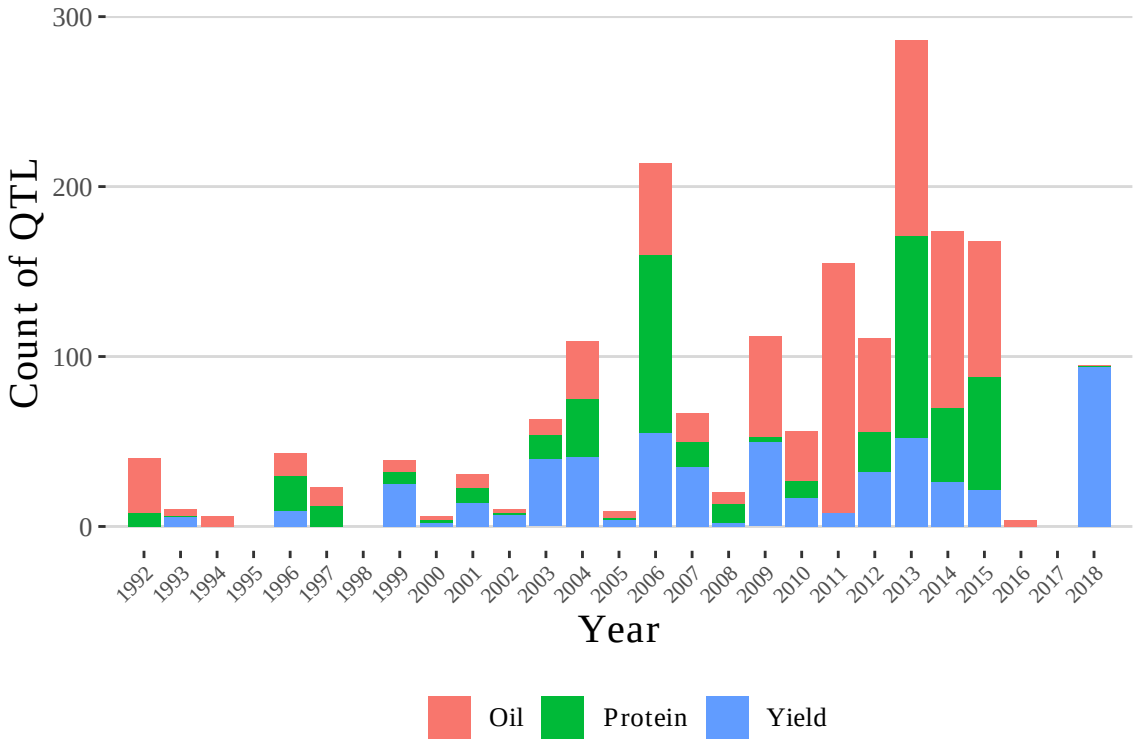

Figure A.1 Counts of Soybean QTL Reported by Year for Seed Oil, Protein, and Yield Traits. Data obtained from SoyBase on 2020-04-21.

Table A.1 Counts of Soybean QTL Discovered by Category. Data obtained from SoyBase on 2020-04-21.

QTL Category	QTL Found
Other Seed	865
Whole Plant	543
Oil	482
Protein	367
Yield	339
Reproductive Period	325
Fungal	314
Leaf-stem	312
Inorganic	214
Nematode	214
Root	133
Pod	101
Insect	77
Misc	57
Viral	21

Table A.2 Counts of published Soybean QTL in the top five subcategories for Protein, Oil, and Yield Traits

Protein QTL		**Oil QTL**		**Yield QTL**	
Category	**Count**	**Category**	**Count**	**Category**	**Count**
Seed protein	209	Seed oil	258	Seed yield	153
Seed Met	18	Seed linolenic	52	Seed set	66
cqSeed protein	15	Seed oleic	35	Seed weight per plant	19
Seed Cys	15	Seed linoleic	33	Row spacing response	16
Seed Thr	9	Seed palmitic	33	Seed yield to Plant height ratio	16

Table A.3 Literature Identifying Confirmed Soybean QTL for Seed Oil, Protein, and Yield traits

Literature Source	QTL Category	QTL Name	LG(GmComposite2003)	Start(cM)	End(cM)	Associated Loci	R-Squared
Fasoula et al. 2004	Oil	cqSeed oil-001	C1	89.72000	91.72	A063_1	
Fasoula et al. 2004	Oil	cqSeed oil-002	H	85.80000	87.80	A566_2	0.083
Fasoula et al. 2004	Oil	cqSeed oil-003	L	30.58000	34.54	Satt313	
Fasoula et al. 2004	Oil	cqSeed oil-003	L	30.58000	34.54	Satt398	
Fasoula et al. 2004	Protein	cqSeed protein-001	E	29.89000	31.89	A454_1	0.123
Fasoula et al. 2004	Protein	cqSeed protein-002				A132_4	0.056
Fasoula et al. 2004	Yield	cqSeed weight-001	G	52.42000	54.42	Satt303	0.08
Fasoula et al. 2004	Yield	cqSeed weight-002	G	44.40000	46.40	Satt263	0.18
Nichols et al. 2006	Oil	cqSeed oil-004	I	35.40000	37.40	Satt496	
Nichols et al. 2006	Protein	cqSeed protein-003	I	35.40000	37.40	Satt496	
Nichols et al. 2006	Yield	cqSeed weight-003	I	35.40000	37.40	Satt496	
Nichols et al. 2006	Yield	cqSeed yield-001	I	35.40001	37.40	Satt496	
Pathan et al. 2013a	Oil	cqSeed oil-005	N	35.77000	39.34	BARC-055637-13558	
Pathan et al. 2013a	Oil	cqSeed oil-005	N	35.77000	39.34	BARC-057823-14942	0.056
Pathan et al. 2013a	Oil	cqSeed oil-006	C2	34.63000	36.55	BARC-059985-16274	
Pathan et al. 2013a	Oil	cqSeed oil-006	C2	34.63000	36.55	BARC-044639-08743	0.093
Pathan et al. 2013a	Oil	cqSeed oil-007	E	16.13000	16.78	BARC-039687-07541	
Pathan et al. 2013a	Oil	cqSeed oil-007	E	16.13000	16.78	BARC-042349-08247	0.053
Pathan et al. 2013a	Oil	cqSeed oil-008	A1	96.72000	102.05	BARC-041257-07953	
Pathan et al. 2013a	Oil	cqSeed oil-008	A1	96.72000	102.05	BARC-060051-16321	0.071

Table A.3 Literature Identifying Confirmed QTL for Seed Oil, Protein, and Yield traits *(continued)*

Literature Source	QTL Category	QTL Name	LG(GmComposite2003)	Start(cM)	End(cM)	Associated Loci	R-Squared
Pathan et al. 2013a	Oil	cqSeed oil-009	C2	34.63000	36.55	BARC-059985-16274	
Pathan et al. 2013a	Oil	cqSeed oil-009	C2	34.63000	36.55	BARC-044639-08743	0.093
Pathan et al. 2013a	Oil	cqSeed oil-010	E	16.13000	16.78	BARC-039687-07541	
Pathan et al. 2013a	Oil	cqSeed oil-010	E	16.13000	16.78	BARC-042349-08247	0.097
Pathan et al. 2013a	Oil	cqSeed oil-011	D1b	117.48000	120.78	BARC-044747-08795	
Pathan et al. 2013a	Oil	cqSeed oil-011	D1b	117.48000	120.78	BARC-054149-12354	0.105
Pathan et al. 2013a	Oil	cqSeed oil-012	A1	96.73000	102.05	BARC-041257-07953	
Pathan et al. 2013a	Oil	cqSeed oil-012	A1	96.73000	102.05	BARC-060051-16321	0.084
Pathan et al. 2013a	Oil	cqSeed oil-013	E	38.78000	41.82	BARC-018901-03270	
Pathan et al. 2013a	Oil	cqSeed oil-013	E	38.78000	41.82	BARC-038977-07417	0.092
Pathan et al. 2013a	Oil	cqSeed oil-014	D1b	114.49000	118.94	BARC-028373-05856	
Pathan et al. 2013a	Oil	cqSeed oil-014	D1b	114.49000	118.94	Sat_198	0.061
Pathan et al. 2013a	Oil	cqSeed oil-015	A1	96.73000	102.05	BARC-041257-07953	
Pathan et al. 2013a	Oil	cqSeed oil-015	A1	96.73000	102.05	BARC-060051-16321	0.057
Pathan et al. 2013a	Oil	cqSeed oil-016	C2	34.34000	51.84	BARC-059997-16280	
Pathan et al. 2013a	Oil	cqSeed oil-016	C2	34.34000	51.84	Sat_336	0.123
Pathan et al. 2013a	Protein	cqSeed protein-004	N	31.85000	35.30	BARC-016199-02307	
Pathan et al. 2013a	Protein	cqSeed protein-004	N	31.85000	35.30	BARC-053313-11792	0.056
Pathan et al. 2013a	Protein	cqSeed protein-005	C2	34.63000	36.55	BARC-059985-16274	
Pathan et al. 2013a	Protein	cqSeed protein-005	C2	34.63000	36.55	BARC-044639-08743	0.053
Pathan et al. 2013a	Protein	cqSeed protein-006	M	18.87000	46.65	BARC-014705-01623	

Table A.3 Literature Identifying Confirmed QTL for Seed Oil, Protein, and Yield traits *(continued)*

Literature Source	QTL Category	QTL Name	LG(GmComposite2003)	Start(cM)	End(cM)	Associated Loci	R-Squared
Pathan et al. 2013a	Protein	cqSeed protein-006	M	18.87000	46.65	BARC-029825-06442	0.068
Pathan et al. 2013a	Protein	cqSeed protein-007	C2	34.34000	40.29	Satt281	
Pathan et al. 2013a	Protein	cqSeed protein-007	C2	34.34000	40.29	BARC-059997-16280	0.048
Pathan et al. 2013a	Protein	cqSeed protein-008	E	9.99000	10.15	BARC-018923-03037	
Pathan et al. 2013a	Protein	cqSeed protein-008	E	9.99000	10.15	BARC-025493-06513	0.075
Pathan et al. 2013a	Protein	cqSeed protein-009	M	43.74000	46.65	BARC-014705-01623	
Pathan et al. 2013a	Protein	cqSeed protein-009	M	43.74000	46.65	BARC-029825-06442	0.07
Pathan et al. 2013a	Protein	cqSeed protein-010	N	15.85000	22.65	BARC-028645-05979	
Pathan et al. 2013a	Protein	cqSeed protein-010	N	15.85000	22.65	BARC-055149-13089	0.067
Pathan et al. 2013a	Protein	cqSeed protein-011	A1	93.23000	96.73	Satt236	
Pathan et al. 2013a	Protein	cqSeed protein-011	A1	93.23000	96.73	BARC-041257-07953	0.064
Pathan et al. 2013a	Protein	cqSeed protein-012	C2	113.50000	128.22	Satt433	
Pathan et al. 2013a	Protein	cqSeed protein-012	C2	113.50000	128.22	BARC-049601-09082	0.076
Pathan et al. 2013a	Protein	cqSeed protein-013	A2	51.50000	53.91	Satt632	
Pathan et al. 2013a	Protein	cqSeed protein-013	A2	51.50000	53.91	BARC-040339-07715	0.142
Pathan et al. 2013a	Protein	cqSeed protein-014	A1	96.73000	102.05	BARC-041257-07953	
Pathan et al. 2013a	Protein	cqSeed protein-014	A1	96.73000	102.05	BARC-060051-16321	0.077
Pathan et al. 2013a	Protein	cqSeed protein-015	C2	34.30000	45.75	BARC-059985-16274	
Pathan et al. 2013a	Protein	cqSeed protein-015	C2	34.30000	45.75	Satt291	0.09
Pathan et al. 2013a	Protein	cqSeed protein-016	A2	51.50000	53.99	Satt632	
Pathan et al. 2013a	Protein	cqSeed protein-016	A2	51.50000	53.99	BARC-010097-00518	0.068

Table A.3 Literature Identifying Confirmed QTL for Seed Oil, Protein, and Yield traits *(continued)*

Literature Source	QTL Category	QTL Name	LG(GmComposite2003)	Start(cM)	End(cM)	Associated Loci	R-Squared
Pathan et al. 2013a	Yield	cqSeed weight-004	C2	42.49000	42.50	BARC-056271-14211	
Pathan et al. 2013a	Yield	cqSeed weight-004	C2	42.49000	42.50	BARC-057469-14770	0.064
Pathan et al. 2013a	Yield	cqSeed weight-005	L	68.36000	70.36	Satt527	
Pathan et al. 2013a	Yield	cqSeed weight-005	L	68.36000	70.36	BARC-047496-12943	0.07
Pathan et al. 2013a	Yield	cqSeed weight-006	D1a	17.52000	25.28	BARC-045297-08928	
Pathan et al. 2013a	Yield	cqSeed weight-006	D1a	17.52000	25.28	Satt184	0.062
Pathan et al. 2013a	Yield	cqSeed weight-007	L	68.36000	70.36	Satt527	
Pathan et al. 2013a	Yield	cqSeed weight-007	L	68.36000	70.36	BARC-047496-12943	0.132
Pathan et al. 2013a	Yield	cqSeed weight-008	C2	34.34000	51.84	BARC-059997-16280	
Pathan et al. 2013a	Yield	cqSeed weight-008	C2	34.34000	51.84	Sat_336	0.059
Pathan et al. 2013a	Yield	cqSeed weight-009	L	68.35000	70.36	BARC-014885-01914	
Pathan et al. 2013a	Yield	cqSeed weight-009	L	68.35000	70.36	Satt527	0.098
Pathan et al. 2013a	Yield	cqSeed weight-010	D1a	17.52000	24.13	BARC-038883-07384	
Pathan et al. 2013a	Yield	cqSeed weight-010	D1a	17.52000	24.13	Satt184	0.079
Pathan et al. 2013a	Yield	cqSeed weight-011	C2	34.34000	51.84	BARC-059997-16280	
Pathan et al. 2013a	Yield	cqSeed weight-011	C2	34.34000	51.84	Sat_336	0.059
Pathan et al. 2013a	Yield	cqSeed weight-012	L	68.35000	70.36	BARC-014885-01914	
Pathan et al. 2013a	Yield	cqSeed weight-012	L	68.35000	70.36	Satt527	0.073
Pathan et al. 2013a	Yield	cqSeed weight-013	C2	27.85000	40.29	BARC-016957-02165	
Pathan et al. 2013a	Yield	cqSeed weight-013	C2	27.85000	40.29	Satt281	0.142

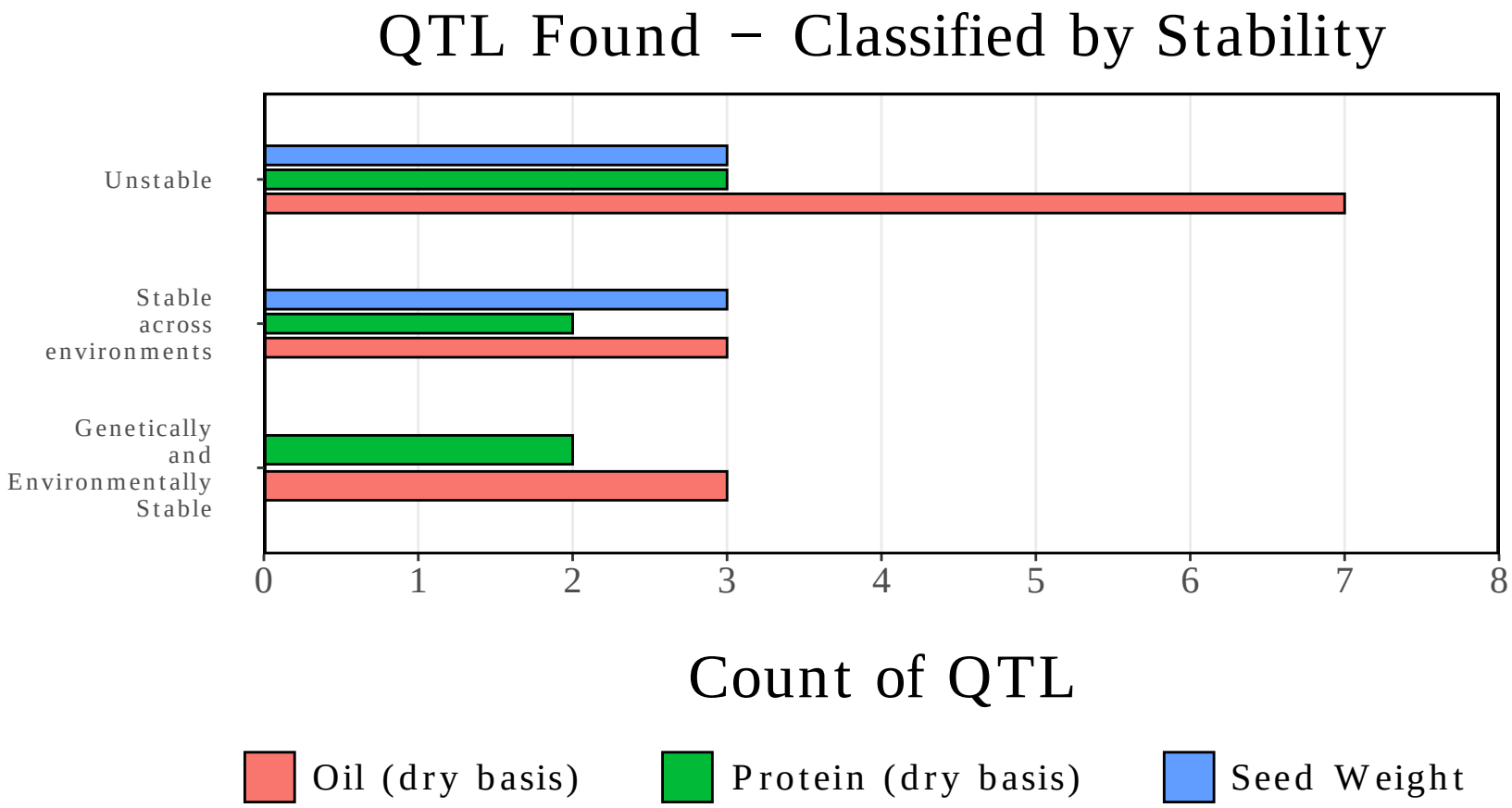

Figure A.2 Counts of QTL found through QTL mapping of soybean mapping populations 201 and 202. QTL counts are colored by trait, and bars are classified according to if they were found in multiple environments and populations, (Genetically and Environmentally stable), if they were found in multiple environments but a single mapping population, (Stable across environments), or if they were found in a single population and environment, (Unstable).

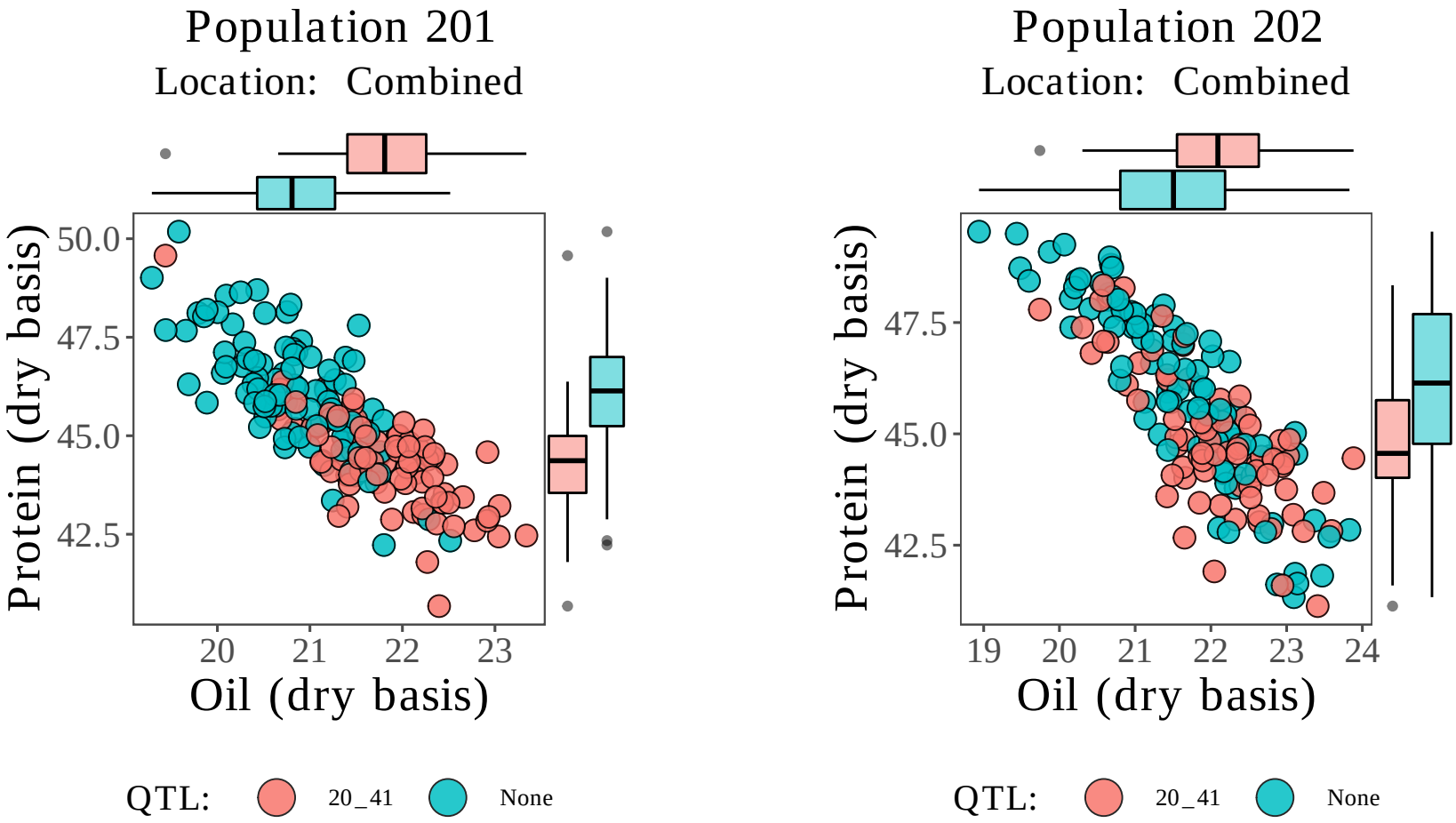

Figure A.3 The seed protein and oil content of RILs of soybean mapping populations 201 and 202 in the combined environment. Points are colored based on if an individual possessed the alleles associated with an increase in oil content at the site of the QTL found on chromosome 20. Lines possessing the increased oil alleles are colored in red while lines without the allele are colored in blue.

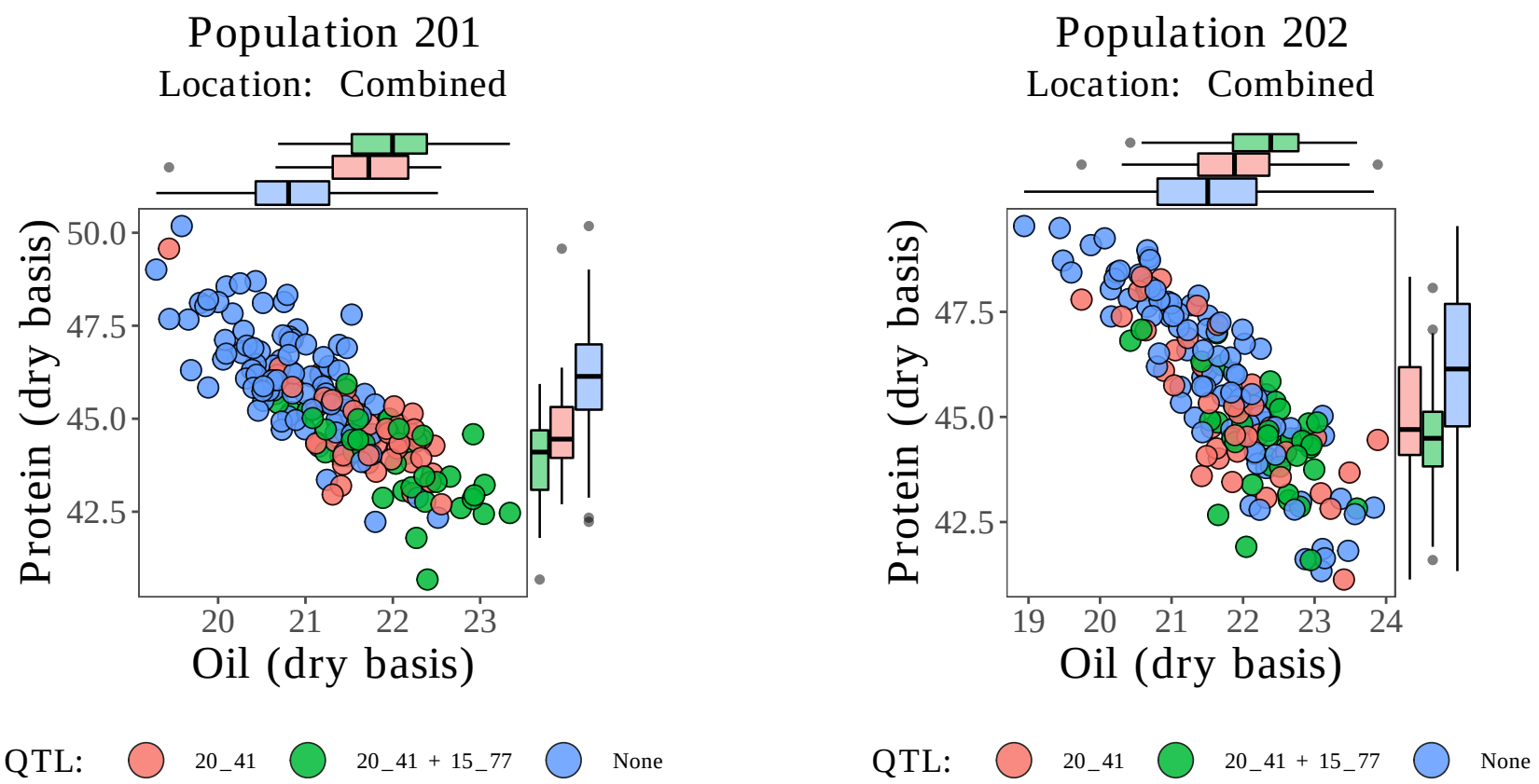

Figure A.4 The seed protein and oil content of RILs of soybean mapping populations 201 and 202 in the combined environment. Points are colored based on if an individual possessed the alleles associated with an increase in oil content at the site of the QTL found on chromosome 20, and the site of the QTL on chromosome 15. Lines possessing the increased oil allele at only chromosome 20 are colored in red, lines possessing the oil alleles at both the QTL on chromosomes 15 and 20 are colored in green, while lines without the alleles are colored in blue.

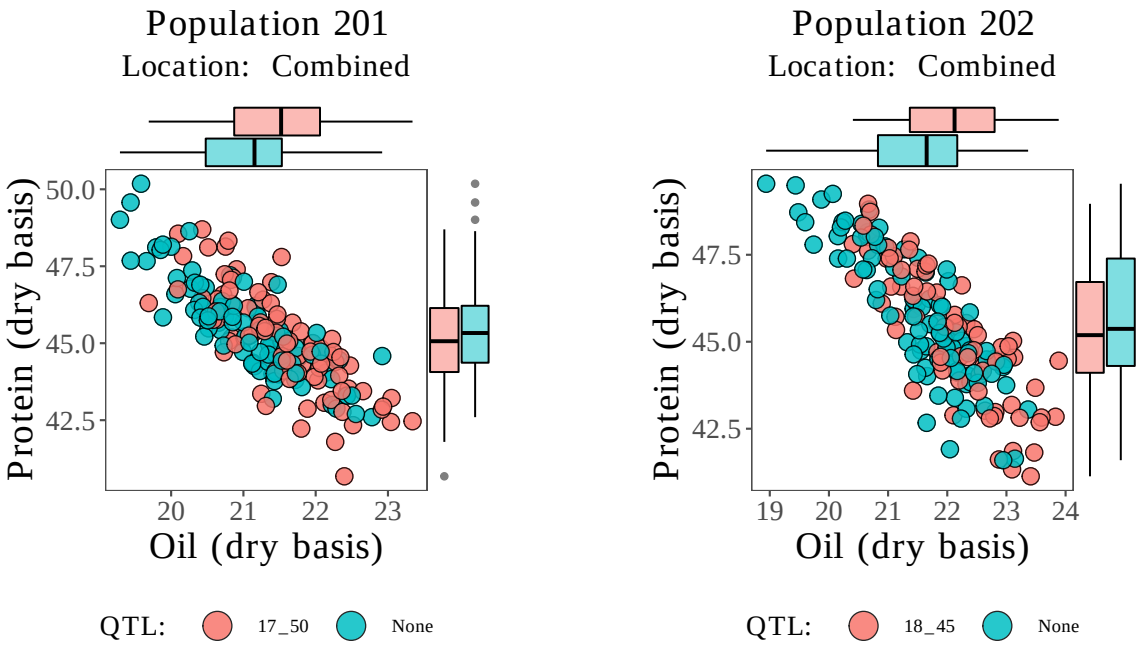

Figure A.5 The seed protein and oil content of RILs of soybean mapping populations 201 and 202 in the combined environment. Points are colored based on if an individual possessed the alleles associated with an increase in oil content at the site of the QTL found on chromosome 17 in population 201, and the site of the QTL on chromosome 18 in population 202. These QTL were associated with an increase in oil and were not significantly associated with a decrease in protein. Points are colored in red if they possessed the alleles associated with an increase in oil, and lines are colored in blue if they did not possess the allele associated with an increase in oil.

Table A.4 ANOVA Table: Yield Test 1 - Height

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 1	Height	Loc	1	3740.8333	3740.8333	422.9787	0
		Genotype	44	1633.3185	37.1209	4.1973	0
		Loc:Genotype	44	649.3333	14.7576	1.6686	0.0108
		Loc:Rep	4	137.4519	34.3630	3.8854	0.0047
		Residuals	176	1556.5481	8.8440		

Table A.5 ANOVA Table: Yield Test 2 - Height

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 2	Height	Loc	1	4200.8333	4200.8333	436.7329	0
		Genotype	44	2252.2074	51.1865	5.3215	0
		Loc:Genotype	44	419.3333	9.5303	0.9908	0.4964
		Loc:Rep	4	309.0963	77.2741	8.0337	0
		Residuals	176	1692.9037	9.6188		

Table A.6 ANOVA Table: Yield Test 3 - Height

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 3	Height	Loc	1	1555.2000	1555.2000	137.4839	0
		Genotype	44	2062.8667	46.8833	4.1446	0
		Loc:Genotype	44	1297.8000	29.4955	2.6075	0
		Loc:Rep	4	316.4444	79.1111	6.9936	0
		Residuals	176	1990.8889	11.3119		

Table A.7 ANOVA Table: Yield Test 4 - Height

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 4	Height	Loc	1	1512.3000	1512.3000	82.7295	0
		Genotype	44	1539.5333	34.9894	1.9141	0.0017
		Loc:Genotype	44	466.8667	10.6106	0.5804	0.9823
		Loc:Rep	4	149.3778	37.3444	2.0429	0.0904
		Residuals	176	3217.2889	18.2801		

Table A.8 ANOVA Table: Yield Test 1 - Hundred Seed Weight

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 1	Hundred Seed Weight	Loc	1	0.4805	0.4805	0.79	0.3765
		Genotype	44	263.4548	5.9876	9.8438	0
		Loc:Genotype	44	24.3070	0.5524	0.9082	0.6316
		Loc:Rep	2	1.2179	0.6089	1.0011	0.3716
		Residuals	88	53.5271	0.6083		

Table A.9 ANOVA Table: Yield Test 2 - Hundred Seed Weight

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 2	Hundred Seed Weight	Loc	1	14.2805	14.2805	22.2117	0
		Genotype	44	270.3850	6.1451	9.558	0
		Loc:Genotype	44	29.7770	0.6767	1.0526	0.4109
		Loc:Rep	2	0.6974	0.3487	0.5424	0.5833
		Residuals	88	56.5776	0.6429		

Table A.10 ANOVA Table: Yield Test 3 - Hundred Seed Weight

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 3	Hundred Seed Weight	Loc	1	75.4930	75.4930	94.8669	0
		Genotype	44	396.2878	9.0065	11.3179	0
		Loc:Genotype	44	46.3393	1.0532	1.3234	0.1328
		Loc:Rep	2	4.0534	2.0267	2.5468	0.0841
		Residuals	88	70.0285	0.7958		

Table A.11 ANOVA Table: Yield Test 4 - Hundred Seed Weight

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 4	Hundred Seed Weight	Loc	1	66.8642	66.8642	60.212	0
		Genotype	44	631.5684	14.3538	12.9258	0
		Loc:Genotype	44	83.0413	1.8873	1.6995	0.0178
		Loc:Rep	2	73.0002	36.5001	32.8688	0
		Residuals	88	97.7221	1.1105		

Table A.12 ANOVA Table: Yield Test 1 - Oil (dry basis)

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 1	Oil (dry basis)	Loc	1	42.4751	42.4751	228.7621	0
		Genotype	44	153.7105	3.4934	18.8149	0
		Loc:Genotype	44	12.0945	0.2749	1.4804	0.0399
		Loc:Rep	4	0.8503	0.2126	1.1448	0.3371
		Residuals	176	32.6785	0.1857		

Table A.13 ANOVA Table: Yield Test 2 - Oil (dry basis)

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 2	Oil (dry basis)	Loc	1	0.0006	0.0006	0.0028	0.9579
		Genotype	44	212.9260	4.8392	21.7306	0
		Loc:Genotype	44	17.0497	0.3875	1.74	0.0064
		Loc:Rep	4	8.4784	2.1196	9.5181	0
		Residuals	176	39.1937	0.2227		

Table A.14 ANOVA Table: Yield Test 3 - Oil (dry basis)

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 3	Oil (dry basis)	Loc	1	4.1490	4.1490	13.6811	3e-04
		Genotype	44	364.9236	8.2937	27.3478	0
		Loc:Genotype	44	18.8400	0.4282	1.4119	0.0618
		Loc:Rep	4	1.0043	0.2511	0.8279	0.509
		Residuals	176	53.3751	0.3033		

Table A.15 ANOVA Table: Yield Test 4 - Oil (dry basis)

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 4	Oil (dry basis)	Loc	1	7.5033	7.5033	28.1969	0
		Genotype	44	340.7009	7.7432	29.0983	0
		Loc:Genotype	44	36.9103	0.8389	3.1524	0
		Loc:Rep	4	23.3466	5.8366	21.9336	0
		Residuals	176	46.8344	0.2661		

Table A.16 ANOVA Table: Yield Test 1 - Protein (dry basis)

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 1	Protein (dry basis)	Loc	1	115.3918	115.3918	166.8249	0
		Genotype	44	1051.8348	23.9053	34.5606	0
		Loc:Genotype	44	40.6994	0.9250	1.3373	0.0971
		Loc:Rep	4	4.0576	1.0144	1.4666	0.2143
		Residuals	176	121.7382	0.6917		

Table A.17 ANOVA Table: Yield Test 2 - Protein (dry basis)

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 2	Protein (dry basis)	Loc	1	4.7071	4.7071	5.0219	0.0263
		Genotype	44	889.5328	20.2167	21.5684	0
		Loc:Genotype	44	70.8352	1.6099	1.7175	0.0076
		Loc:Rep	4	32.4645	8.1161	8.6588	0
		Residuals	176	164.9697	0.9373		

Table A.18 ANOVA Table: Yield Test 3 - Protein (dry basis)

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 3	Protein (dry basis)	Loc	1	13.6373	13.6373	15.8571	1e-04
		Genotype	44	1443.0032	32.7955	38.1338	0
		Loc:Genotype	44	56.2799	1.2791	1.4873	0.0381
		Loc:Rep	4	10.0444	2.5111	2.9198	0.0227
		Residuals	176	151.3621	0.8600		

Table A.19 ANOVA Table: Yield Test 4 - Protein (dry basis)

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 4	Protein (dry basis)	Loc	1	1.5793	1.5793	2.1254	0.1467
		Genotype	44	1701.5655	38.6719	52.0438	0
		Loc:Genotype	44	75.1983	1.7091	2.3	1e-04
		Loc:Rep	4	52.8008	13.2002	17.7645	0
		Residuals	176	130.7794	0.7431		

Table A.20 ANOVA Table: Yield Test 1 - Protein + Oil

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 1	Protein + Oil	Loc	1	17.8487	17.8487	62.8938	0
		Genotype	44	525.3668	11.9402	42.0738	0
		Loc:Genotype	44	19.1487	0.4352	1.5335	0.028
		Loc:Rep	4	7.3581	1.8395	6.482	1e-04
		Residuals	176	49.9471	0.2838		

Table A.21 ANOVA Table: Yield Test 2 - Protein + Oil

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 2	Protein + Oil	Loc	1	4.8160	4.8160	12.0054	7e-04
		Genotype	44	364.0695	8.2743	20.6262	0
		Loc:Genotype	44	27.9397	0.6350	1.5829	0.0199
		Loc:Rep	4	9.2027	2.3007	5.7351	2e-04
		Residuals	176	70.6033	0.4012		

Table A.22 ANOVA Table: Yield Test 3 - Protein + Oil

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 3	Protein + Oil	Loc	1	2.7422	2.7422	8.8248	0.0034
		Genotype	44	504.7162	11.4708	36.9154	0
		Loc:Genotype	44	22.5830	0.5132	1.6517	0.0122
		Loc:Rep	4	11.7058	2.9264	9.4179	0
		Residuals	176	54.6890	0.3107		

Table A.23 ANOVA Table: Yield Test 4 - Protein + Oil

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 4	Protein + Oil	Loc	1	2.1978	2.1978	6.6498	0.0107
		Genotype	44	704.4646	16.0106	48.4422	0
		Loc:Genotype	44	24.4848	0.5565	1.6837	0.0097
		Loc:Rep	4	14.3416	3.5854	10.8481	0
		Residuals	176	58.1694	0.3305		

Table A.24 ANOVA Table: Yield Test 1 - Yield

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 1	Yield	Loc	1	14242043.4	14242043.403	1944.013	0
		Genotype	44	832763.7	18926.448	2.5834	0
		Loc:Genotype	44	578864.5	13156.012	1.7958	0.0042
		Loc:Rep	4	749567.8	187391.961	25.5787	0
		Residuals	176	1289394.5	7326.105		

Table A.25 ANOVA Table: Yield Test 2 - Yield

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 2	Yield	Loc	1	11893186.0	11893186.009	1199.0695	0
		Genotype	44	1548925.1	35202.844	3.5491	0
		Loc:Genotype	44	704694.7	16015.788	1.6147	0.0159
		Loc:Rep	4	255834.5	63958.623	6.4483	1e-04
		Residuals	176	1745687.5	9918.679		

Table A.26 ANOVA Table: Yield Test 3 - Yield

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 3	Yield	Loc	1	15283472.5	15283472.47	717.551	0
		Genotype	44	2428801.0	55200.02	2.5916	0
		Loc:Genotype	44	930738.2	21153.14	0.9931	0.4924
		Loc:Rep	4	409023.7	102255.93	4.8009	0.0011
		Residuals	176	3748710.6	21299.49		

Table A.27 ANOVA Table: Yield Test 4 - Yield

Test	Trait	Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yield Test 4	Yield	Loc	1	12029849.5	12029849.45	770.8993	0
		Genotype	44	1679461.4	38169.58	2.446	0
		Loc:Genotype	44	946801.1	21518.21	1.3789	0.0757
		Loc:Rep	4	612186.6	153046.66	9.8076	0
		Residuals	176	2746472.3	15604.96		

Table A.28 Mapping Population Trait Summaries - Caswell

Population	Trait	Loc	Average	SD	Range
Oil Mapping Pop 201	Hundred Seed Weight	Cas	16.32	1.79	12.05 - 20.9
	Oil (dry basis)		21.51	1.07	18.65 - 24.13
	Protein (dry basis)		45.48	2.08	38.44 - 51.08
Oil Mapping Pop 202	Hundred Seed Weight		17.98	1.66	11.65 - 22.5
	Oil (dry basis)		21.97	1.11	18.77 - 24.92
	Protein (dry basis)		45.75	2.07	39.16 - 50.55

Table A.29 Mapping Population Trait Summaries - Clayton

Population	Trait	Loc	Average	SD	Range
Oil Mapping Pop 201	Hundred Seed Weight	Cla	15.87	1.37	12.55 - 19.95
	Oil (dry basis)		20.34	0.78	18.45 - 22.36
	Protein (dry basis)		47.14	1.57	43.25 - 51.65
Oil Mapping Pop 202	Hundred Seed Weight		15.18	1.57	9.6 - 20.35
	Oil (dry basis)		20.59	0.93	17.91 - 22.73
	Protein (dry basis)		46.79	1.79	42.46 - 51.08

Table A.30 Mapping Population Trait Summaries - Combined Data

Population	Trait	Loc	Average	SD	Range
Oil Mapping Pop 201	Hundred Seed Weight	Combined	15.63	1.30	12.23 - 19.58
	Oil (dry basis)		21.28	0.79	19.26 - 23.34
	Protein (dry basis)		45.28	1.61	40.68 - 50.18
Oil Mapping Pop 202	Hundred Seed Weight		16.44	1.44	10.57 - 20.28
	Oil (dry basis)		21.76	0.98	18.94 - 25.23
	Protein (dry basis)		45.52	1.96	39.3 - 49.94

Table A.31 Mapping Population Trait Summaries - Plymouth

Population	Trait	Loc	Average	SD	Range
Oil Mapping Pop 201	Hundred Seed Weight	Ply	14.70	1.66	10.85 - 18.55
	Oil (dry basis)		22.00	1.15	19.27 - 25.21
	Protein (dry basis)		43.19	2.89	34.2 - 49.52
Oil Mapping Pop 202	Hundred Seed Weight		16.17	1.82	10.45 - 20.61
	Oil (dry basis)		22.68	1.31	18.73 - 25.76
	Protein (dry basis)		43.79	2.98	36.16 - 50.26

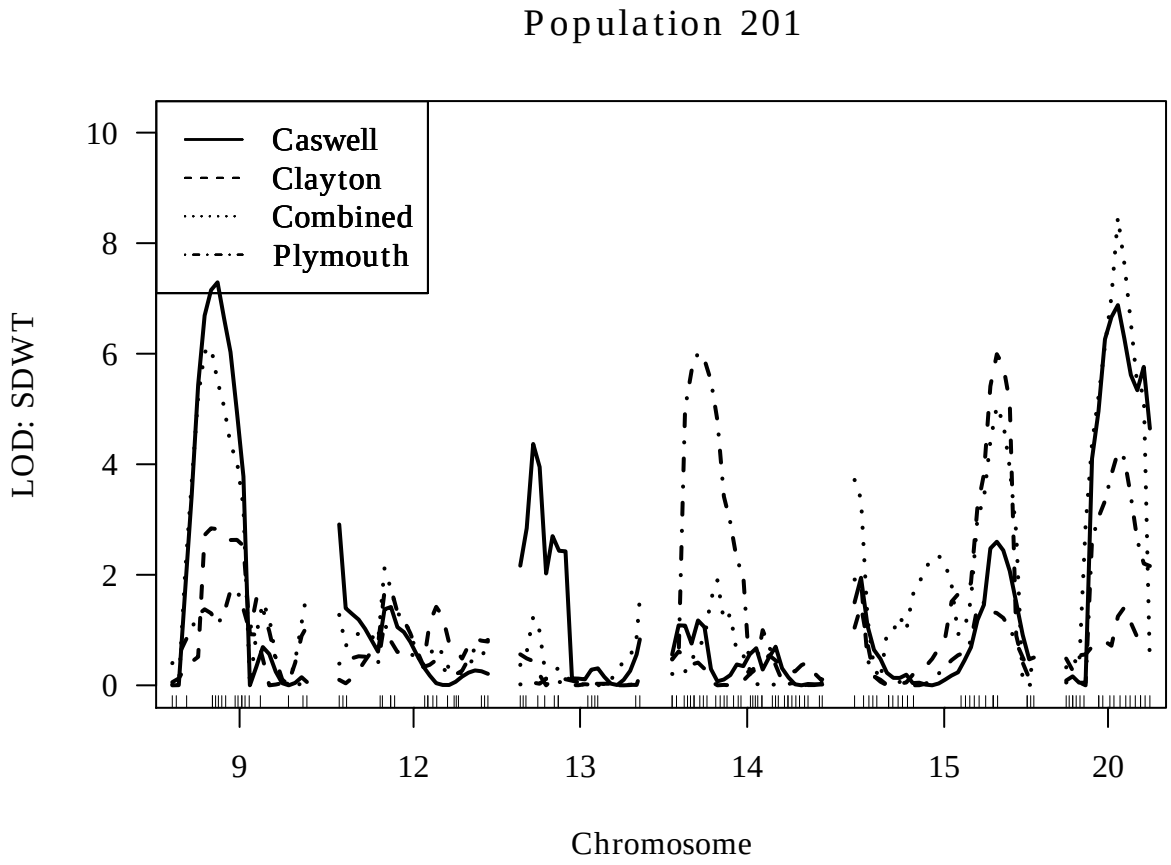

Figure A.6 LOD profiles for hundred seed weight (SDWT) for each environment for mapping population 201.

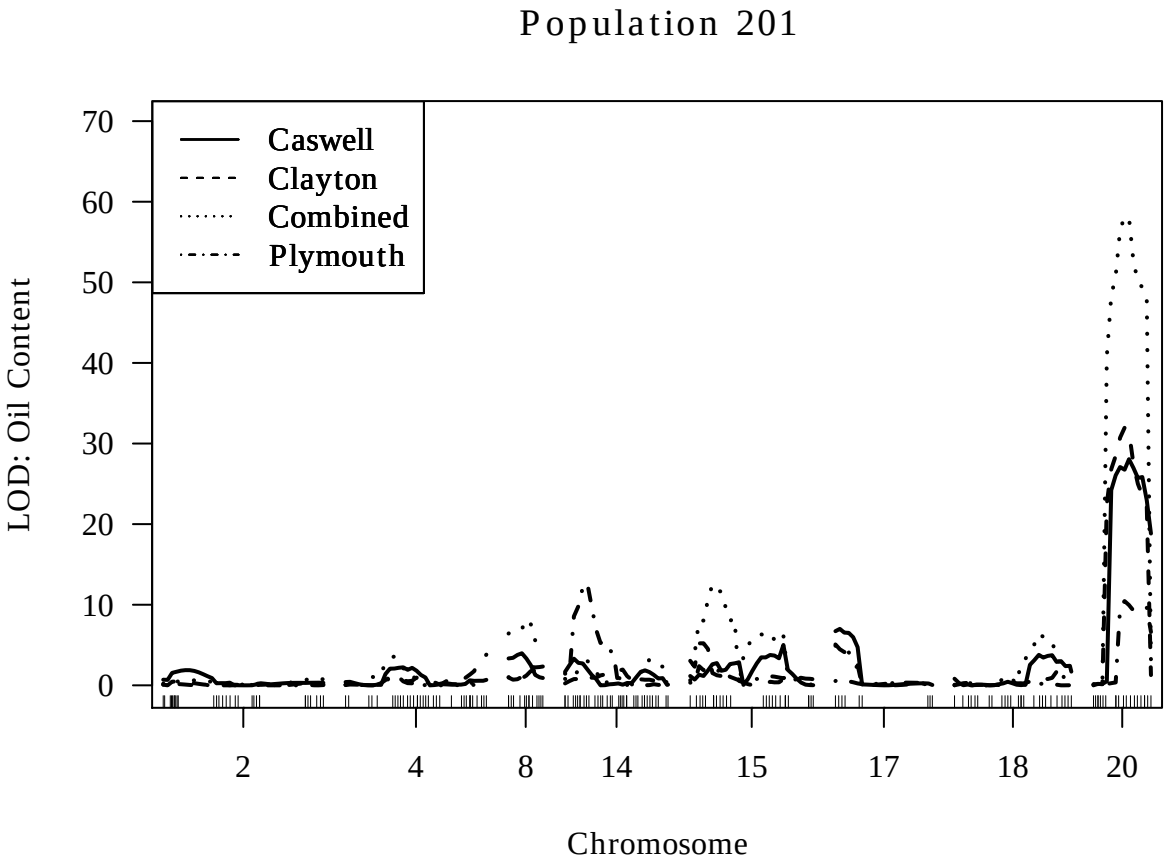

Figure A.7 LOD profiles for seed oil for each environment for mapping population 201.

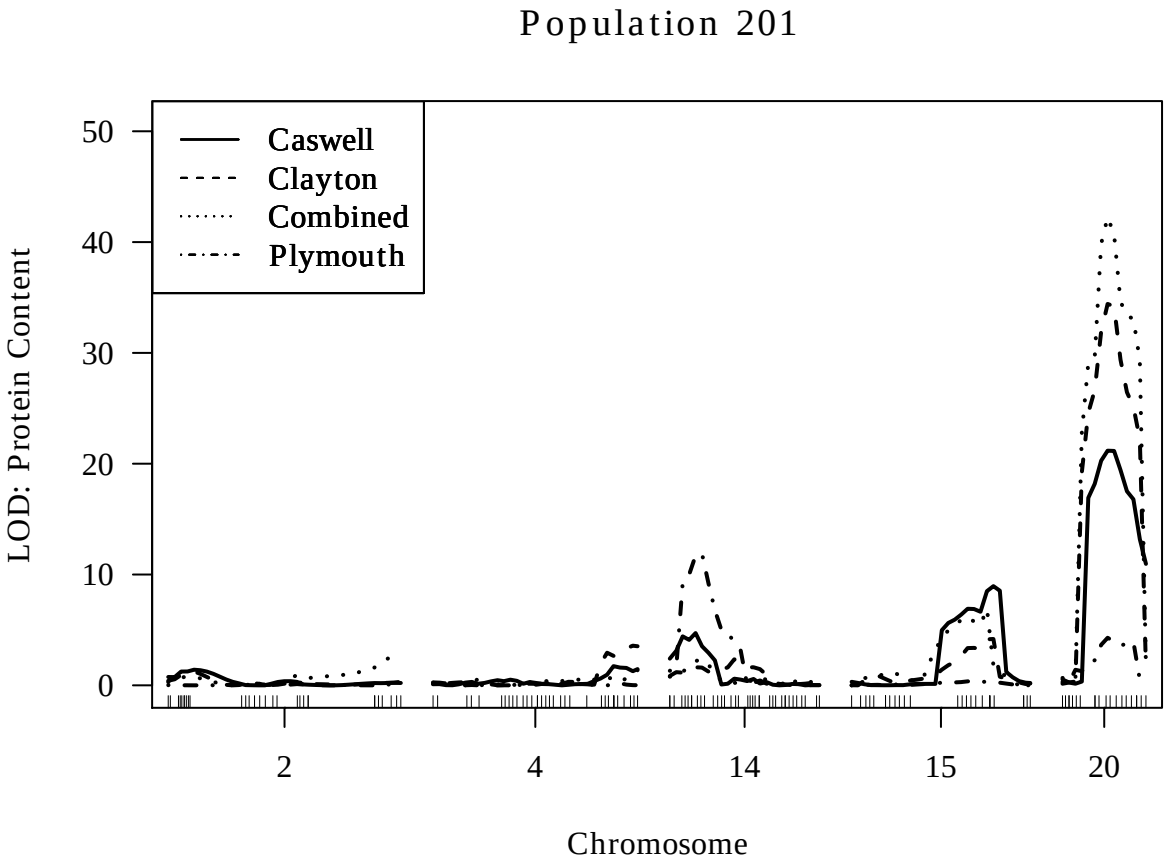

Figure A.8 LOD profiles for hundred seed protein for each environment for mapping population 201.

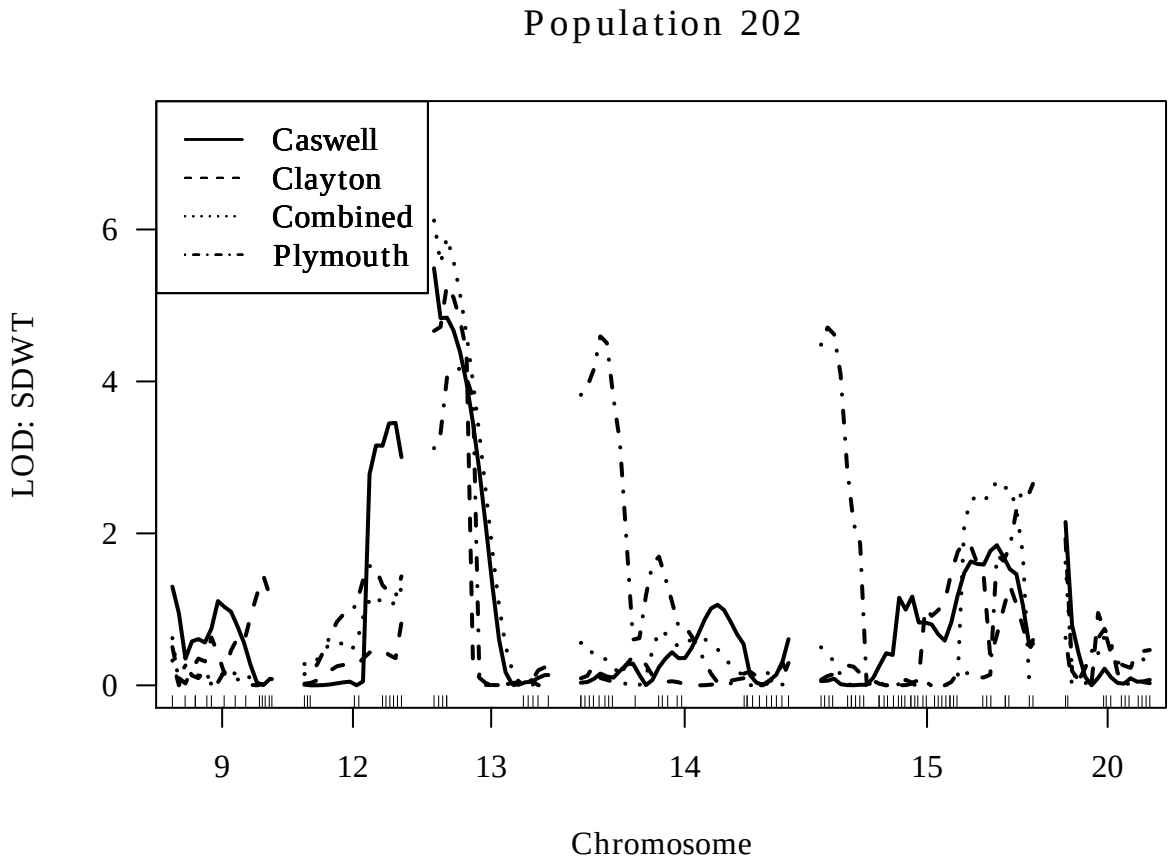

Figure A.9 LOD profiles for hundred seed weight (SDWT) for each environment for mapping population 202.

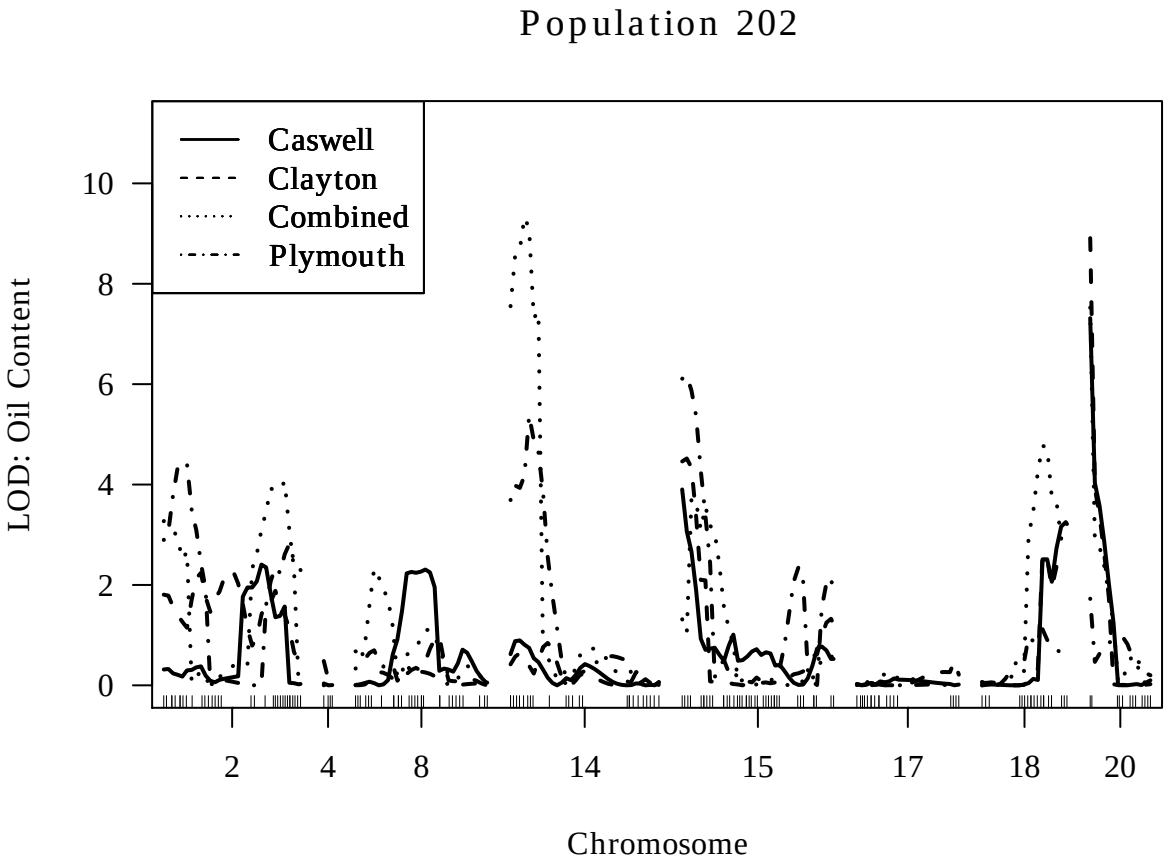

Figure A.10 LOD profiles for seed oil for each environment for mapping population 202.

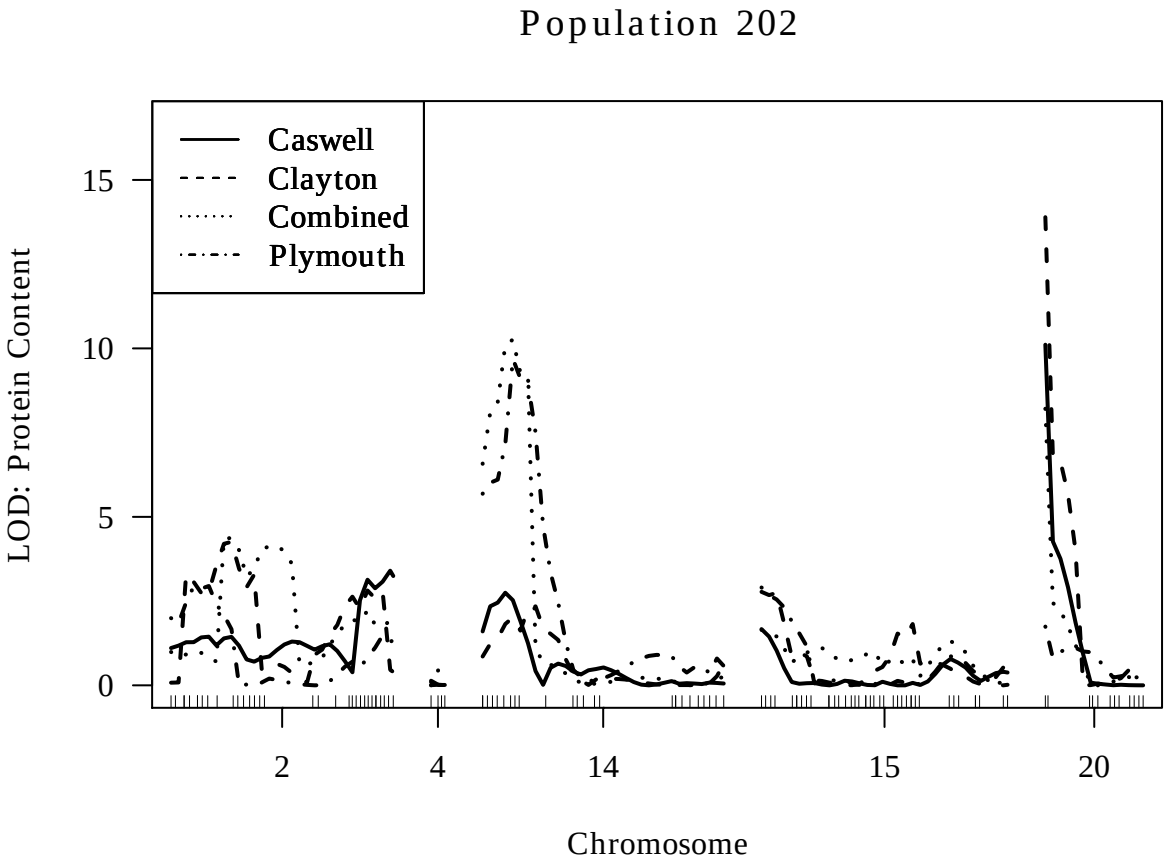

Figure A.11 LOD profiles for seed protein for each environment for mapping population 202.

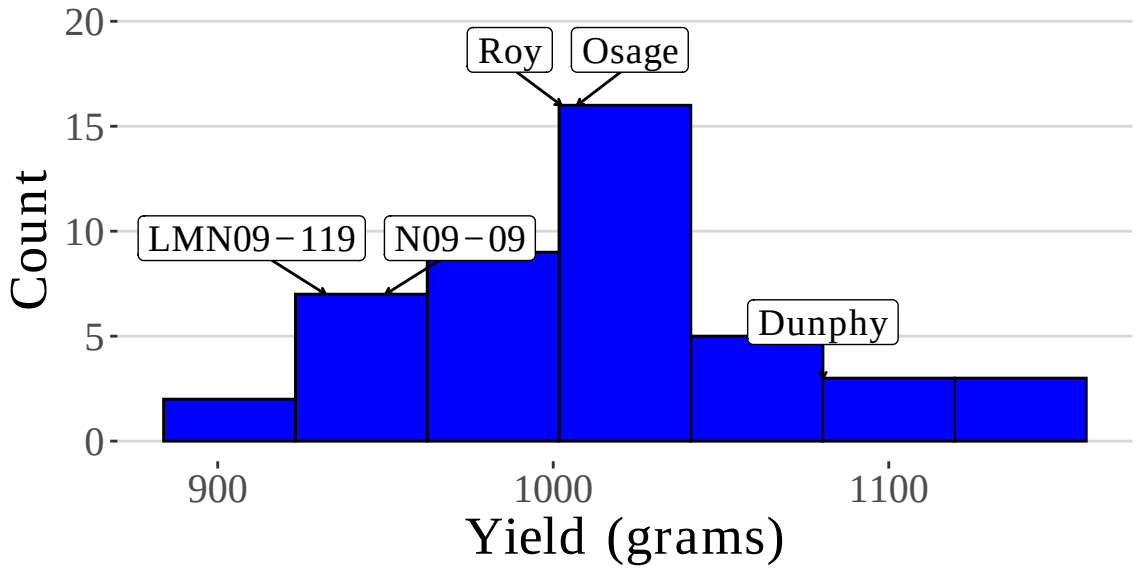

Figure A.12 Distribution of Yield for Yield Test 1. Yield checks and population parents are labeled.

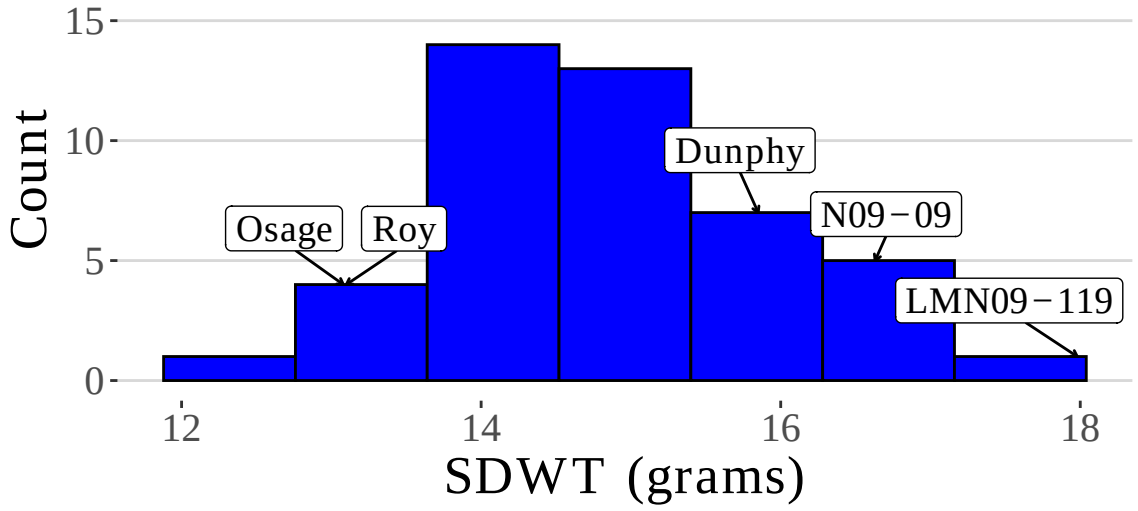

Figure A.13 Distribution of Hundred Seed Weight for Yield Test 1. Yield checks and population parents are labeled.

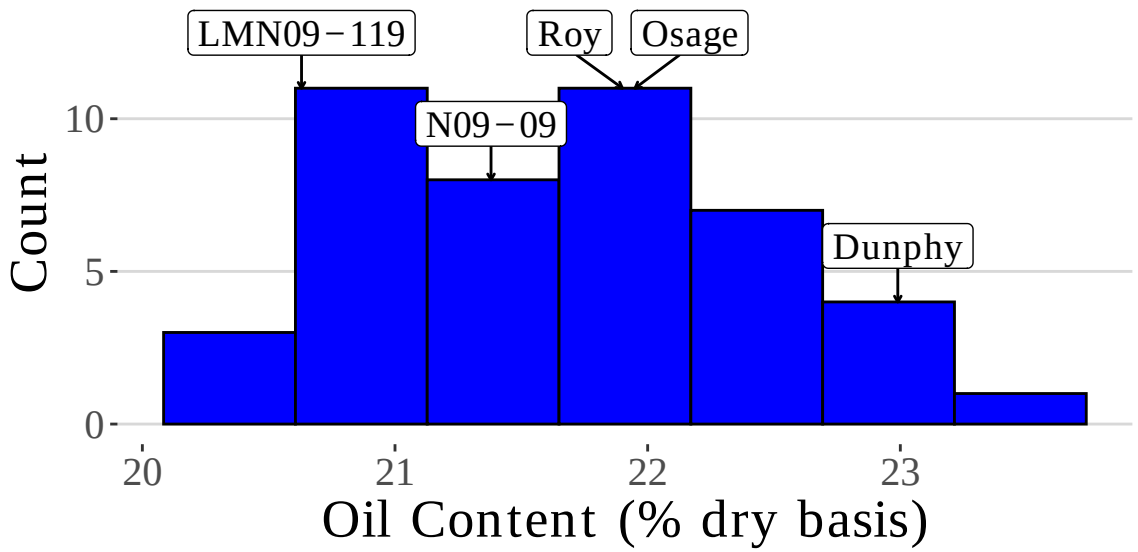

Figure A.14 Distribution of Oil Content for Yield Test 1. Yield checks and population parents are labeled.

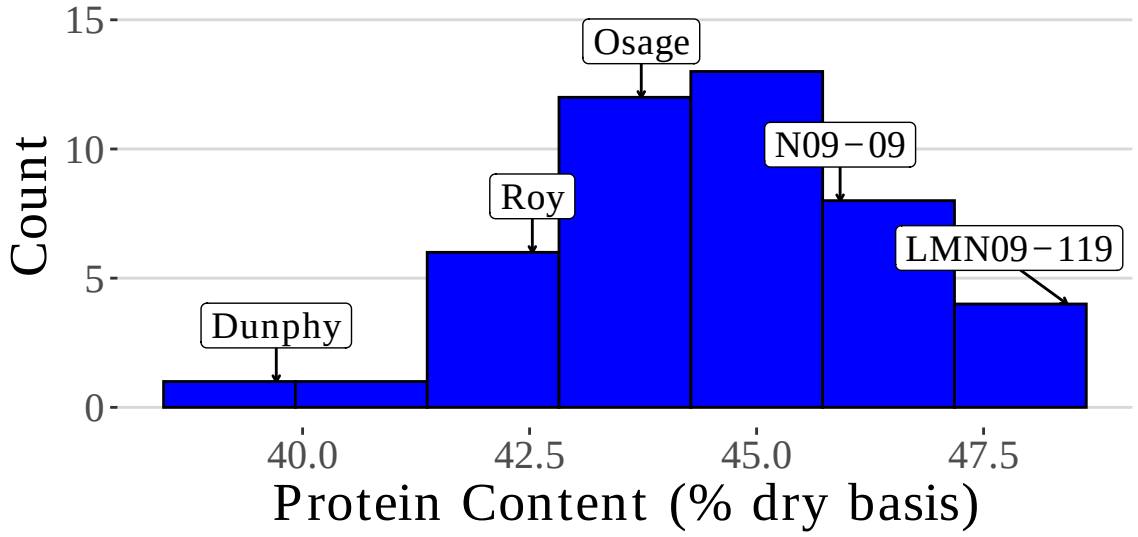

Figure A.15 Distribution of Protein Content for Yield Test 1. Yield checks and population parents are labeled.

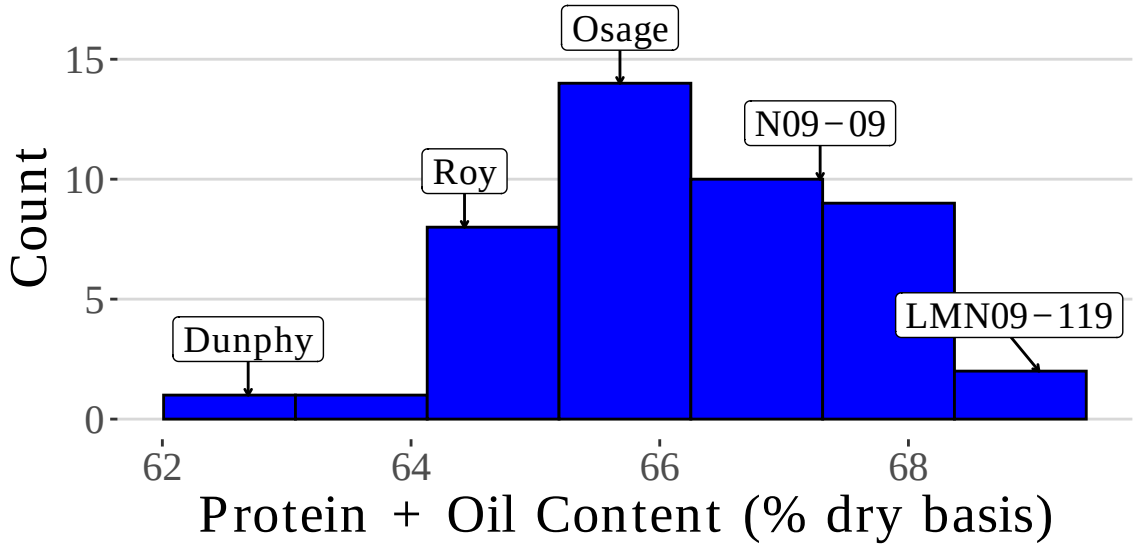

Figure A.16 Distribution of Protein + Oil for Yield Test 1. Yield checks and population parents are labeled.

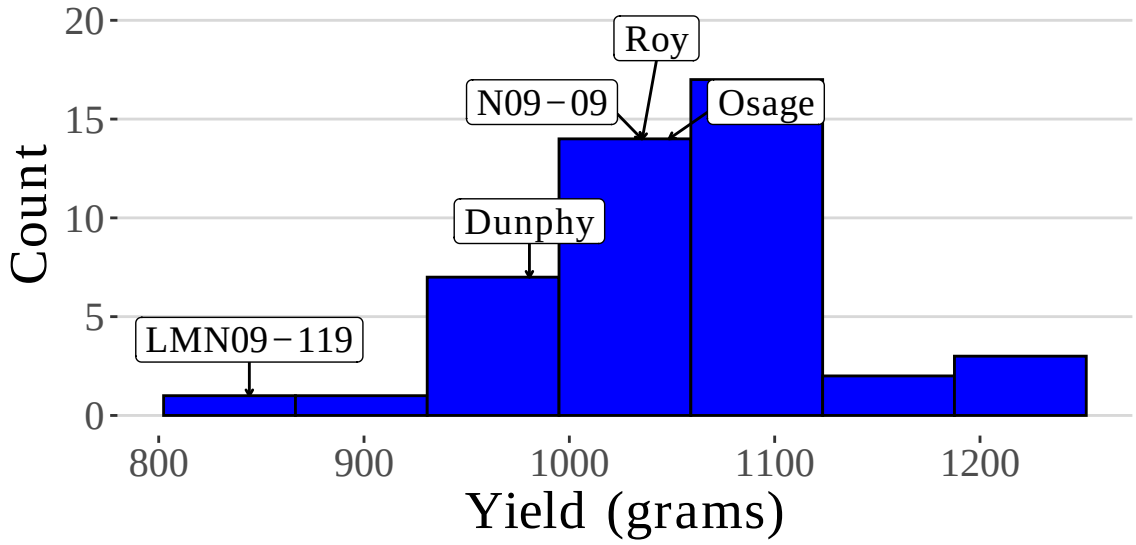

Figure A.17 Distribution of Yield for Yield Test 2. Yield checks and population parents are labeled.

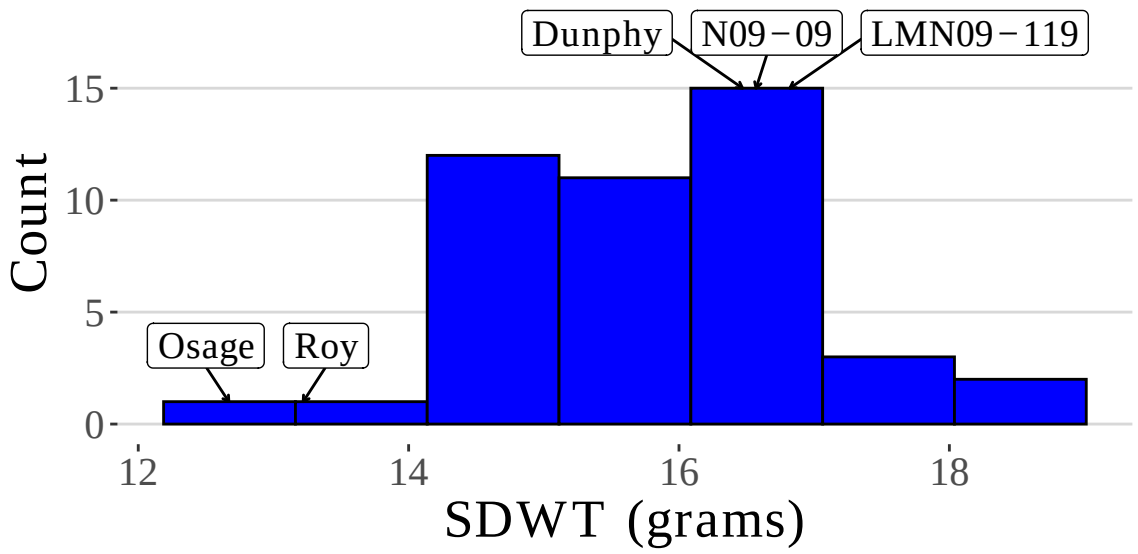

Figure A.18 Distribution of Hundred Seed Weight for Yield Test 2. Yield checks and population parents are labeled.

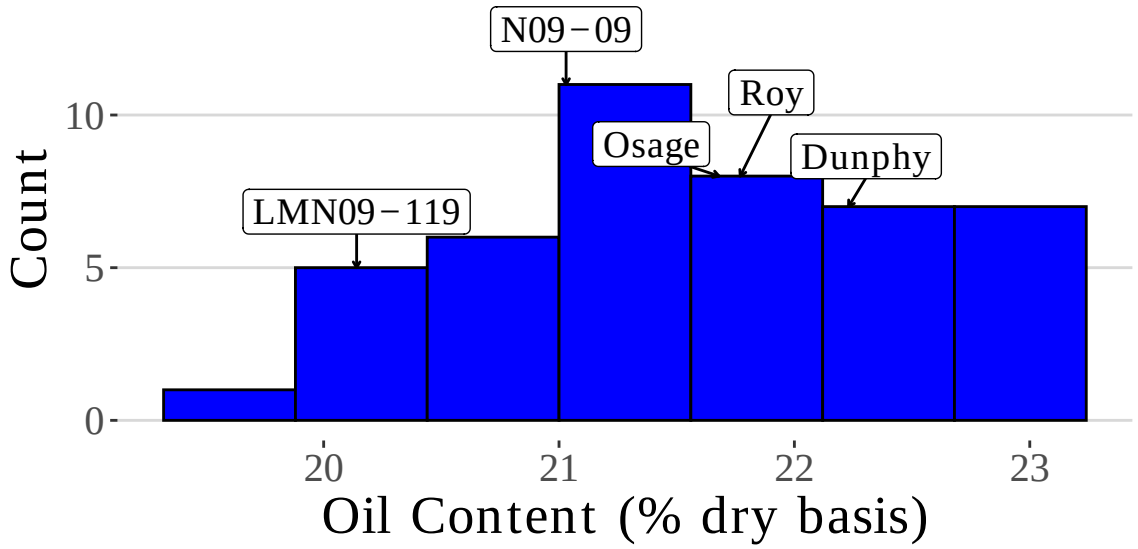

Figure A.19 Distribution of Oil Content for Yield Test 2. Yield checks and population parents are labeled.

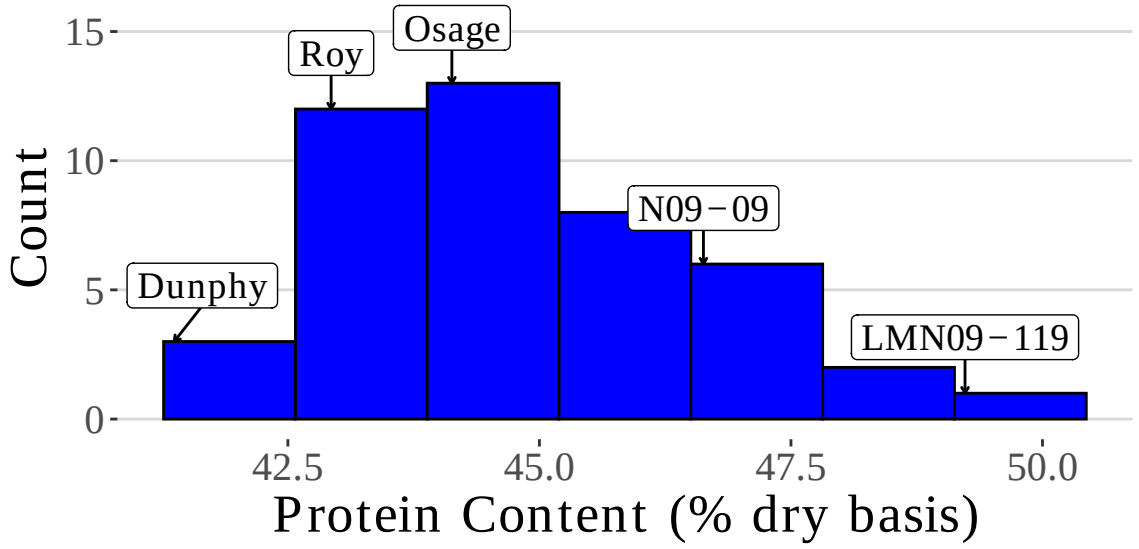

Figure A.20 Distribution of Protein Content for Yield Test 2. Yield checks and population parents are labeled.

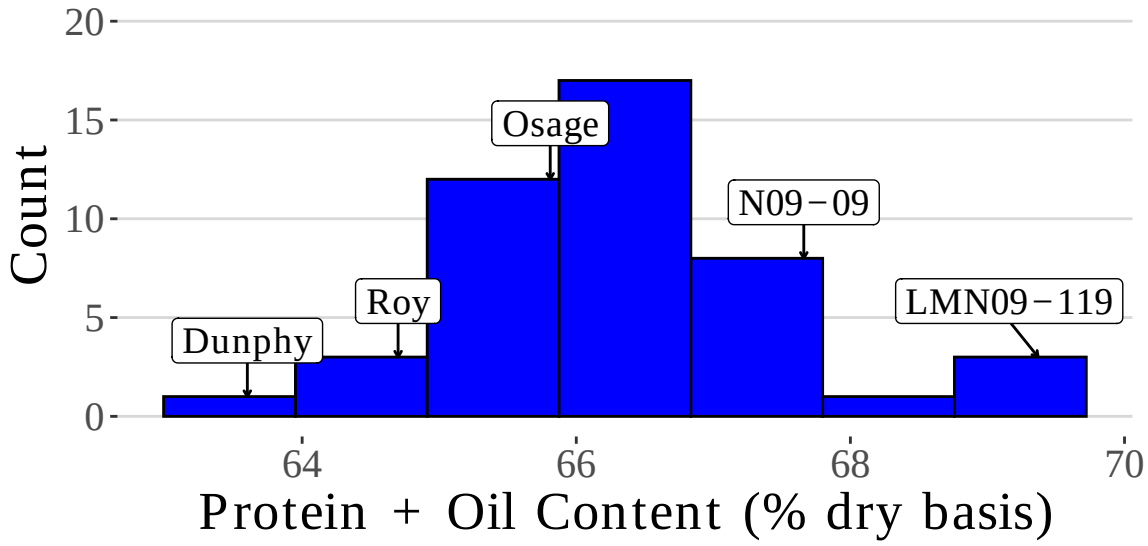

Figure A.21 Distribution of Protein + Oil for Yield Test 2. Yield checks and population parents are labeled.

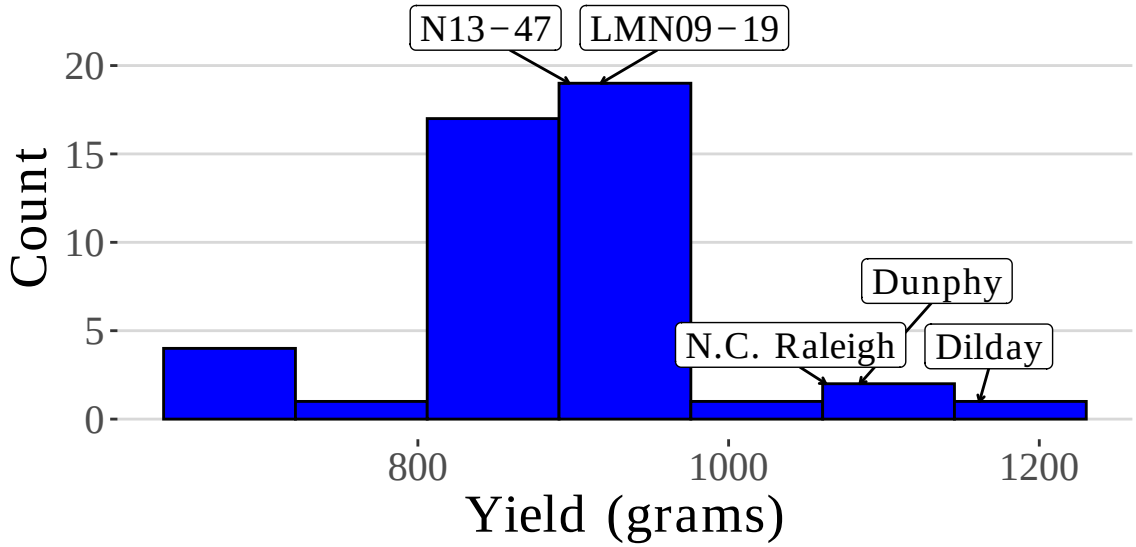

Figure A.22 Distribution of Yield for Yield Test 3. Yield checks and population parents are labeled.

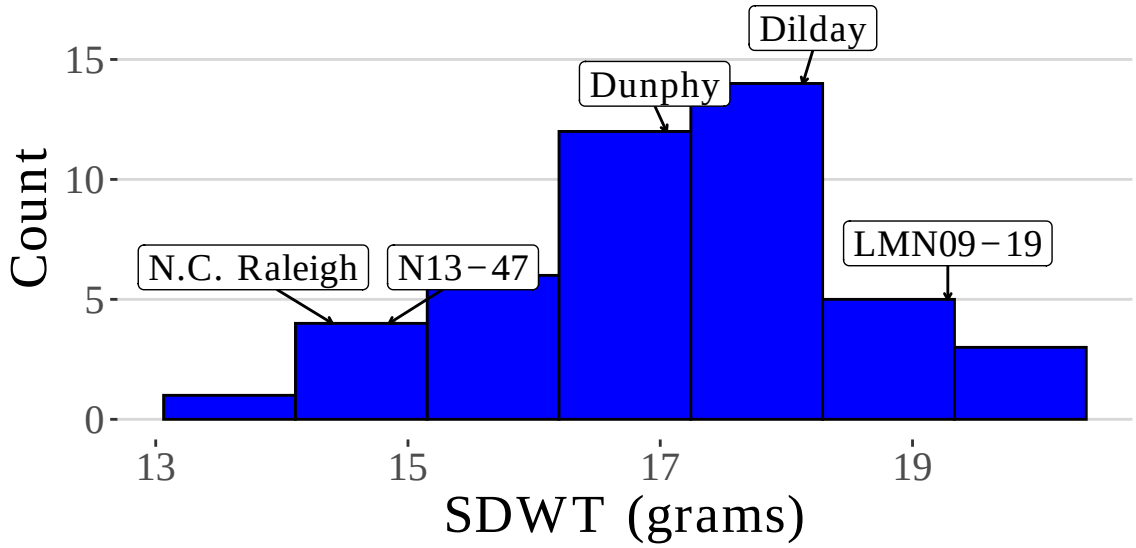

Figure A.23 Distribution of Hundred Seed Weight for Yield Test 3. Yield checks and population parents are labeled.

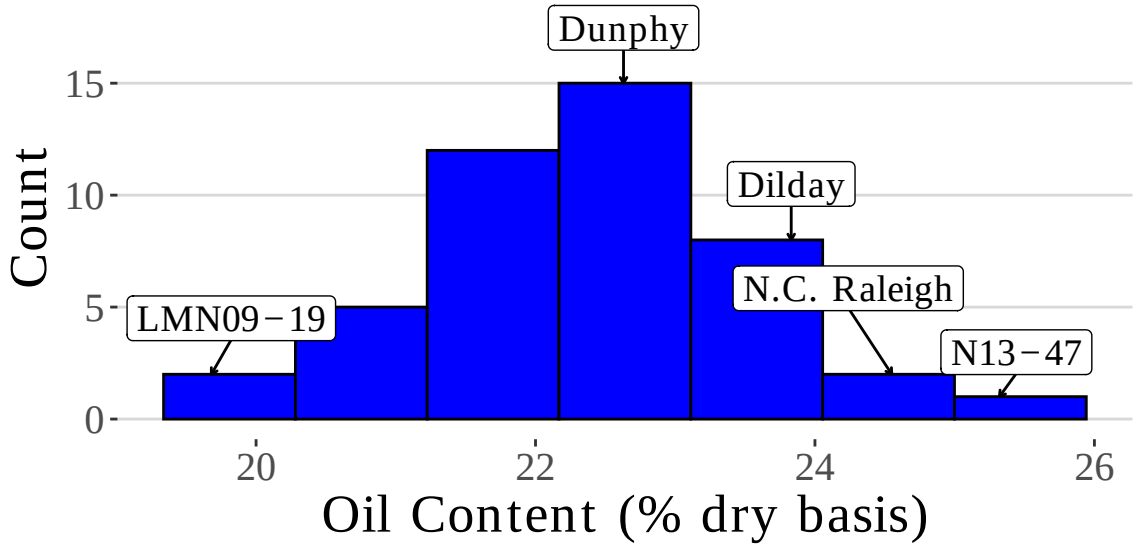

Figure A.24 Distribution of Oil Content for Yield Test 3. Yield checks and population parents are labeled.

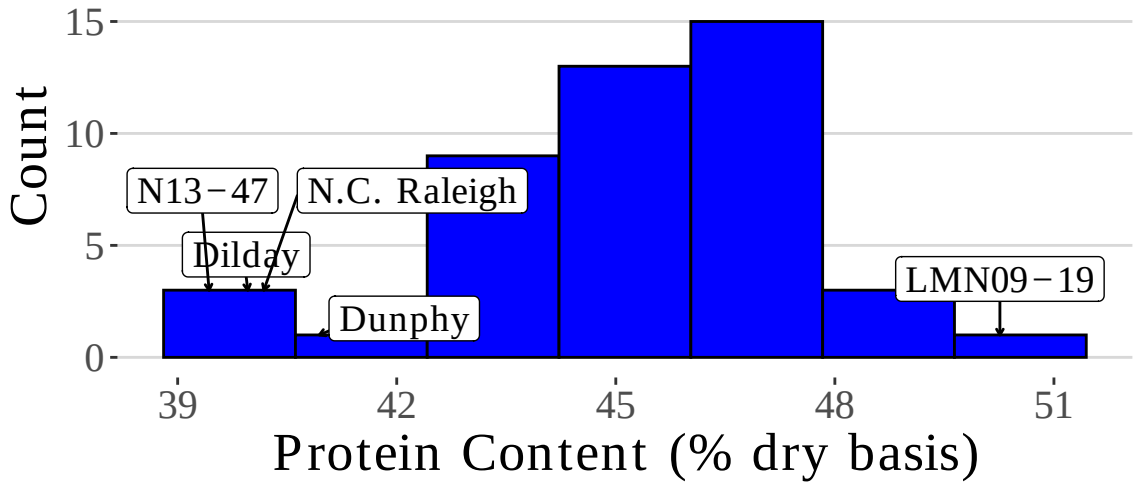

Figure A.25 Distribution of Protein Content for Yield Test 3. Yield checks and population parents are labeled.

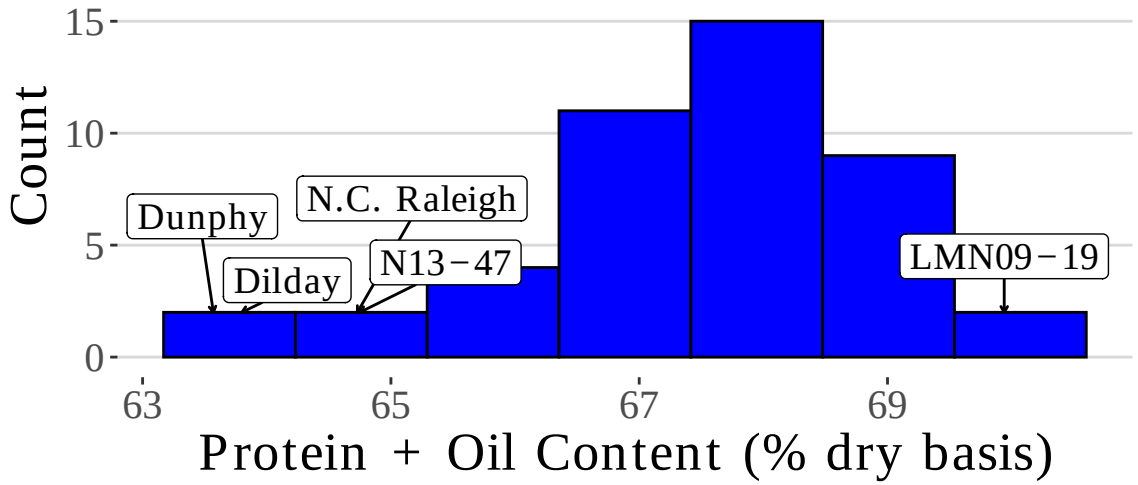

Figure A.26 Distribution of Protein + Oil for Yield Test 3. Yield checks and population parents are labeled.

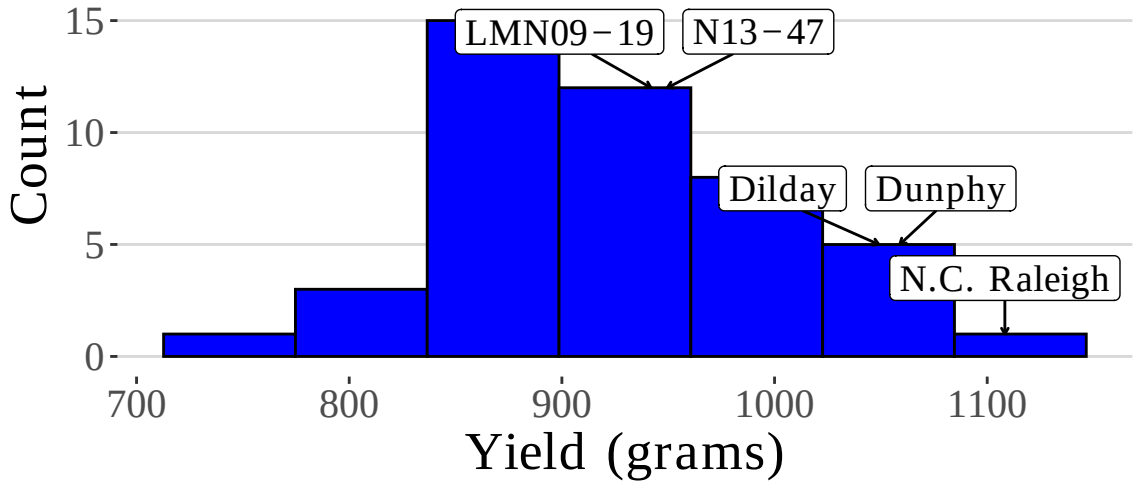

Figure A.27 Distribution of Yield for Yield Test 4. Yield checks and population parents are labeled.

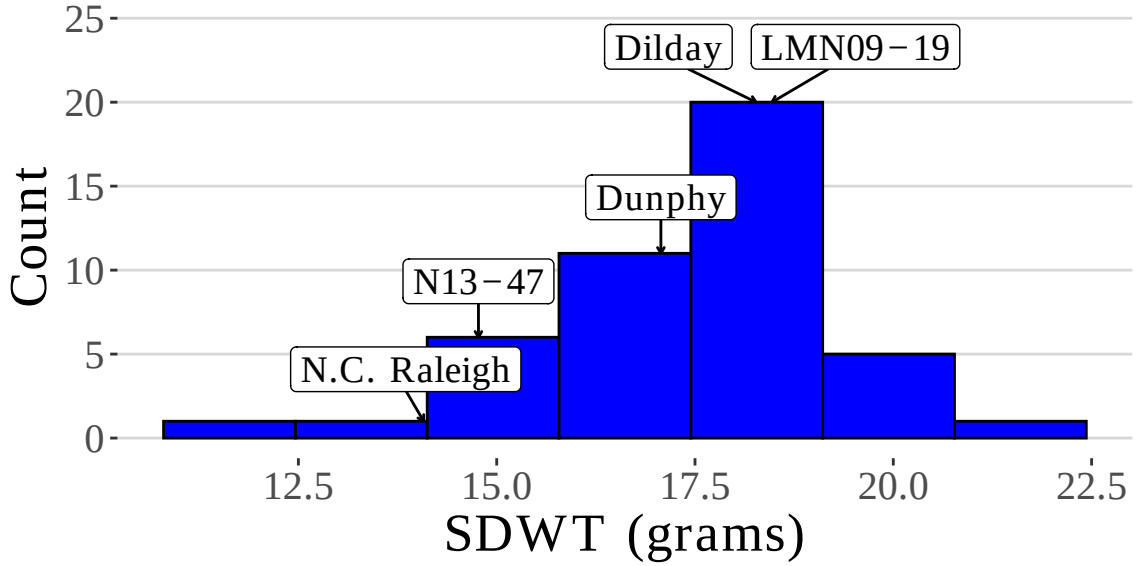

Figure A.28 Distribution of Hundred Seed Weight for Yield Test 4. Yield checks and population parents are labeled.

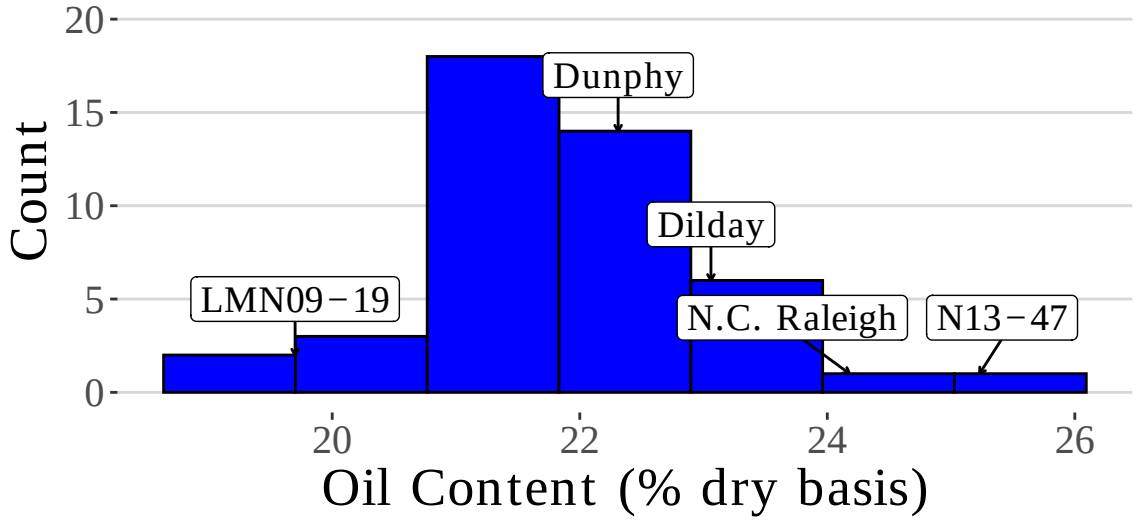

Figure A.29 Distribution of Oil Content for Yield Test 4. Yield checks and population parents are labeled.

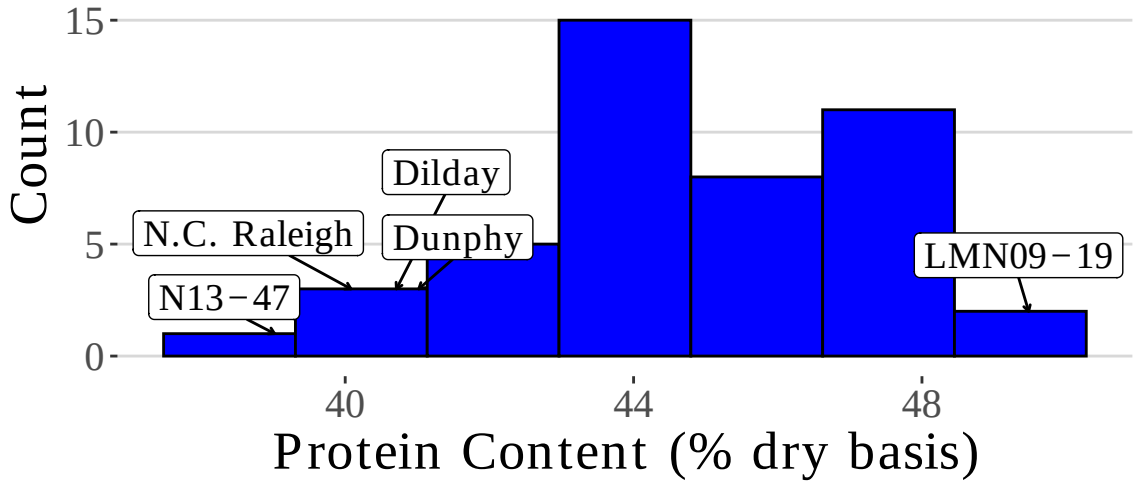

Figure A.30 Distribution of Protein Content for Yield Test 4. Yield checks and population parents are labeled.

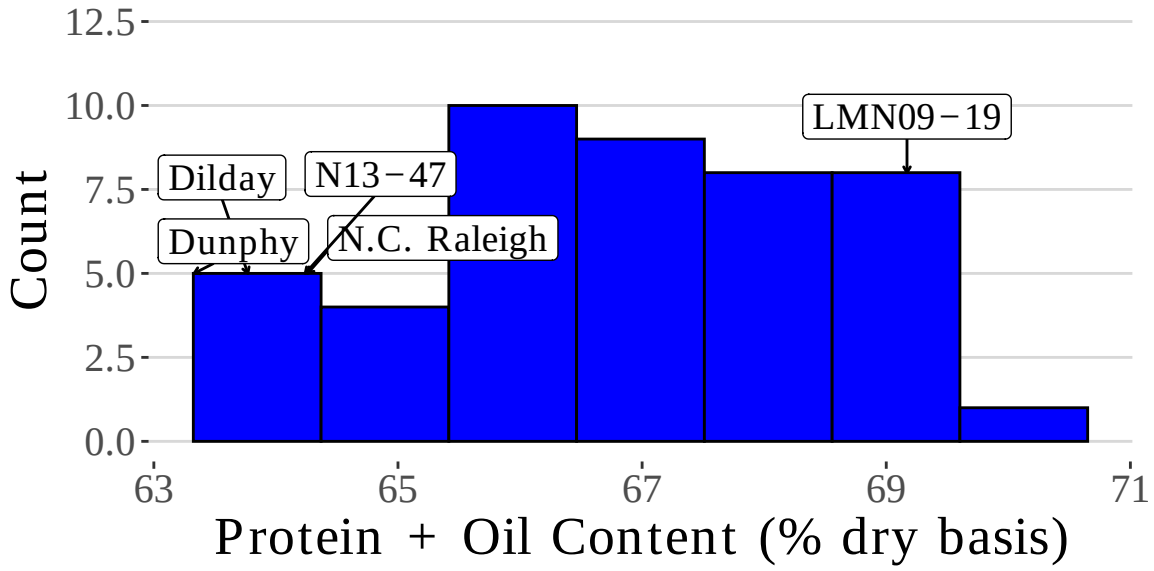

Figure A.31 Distribution of Protein + Oil for Yield Test 4. Yield checks and population parents are labeled.

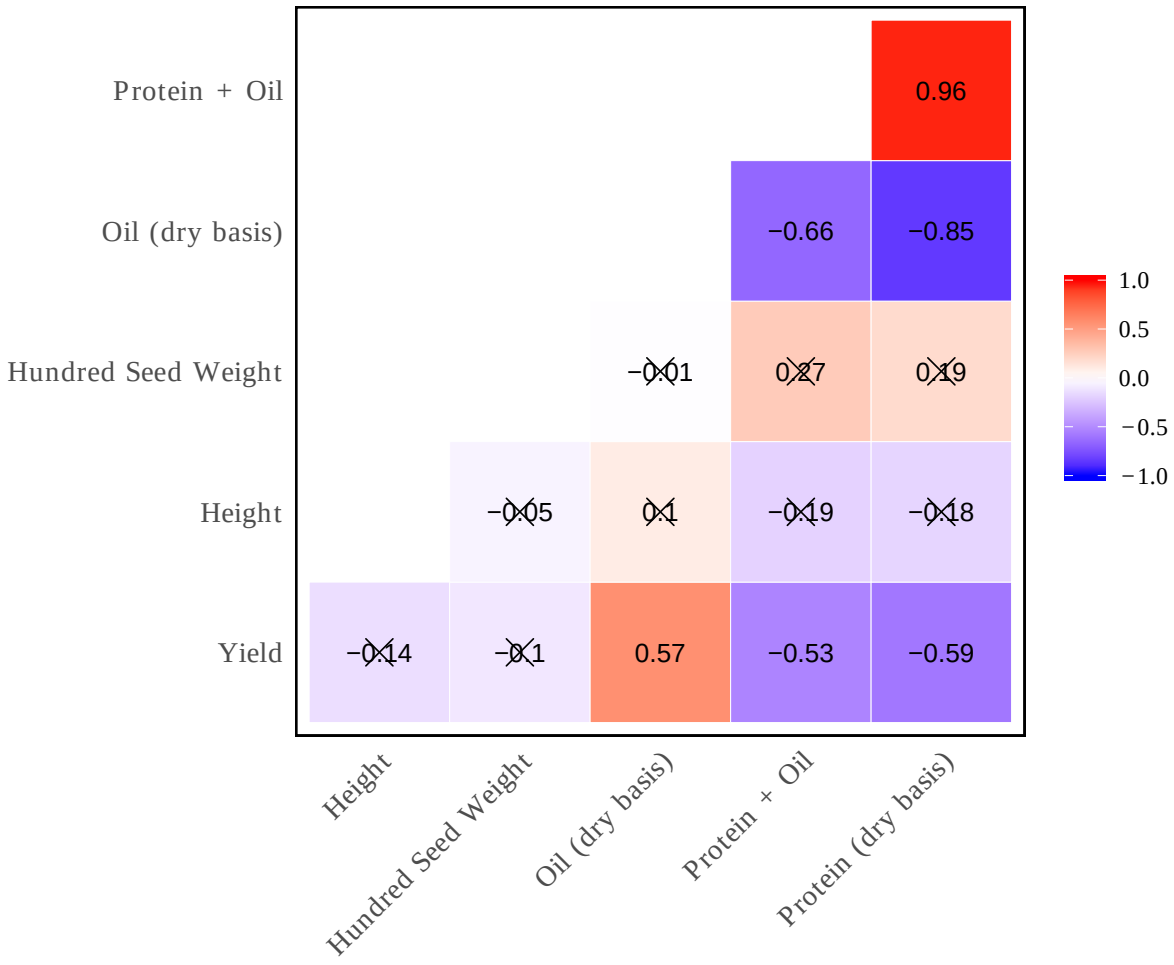

Figure A.32 Pairwise Trait Correlations for Yield Test 1. Non-significant, ($p > 0.05$), correlation coefficients are identified with an "X".

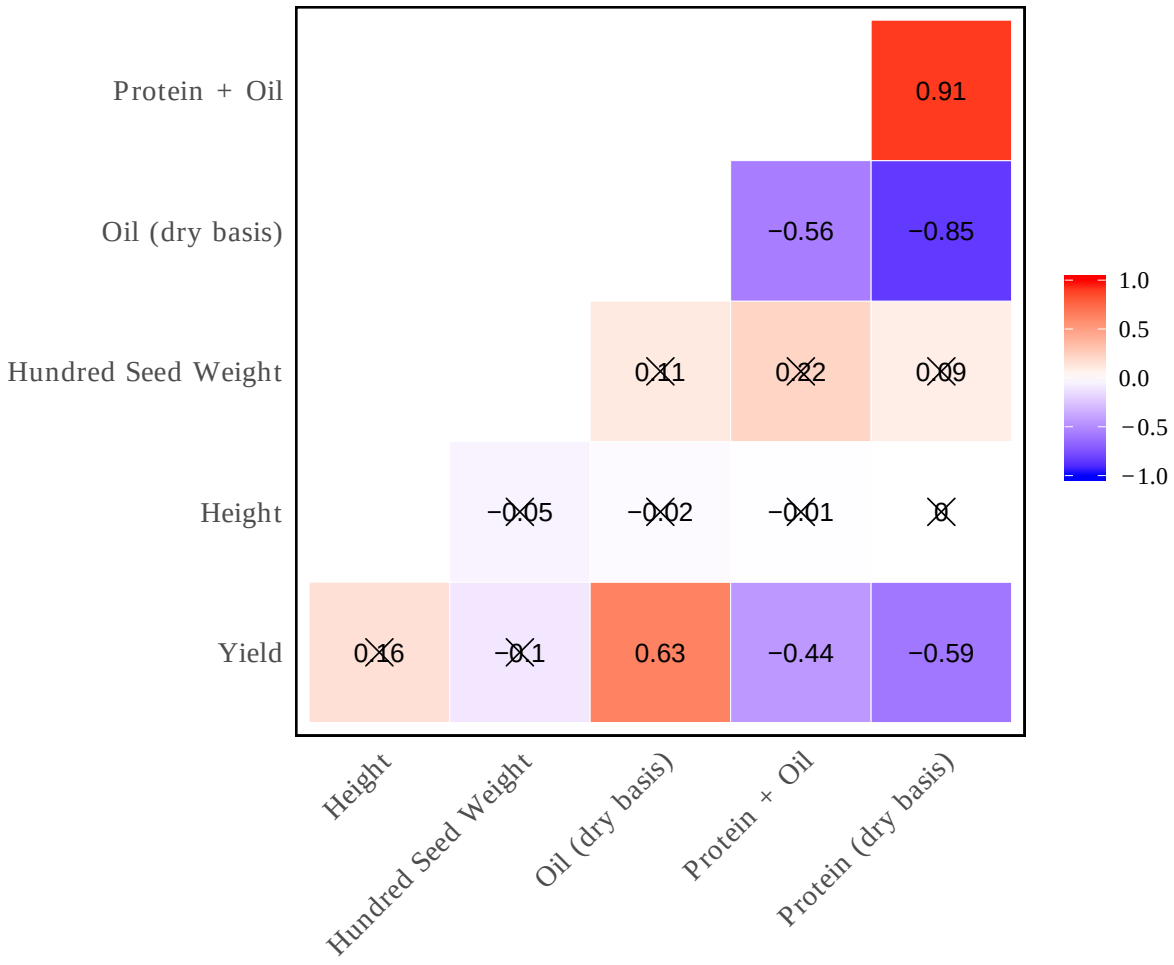

Figure A.33 Pairwise Trait Correlations for Yield Test 2. Non-significant, ($p > 0.05$), correlation coefficients are identified with an "X".

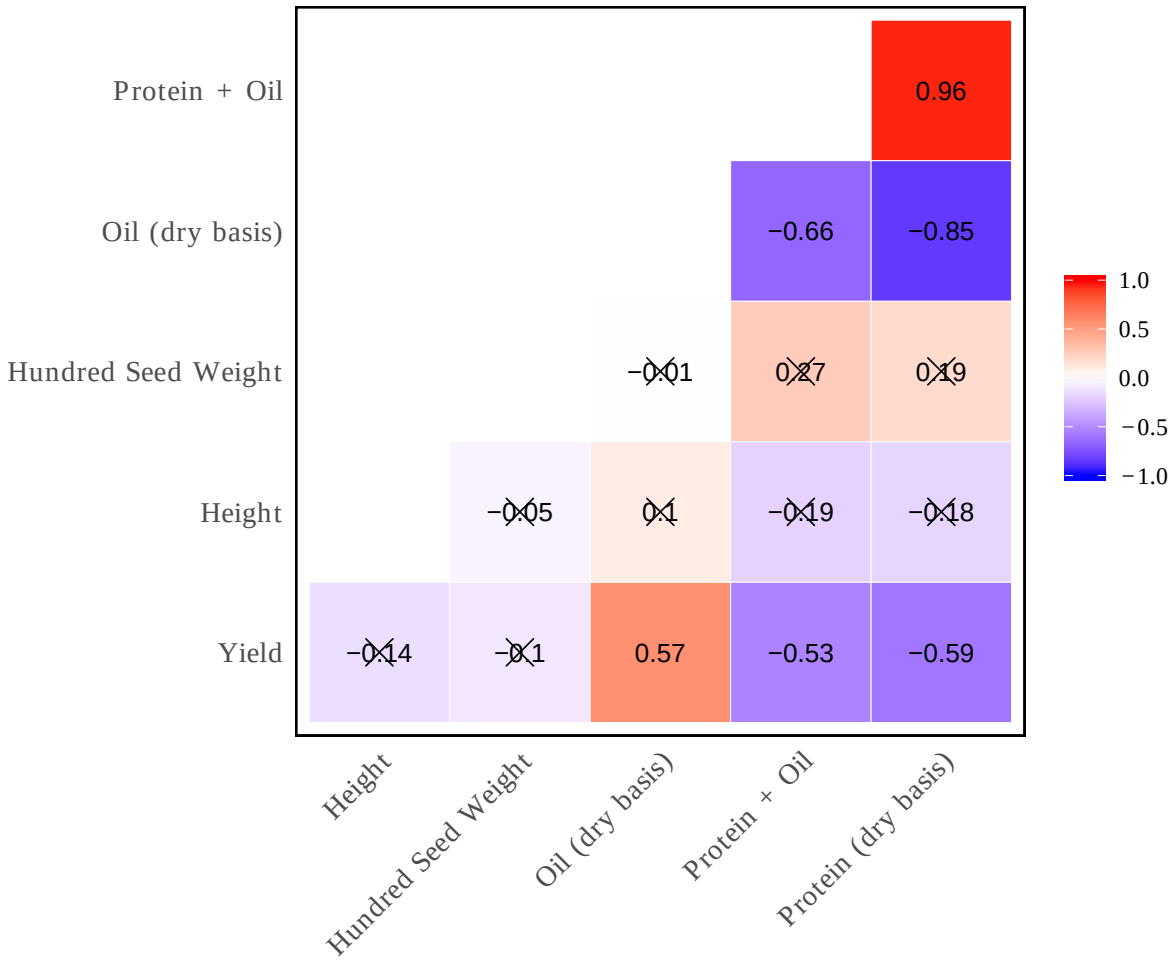

Figure A.34 Pairwise Trait Correlations for Yield Test 3. Non-significant, ($p > 0.05$), correlation coefficients are identified with an "X".

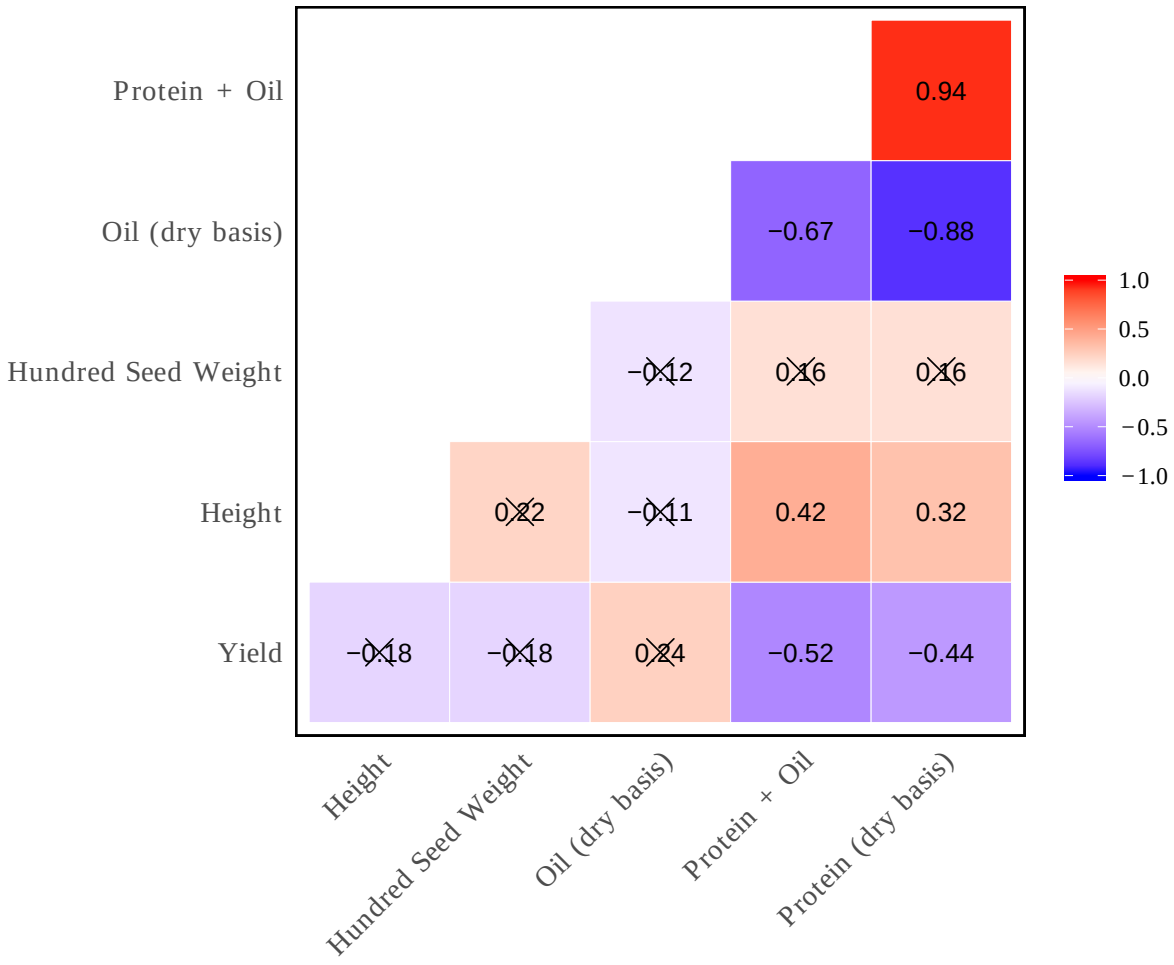

Figure A.35 Pairwise Trait Correlations for Yield Test 4. Non-significant, ($p > 0.05$), correlation coefficients are identified with an "X".

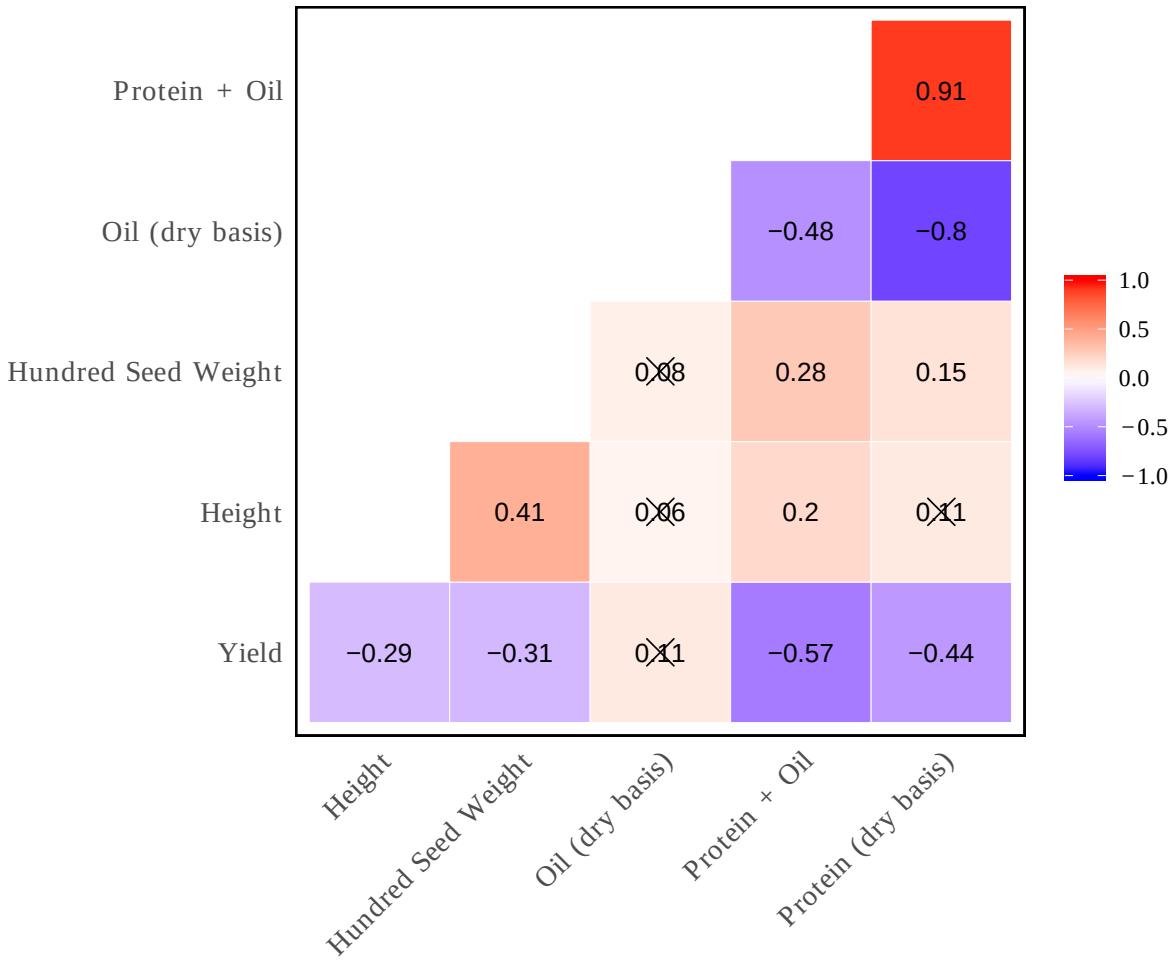

Figure A.36 Pairwise Trait Correlations for All Tests. Non-significant, ($p > 0.05$), correlation coefficients are identified with an "X".

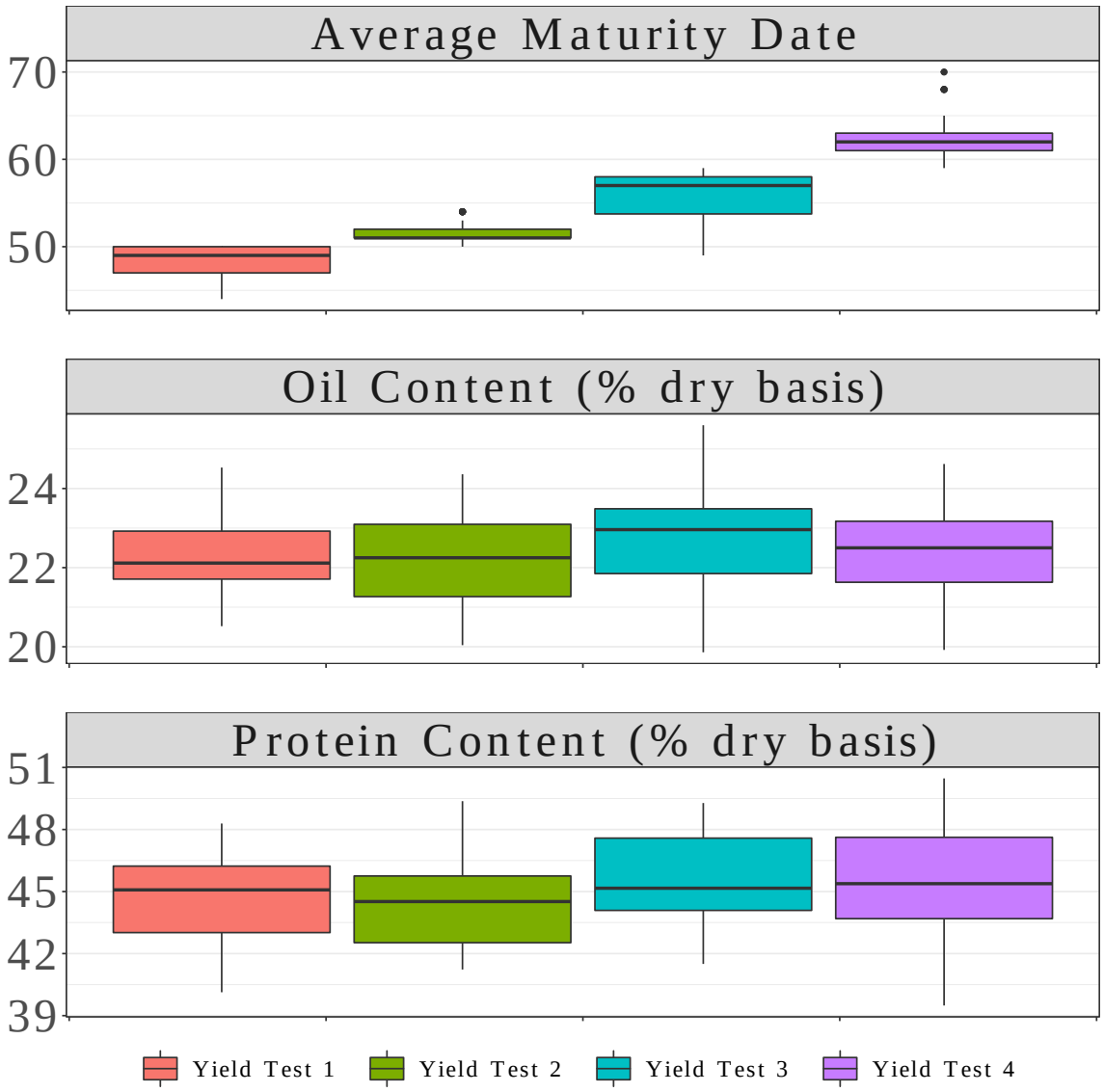

Figure A.37 Trait distributions for the lines included in soybean yield tests 1-4 in 2019. The data used in this plot was collected from plant rows in 2018. Maturity date is measured as the number of days after September 1 when 90% of the plants in a plot have reached the R8 growth stage. Protein and oil content are measured on a % dry basis.

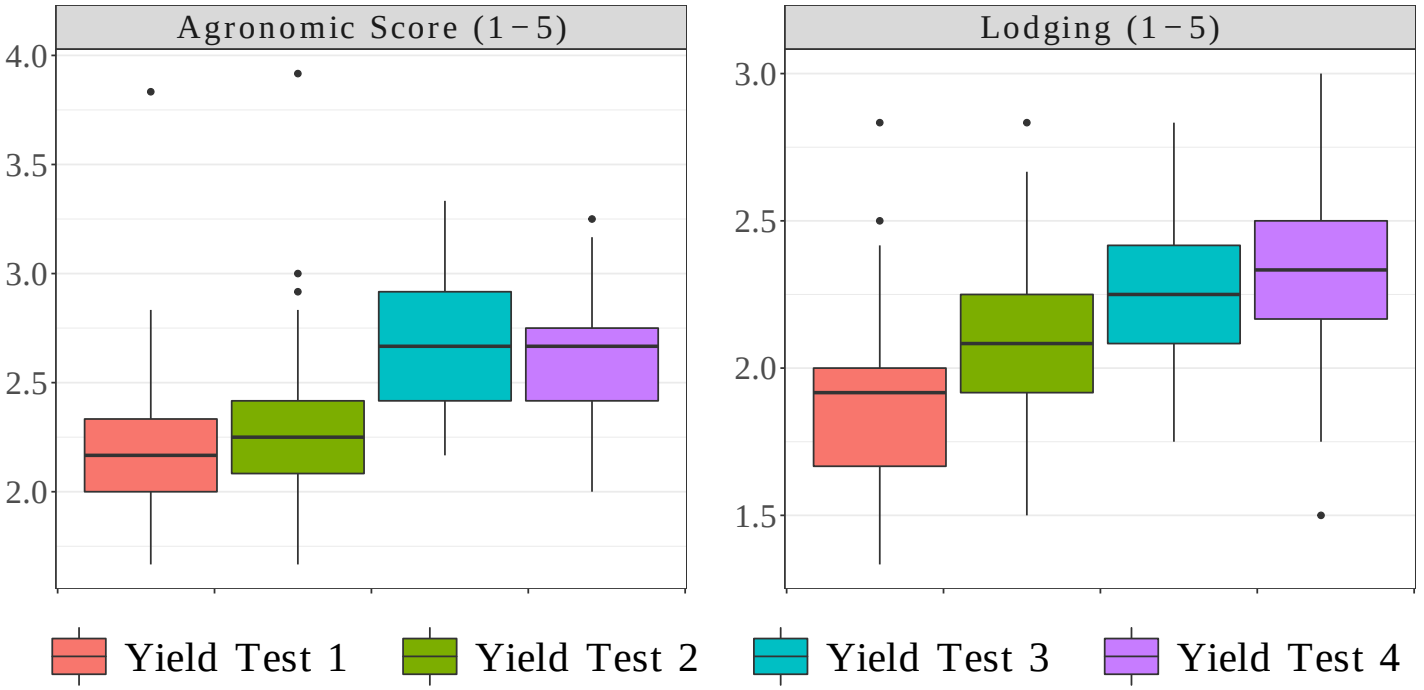

Figure A.38 Agronomic trait distributions for soybean yield tests 1-4 in 2019. The data used in this plot was collected from yield trials in two locations in 2019. Lodging and agronomic score are measured on a scale of 1-5 where a score of 1 indicates optimal agronomic traits and completely upright plants and a score of 5 indicates poor agronomic qualities and prostrate plants for the two traits, respectively.

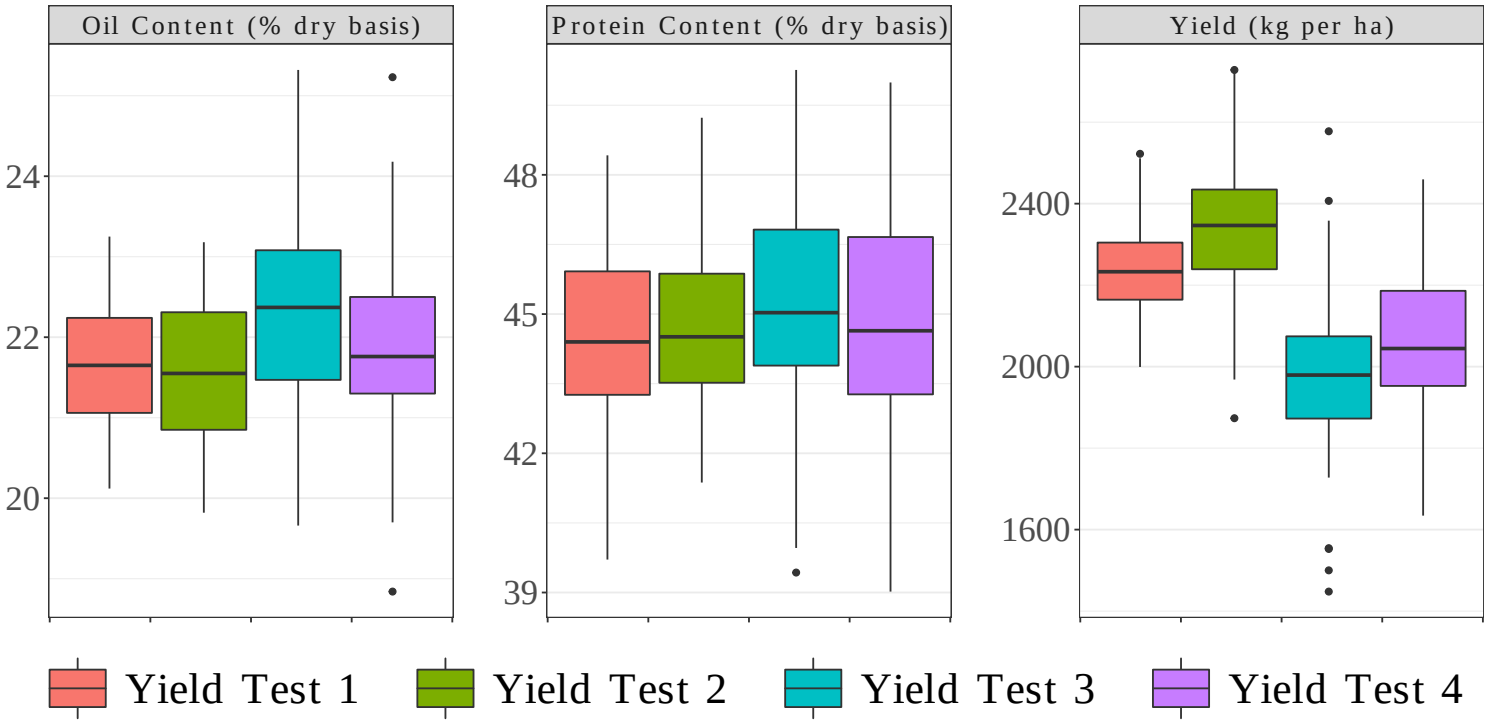

Figure A.39 Seed trait distributions for soybean yield tests 1-4 in 2019. The data used in this plot was collected from yield trials in two locations in 2019. Seed protein and seed oil are measured on a % dry basis, yield is measured in units of kg/ha.

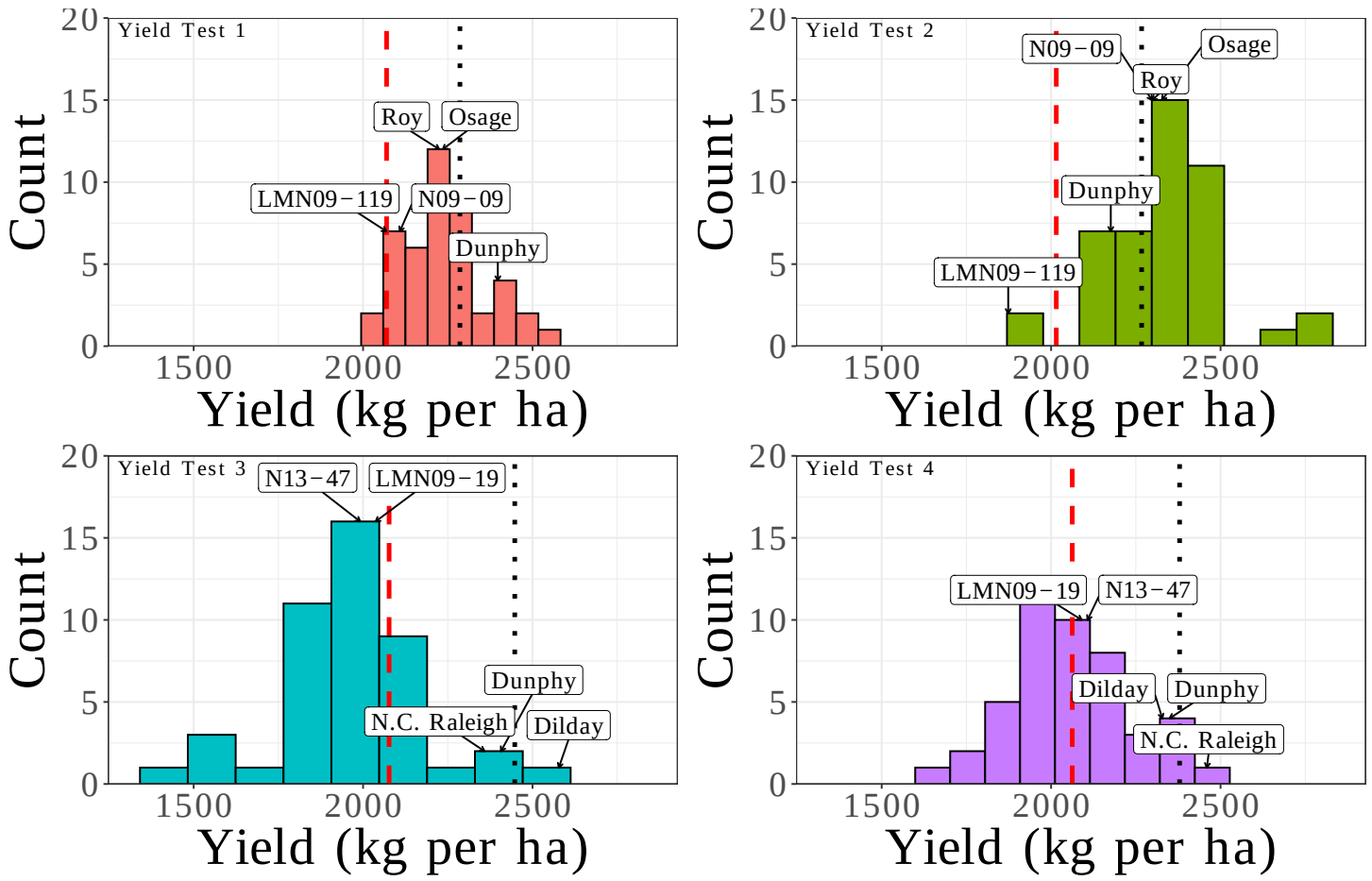

Figure A.40 Comparison of yield between soybean yield tests 1-4. Each test was grown in two locations in North Carolina. Parent and check cultivars are labeled in each distribution. The dotted line indicates the average yield of the check cultivars and the dashed line represents one least-significant difference less than this average.

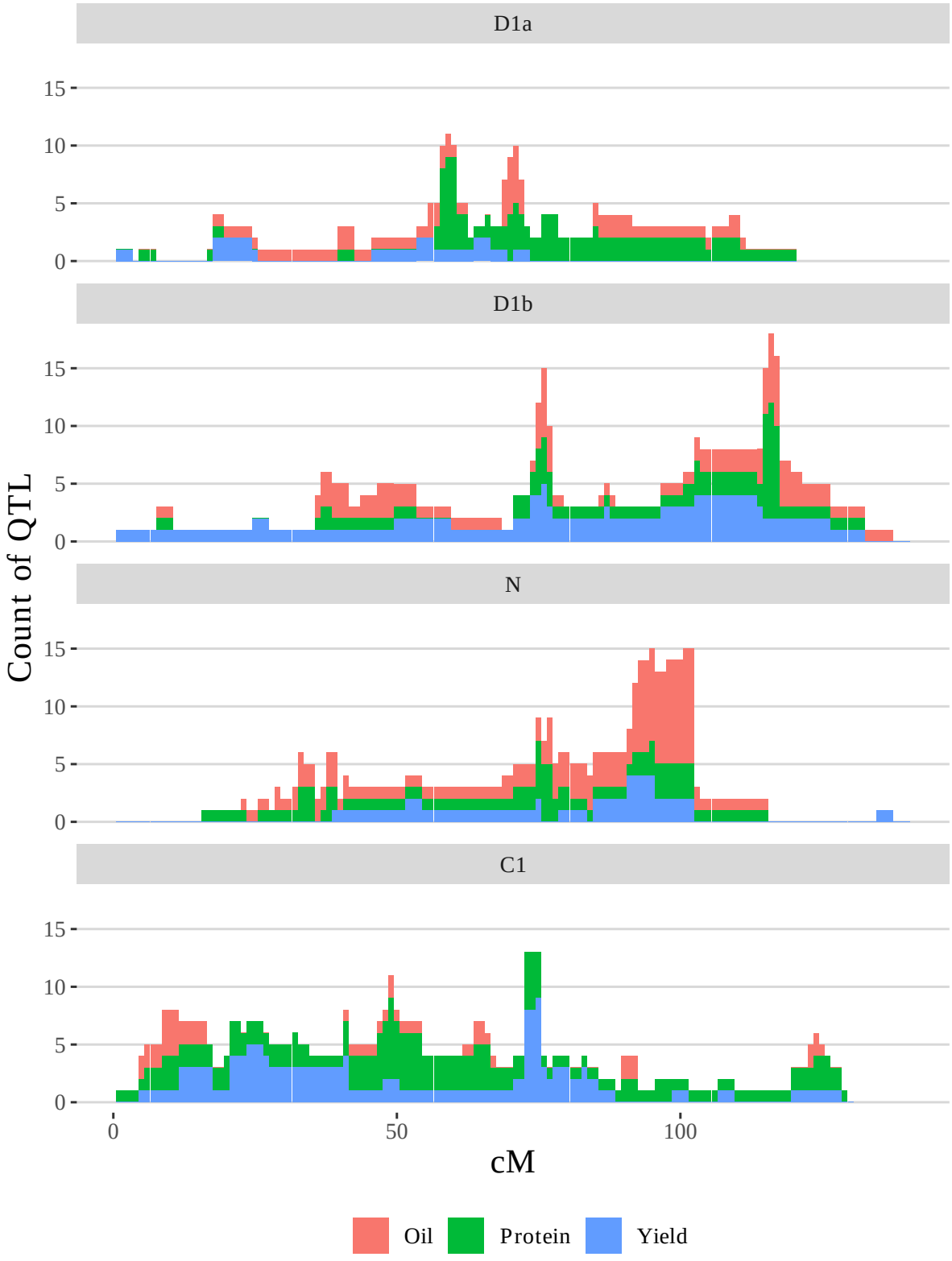

Figure A.41 Count of Previously Discovered Soybean QTL for Oil, Protein, and Yield Traits for Chromosomes 1-4 of the Soybean GmConsensus 4.0 Map.

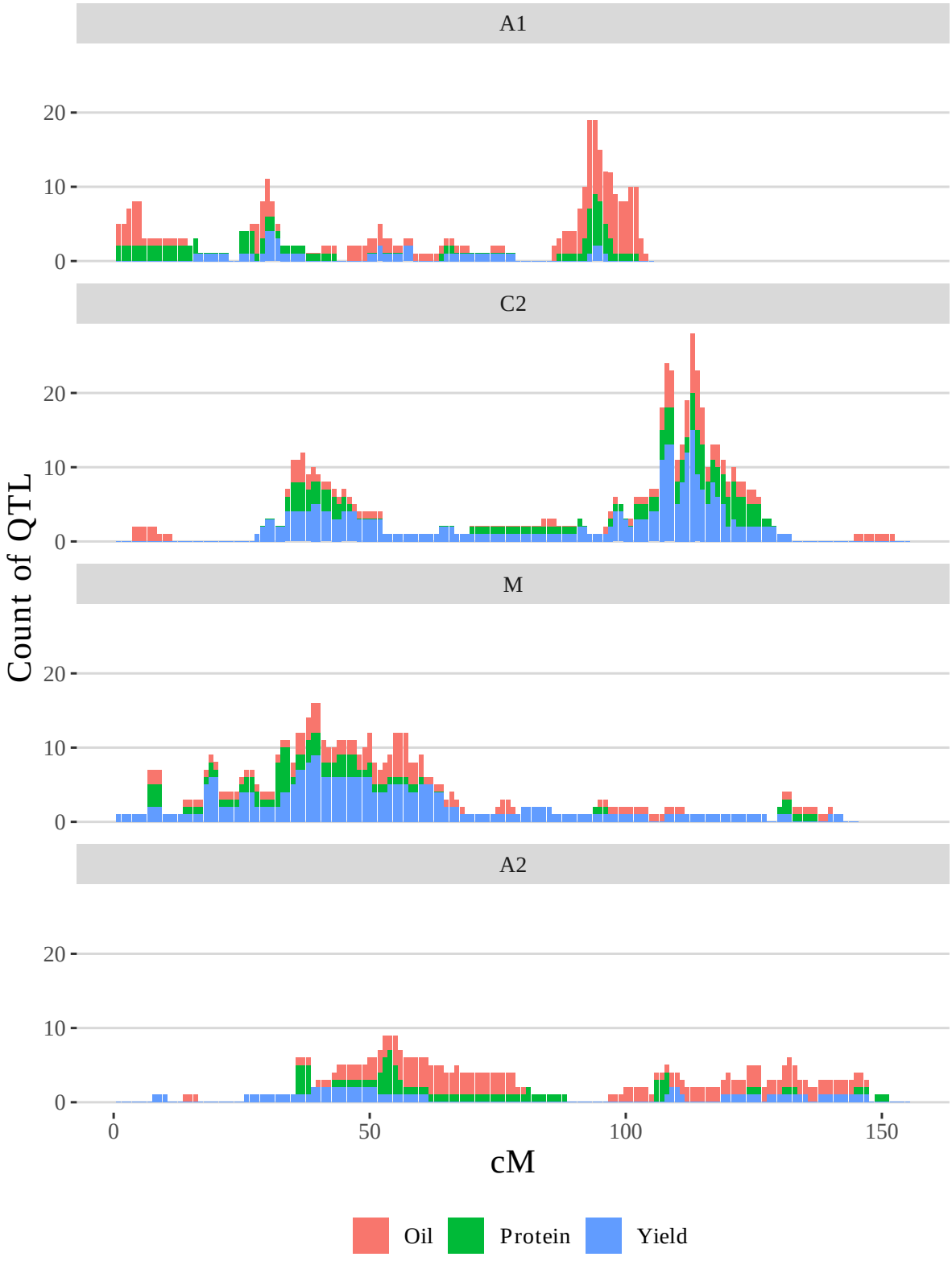

Figure A.42 Count of Previously Discovered Soybean QTL for Oil, Protein, and Yield Traits for Chromosomes 5-8 of the Soybean GmConsensus 4.0 Map.

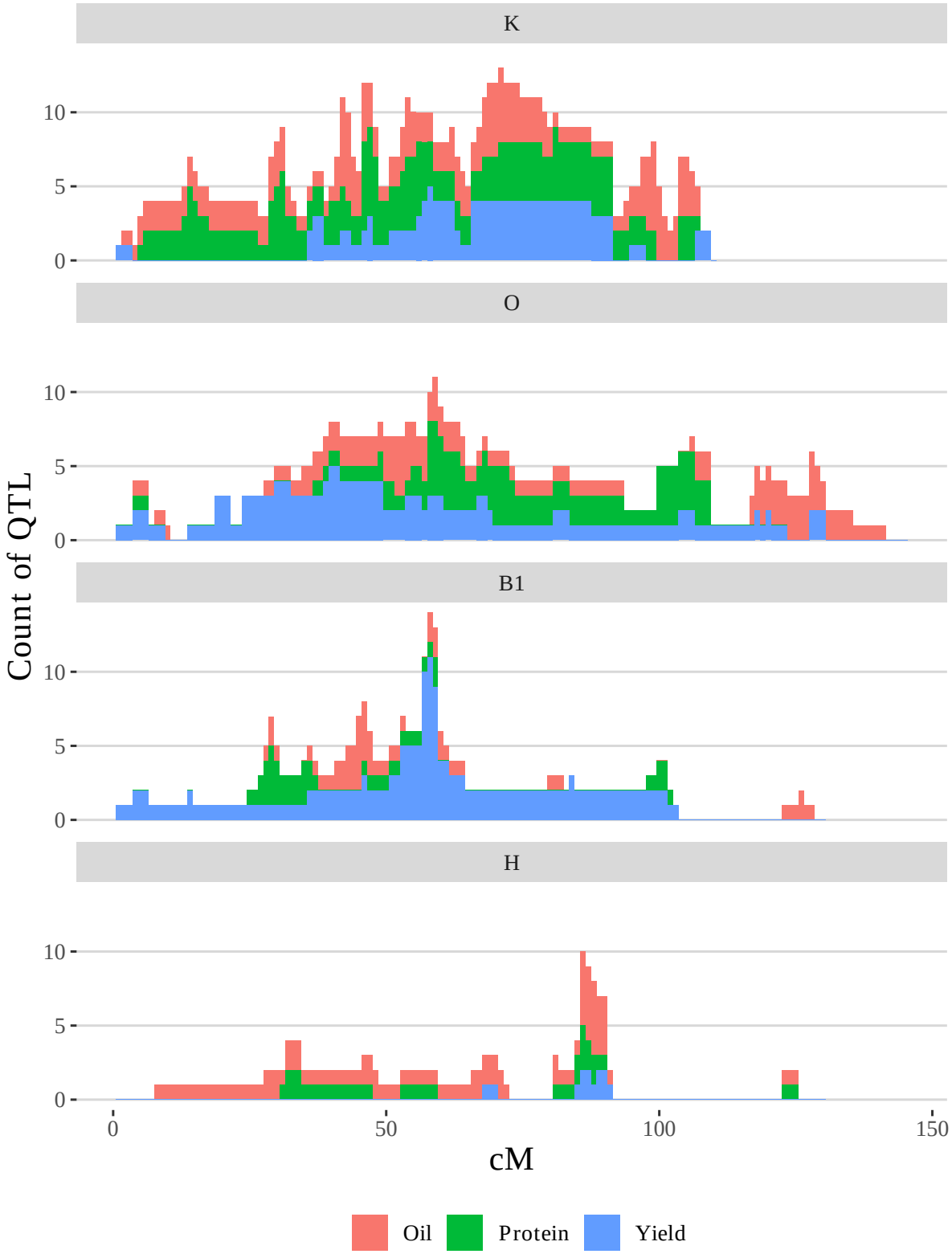

Figure A.43 Count of Previously Discovered Soybean QTL for Oil, Protein, and Yield Traits for Chromosomes 9-12 of the Soybean GmConsensus 4.0 Map.

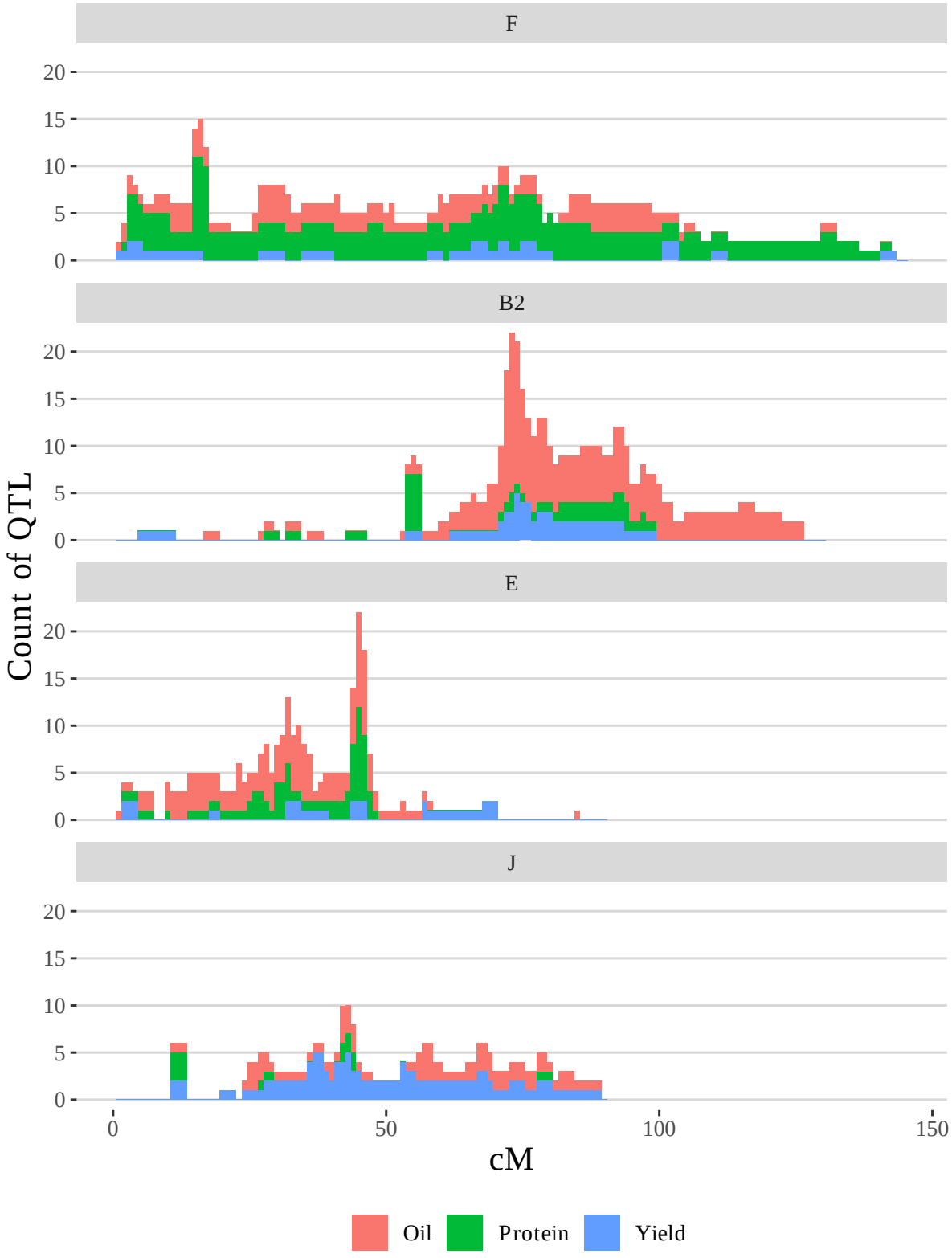

Figure A.44 Count of Previously Discovered Soybean QTL for Oil, Protein, and Yield Traits for Chromosomes 13-16 of the Soybean GmConsensus 4.0 Map.

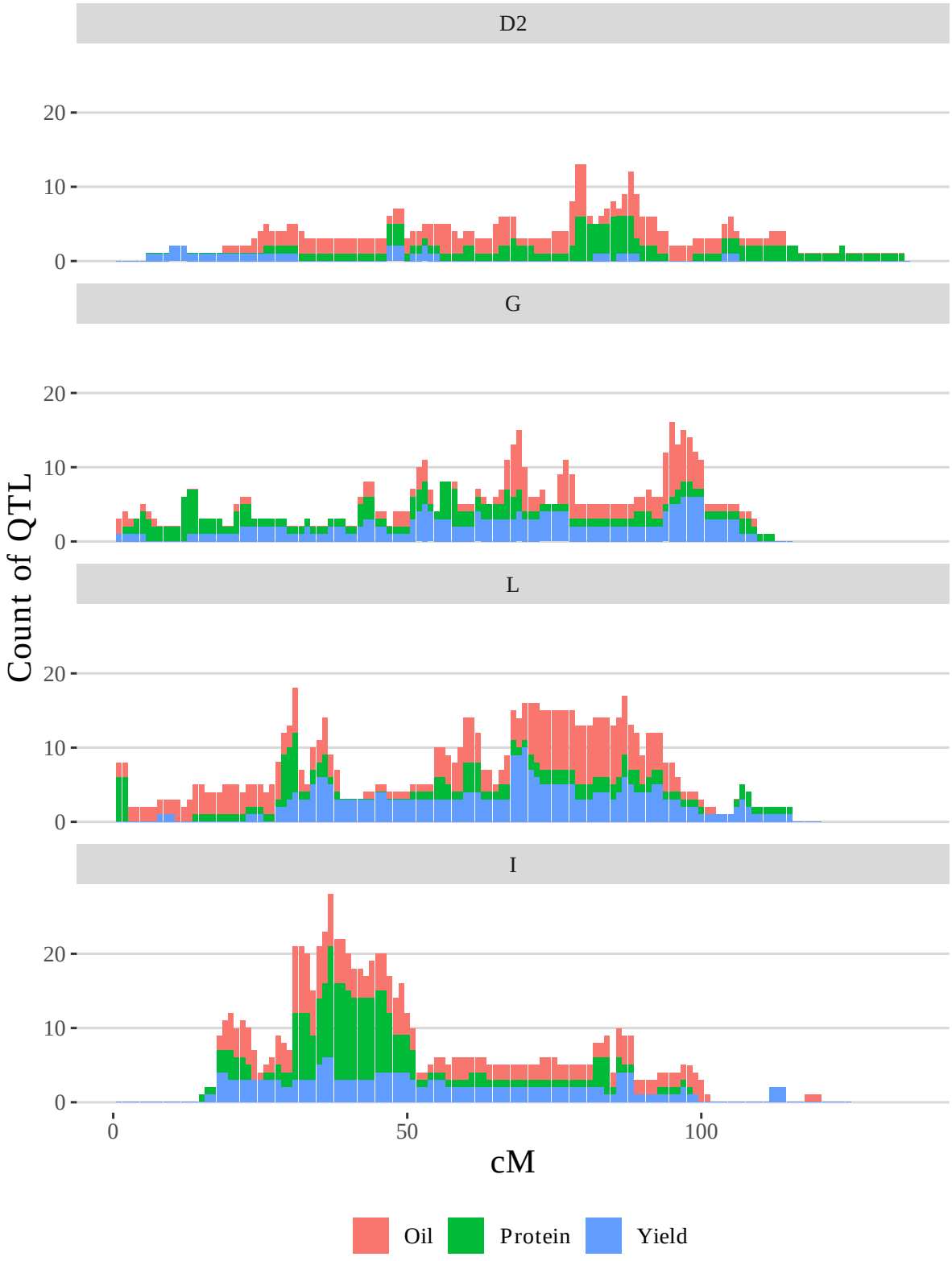

Figure A.45 Count of Previously Discovered Soybean QTL for Oil, Protein, and Yield Traits for Chromosomes 17-20 of the Soybean GmConsensus 4.0 Map.

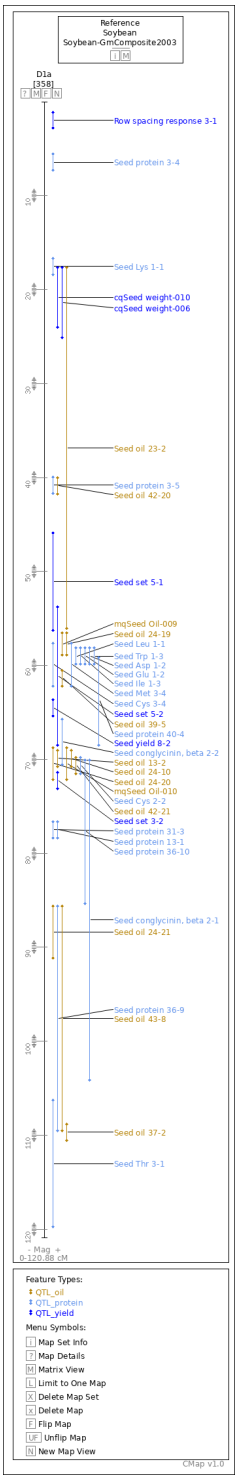

Figure A.46 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group D1a (chromosome 1). Downloaded from Soybase on 2020-06-04.

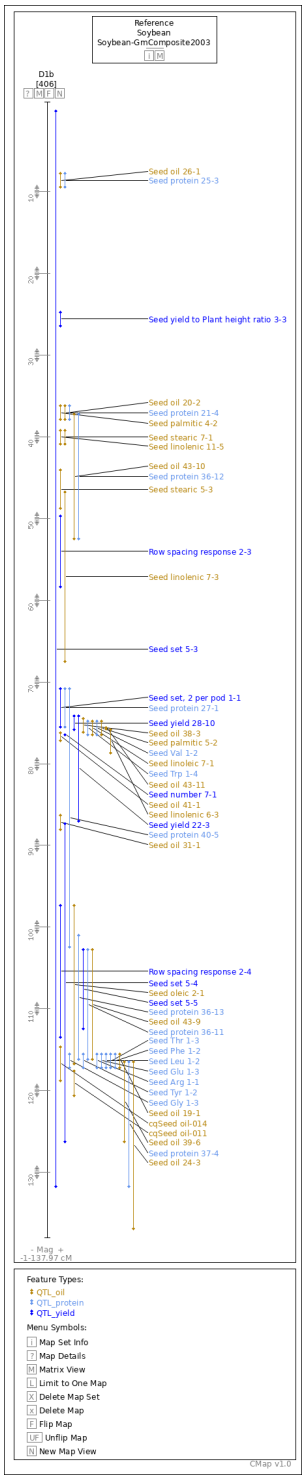

Figure A.47 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group D1b (chromosome 2). Downloaded from Soybase on 2020-06-04.

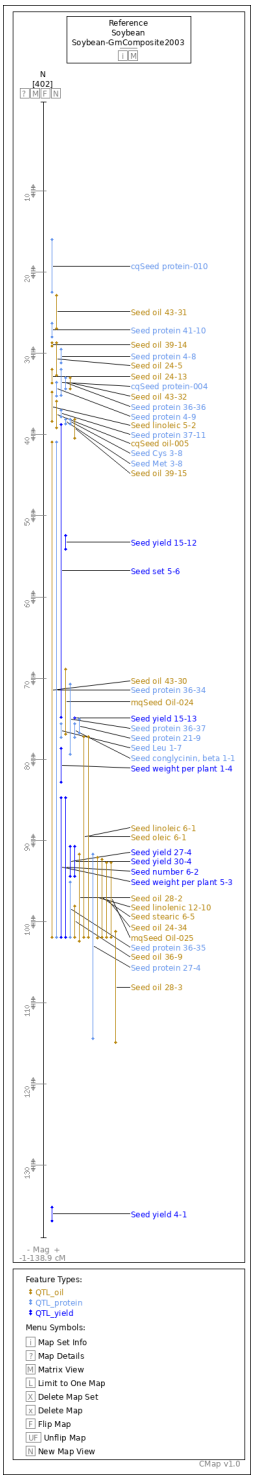

Figure A.48 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group N (chromosome 3). Downloaded from Soybase on 2020-06-04.

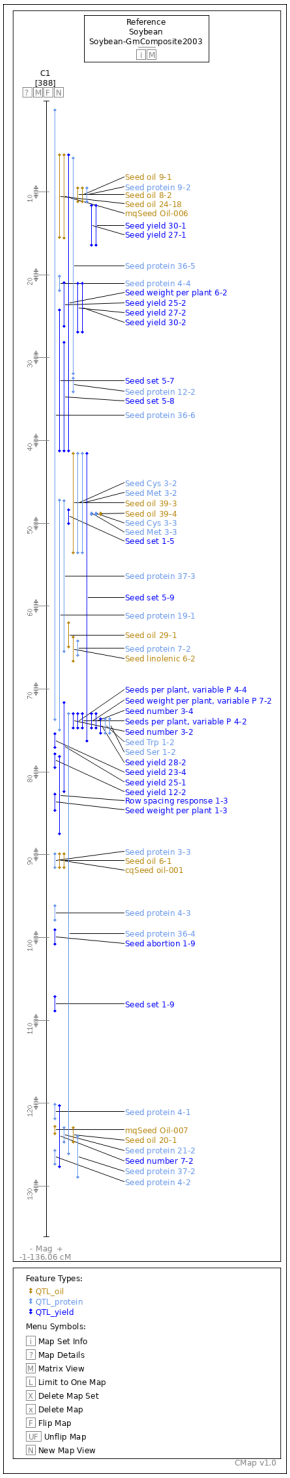

Figure A.49 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group C1 (chromosome 4). Downloaded from Soybase on 2020-06-04.

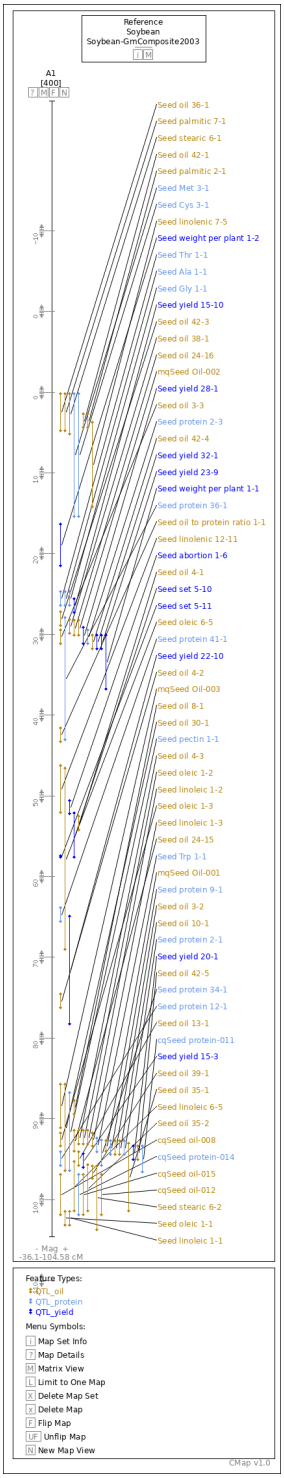

Figure A.50 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group A1 (chromosome 5). Downloaded from Soybase on 2020-06-04.

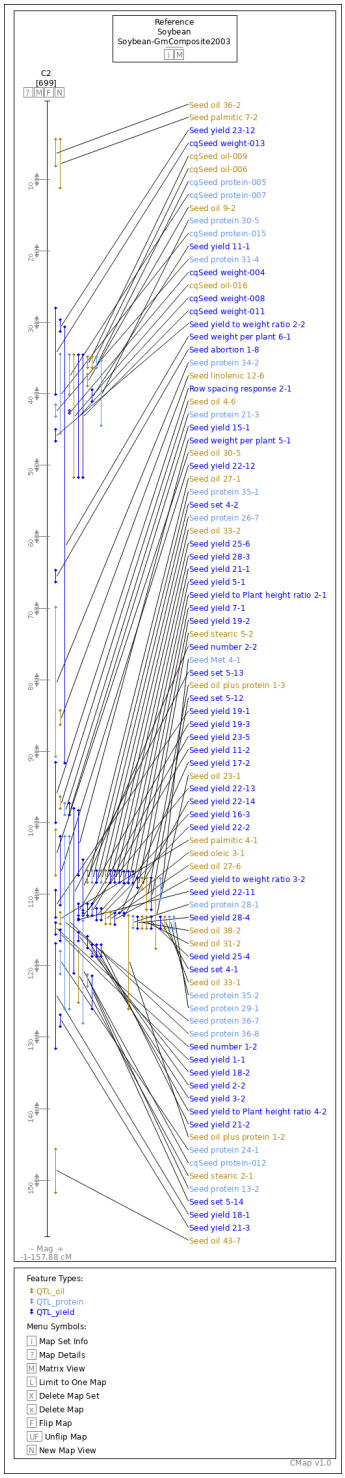

Figure A.51 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group C2 (chromosome 6). Downloaded from Soybase on 2020-06-04.

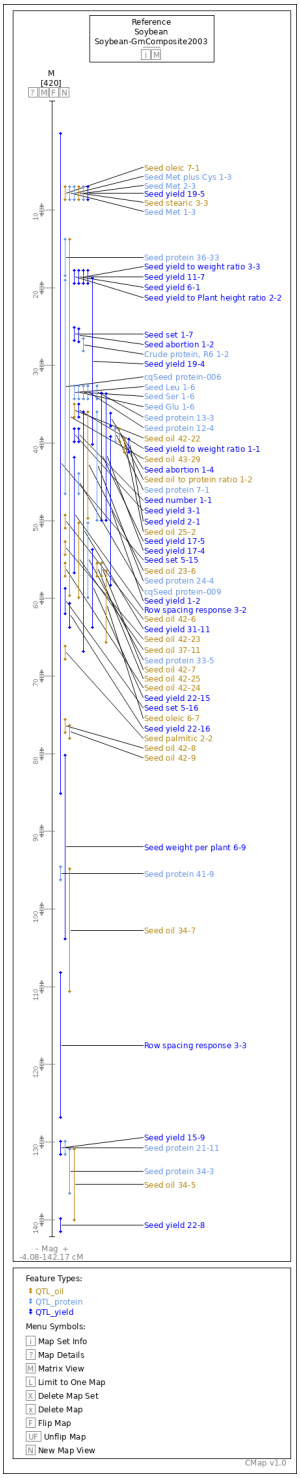

Figure A.52 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group M (chromosome 7). Downloaded from Soybase on 2020-06-04.

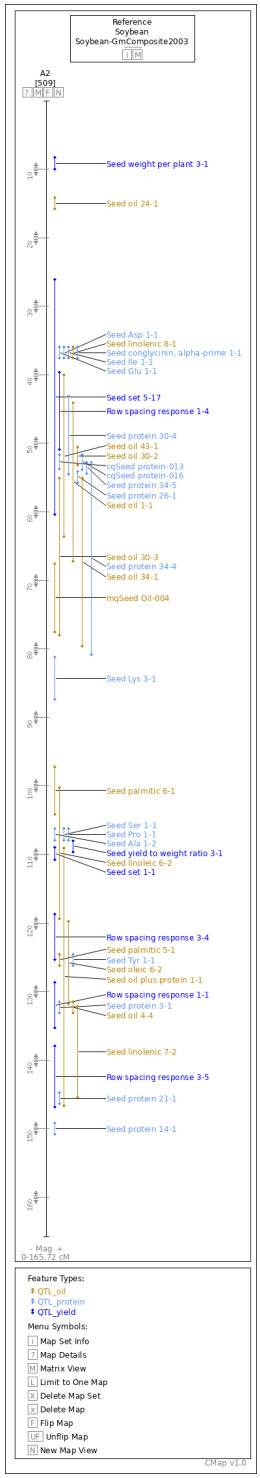

Figure A.53 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group A2 (chromosome 8). Downloaded from Soybase on 2020-06-04.

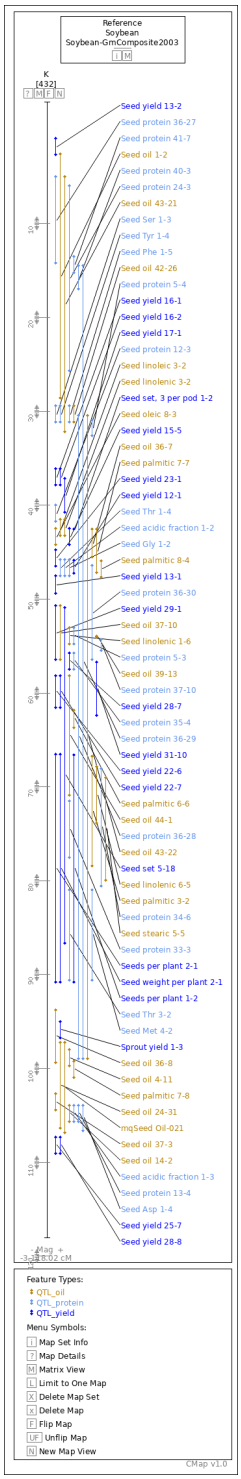

Figure A.54 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group K (chromosome 9). Downloaded from Soybase on 2020-06-04.

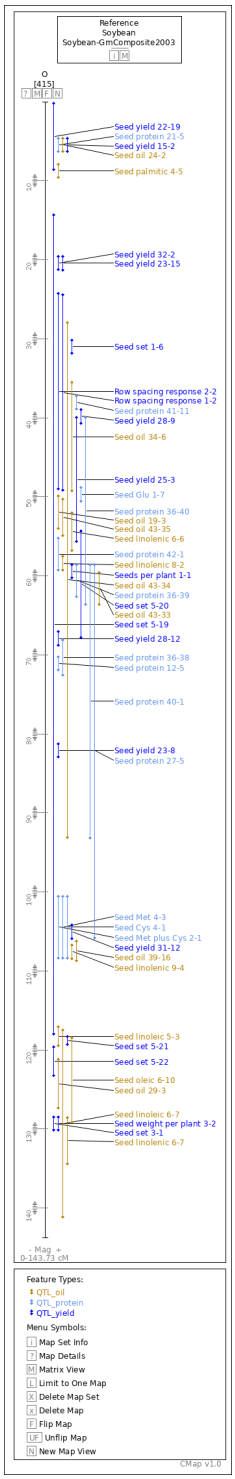

Figure A.55 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group O (chromosome 10). Downloaded from Soybase on 2020-06-04.

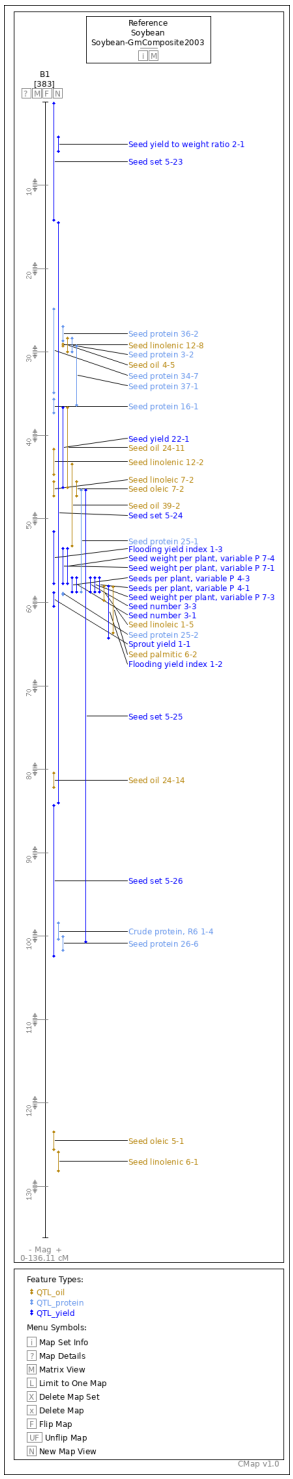

Figure A.56 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group B1 (chromosome 11). Downloaded from Soybase on 2020-06-04.

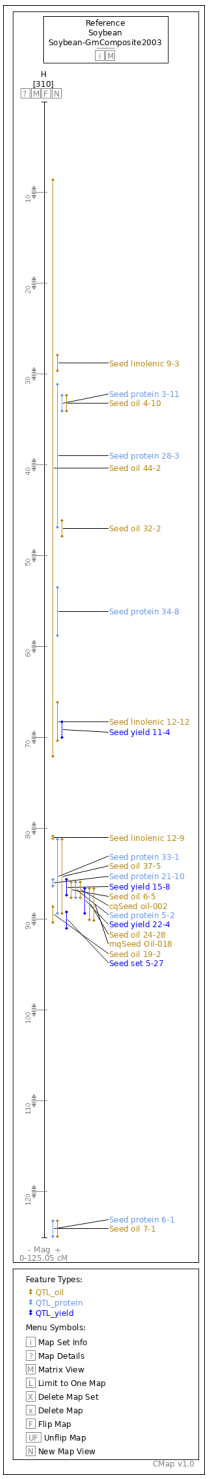

Figure A.57 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group H (chromosome 12). Downloaded from Soybase on 2020-06-04.

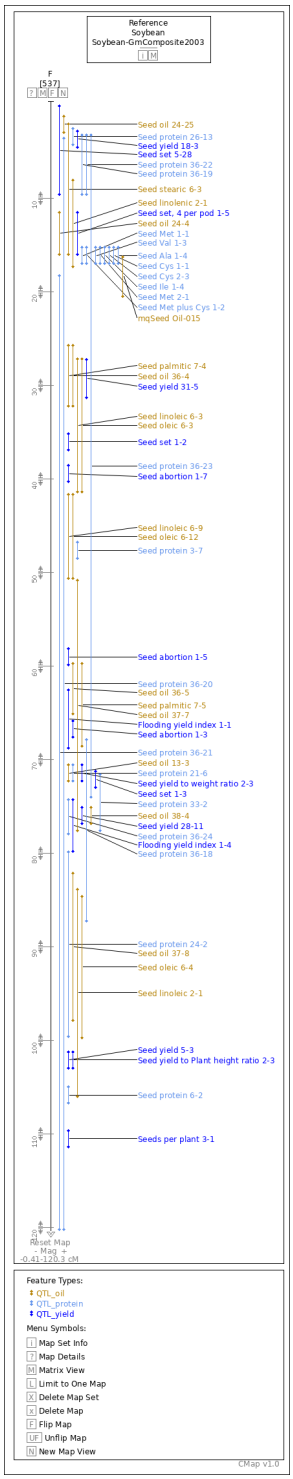

Figure A.58 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group F (chromosome 13). Downloaded from Soybase on 2020-06-04.

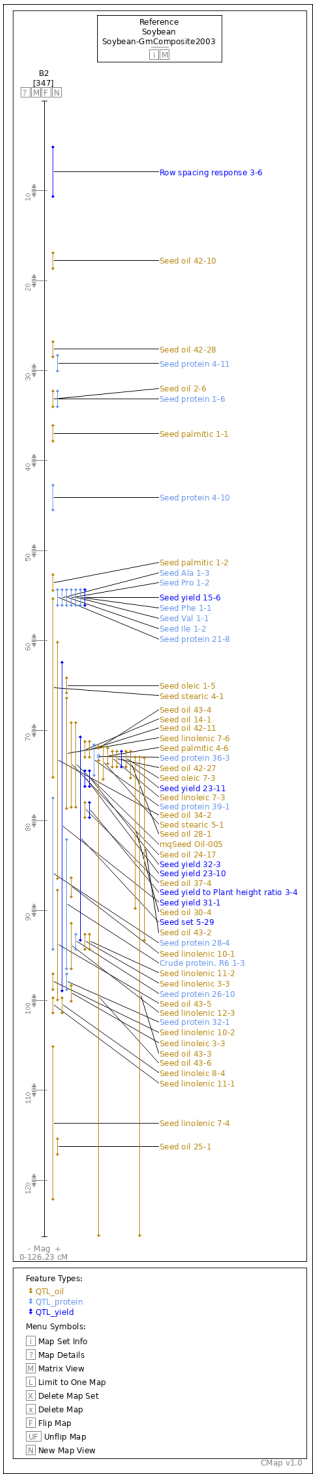

Figure A.59 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group B2 (chromosome 14). Downloaded from Soybase on 2020-06-04.

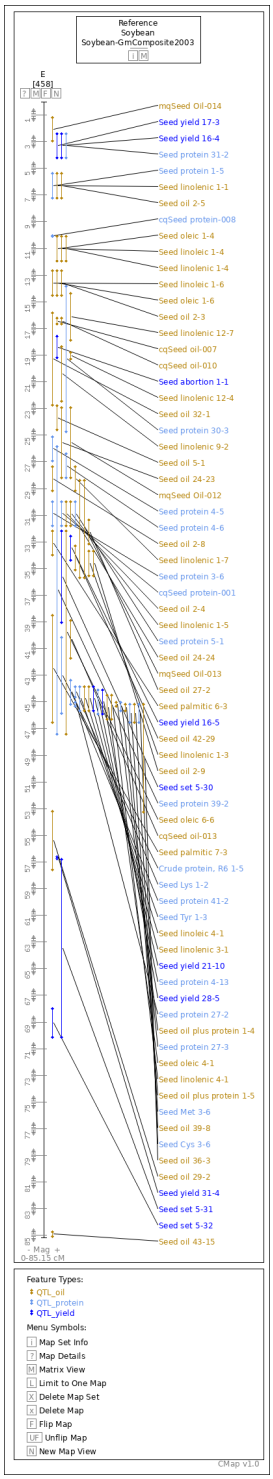

Figure A.60 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group E (chromosome 15). Downloaded from Soybase on 2020-06-04.

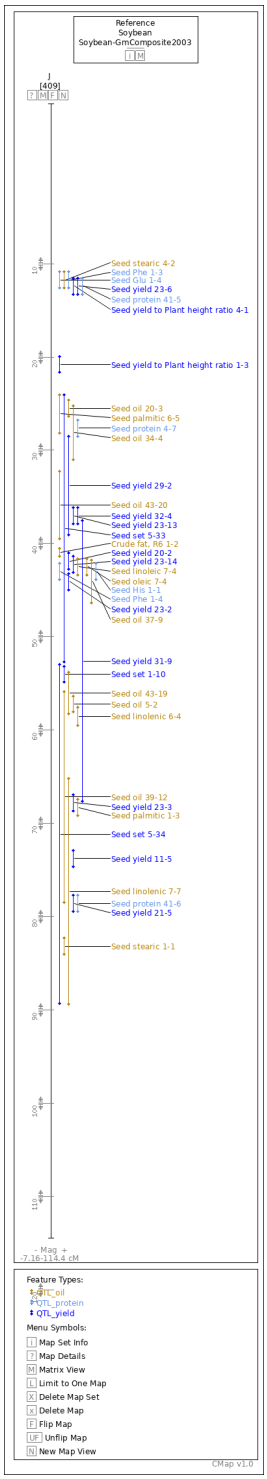

Figure A.61 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group J (chromosome 16). Downloaded from Soybase on 2020-06-04.

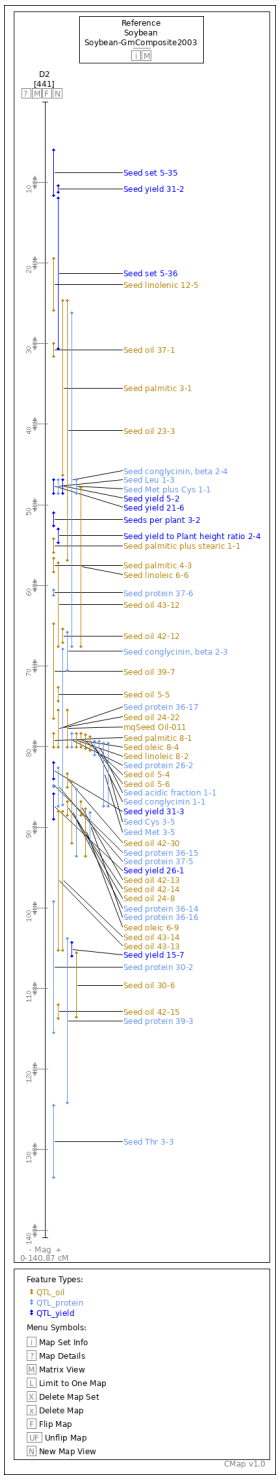

Figure A.62 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group A1 (chromosome 17). Downloaded from Soybase on 2020-06-04.

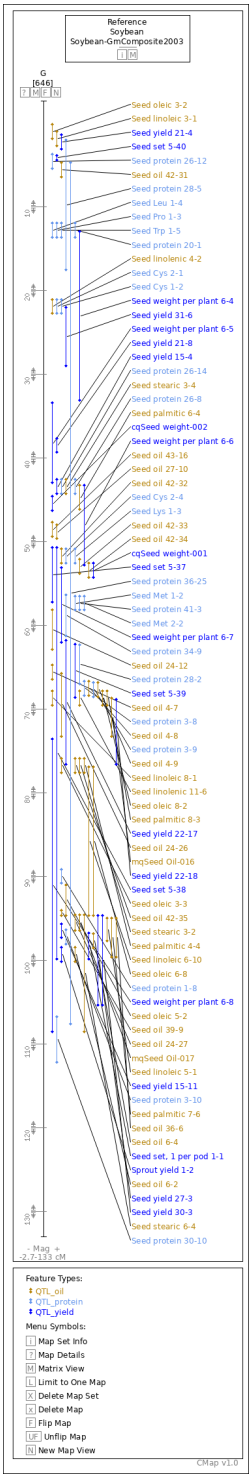

Figure A.63 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group G (chromosome 18). Downloaded from Soybase on 2020-06-04.

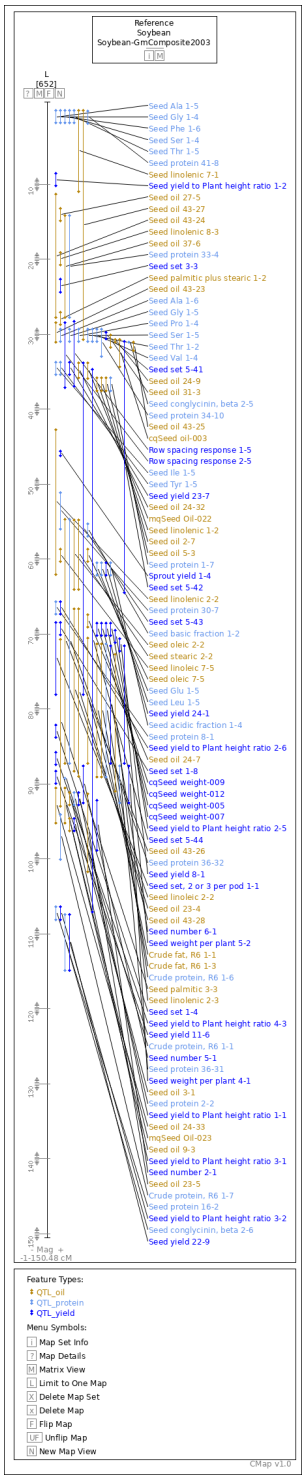

Figure A.64 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group L (chromosome 19). Downloaded from Soybase on 2020-06-04.

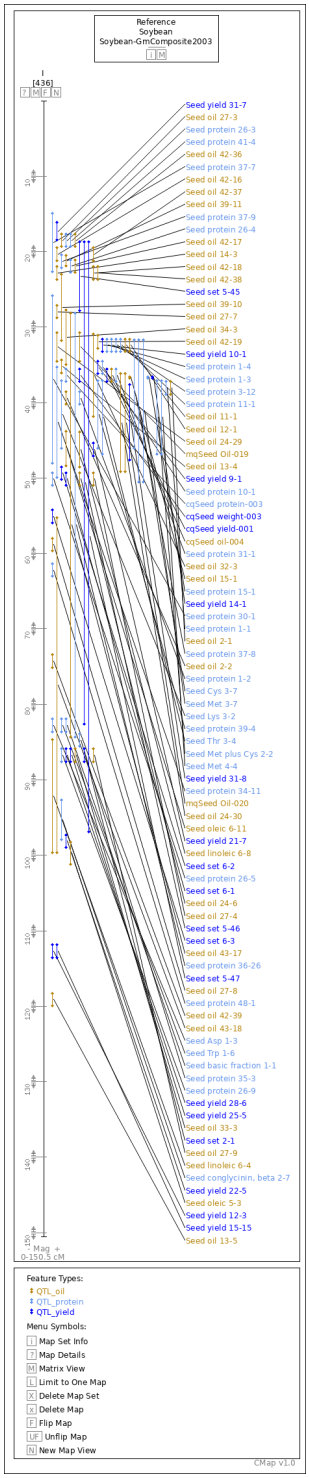

Figure A.65 Positions of Soybean QTL for Protein, Oil, and Yield traits on Linkage group I (chromosome 20). Downloaded from Soybase on 2020-06-04.