Quantitative trait loci analysis of a RIL soybean population to determine chromosomal regions governing seed protein, oil, and linolenic acid content

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Ronald Edward Radish Moore
August 2020

Copyright © 2020 by Ronald Edward Radish Moore
All rights reserved.

ACKNOWLEDGEMENTS

Special thanks to my advisor, Dr. Vince Pantalone, along with members of my committee for continued support of my project. I would also like to thank members of the Bean Team, including staff and undergraduates for their help and friendship. This work could not have been completed without the support from staff at East Tennessee Research and Education Center, Highland Rim Research and Education Center, and Research Education Center at Milan. Funding from the Tennessee Soybean Promotion Board and from United Soybean Board project# 2020-152-0113 made this research possible.

ABSTRACT

180 $F_{4:6}$ [fourth filial generation advanced to sixth filial generation] recombinant inbred lines (RILs) segregating for protein, oil, and fatty acids were produced from a cross between TN12-4098 and TN13-4303. These lines were grown across three locations spread horizontally across Tennessee at: Research Education Center at Milan (RECM), Highland Rim Research and Education Center (HRREC), and East Tennessee Research and Education Center (ETREC) in 2018 and 2019. 21 quantitative trait loci (QTL) spanning 7 chromosomes were found using WinQTLCart2.5 for traits, including days after planting (DAP), height, lodging, yield, protein, oil, linolenic acid, and meal protein.

TABLE OF CONTENTS

Introduction 1
Chapter One Literature Review 2
Abstract 3
Protein 3
Oil 5
Linolenic Acid 6
Objectives 9
References 10
Chapter Two Agronomic and Seed Quality Traits 15
Abstract 16
Introduction 16
Population Structure and Field Layout 18
Phenotyping 19
Field Notes 19
NIR Analysis (Protein and Oil) 19
Fatty Acid Analysis 20
Statistical Analysis 21
SAS 9.4 21
Microsoft Excel 22
Results and Discussion 22
References 25
Appendix A 29
Chapter Three QTL Analysis 44
Abstract 45
Introduction 45
Materials and Methods 46
Genotyping 46
GenomeStudio 2.0 47
Linkage Mapping 48
Microsoft Excel 49
WinQTLCart2.5 49
Results and Discussion 50
References 55
Appendix B 58
Conclusion 98
Vita 100

LIST OF TABLES

Table 2.1 History of RIL population development 30

Table 2.2 RIL agronomic traits separated by yield 31

Table 2.3 Top 10% of RILs for protein content 36

Table 2.4 Top 10% of RILs for oil content 37

Table 2.5 Top 10% of RILs for linolenic acid content 38

Table 2.6 Top 10% of RILs for meal protein 39

Table 2.7 Top 10% of RILs for yield 40

Table 2.8 Correlation between major traits of interest 43

Table 3.1 List of QTLs along with respective traits and attributes 97

LIST OF FIGURES

Figure 2.1 Oil frequency distribution..... 41

Figure 2.2 Protein frequency distribution 41

Figure 2.3 Meal protein frequency distribution 42

Figure 2.4 Linolenic acid frequency distribution..... 42

Figure 3.1 Incorrect GenomeStudio 2.0 SNP call..... 59

Figure 3.2 Correct GenomeStudio 2.0 SNP call 60

Figure 3.3 Raw chromosome 18 heatmap..... 61

Figure 3.4 Revised chromosome 18 heatmap 62

Figure 3.5 Linkage map of chromosome 1 63

Figure 3.6 Linkage map of chromosome 2 64

Figure 3.7 Linkage map of chromosome 3 65

Figure 3.8 Linkage map of chromosome 4 66

Figure 3.9 Linkage map of chromosome 5 67

Figure 3.10 Linkage map of chromosome 6 68

Figure 3.11 Linkage map of chromosome 7 69

Figure 3.12 Linkage map of chromosome 8 70

Figure 3.13 Linkage map of chromosome 9 71

Figure 3.14 Linkage map of chromosome 10 72

Figure 3.15 Linkage map of chromosome 11 73

Figure 3.16 Linkage map of chromosome 12 74

Figure 3.17 Linkage map of chromosome 13 75

Figure 3.18 Linkage map of chromosome 14 76

Figure 3.19 Linkage map of chromosome 15 77

Figure 3.20 Linkage map of chromosome 16 78

Figure 3.21 Linkage map of chromosome 17 79

Figure 3.22 Linkage map of chromosome 18 80

Figure 3.23 Linkage map of chromosome 19 81

Figure 3.24 Linkage map of chromosome 20 82

Figure 3.25 Two protein QTL located on chromosome 1 83

Figure 3.26 One oil QTL located on chromosome 1 84

Figure 3.27 Two protein QTL located on chromosome 3 84

Figure 3.28 One oil QTL located on chromosome 3 85

Figure 3.29 Three meal protein QTL located on chromosome 3..... 85

Figure 3.30 One linolenic acid QTL located on chromosome 5..... 86

Figure 3.31 One protein QTL located on chromosome 7 86

Figure 3.32 One yield QTL located on chromosome 14 87

Figure 3.33 One protein QTL located on chromosome 14 87

Figure 3.34 One oil QTL located on chromosome 14 88

Figure 3.35 One linolenic acid QTL located on chromosome 14..... 88

Figure 3.36 One meal protein QTL located on chromosome 14 89

Figure 3.37 One linolenic acid QTL located on chromosome 18..... 89

Figure 3.38 One DAP QTL located on chromosome 19 90

Figure 3.39 Two lodging QTL located on chromosome 19..... 90

Figure 3.40 One height QTL located on chromosome 19 91

Figure 3.41 Chromosome 1 linkage map with highlighted QTLs 92

Figure 3.42 Chromosome 3 Linkage map with highlighted QTLs..... 93

Figure 3.43 Chromosome 5 Linkage map with highlighted QTLs..... 94

Figure 3.44 Chromosome 7 Linkage map with highlighted QTLs..... 94

Figure 3.45 Chromosome 14 Linkage map with highlighted QTLs..... 95

Figure 3.46 Chromosome 18 Linkage map with highlighted QTLs..... 96

Figure 3.47 Chromosome 19 Linkage map with highlighted QTLs..... 97

INTRODUCTION

Soybean [Glycine max (L.) Merrill] is an annual legume cultivated for its seed products, specifically protein and oil. Soybean has a high seed protein content, and is the top cultivated crop used for vegetable oil (Qiu and Chang, 2010). In 2018 alone, Tennessee farmers planted 687,966 hectares of soybeans (USDA, 2019). Soybean protein is in high demand, as there is a global shortage in animal feed protein to meet production levels (Kim et al., 2019). Markets for soybean protein include tofu, edamame, and animal feed, with animal feed the dominant market. Protein derived from soybean seed is high quality and captures attention from markets for being plant-based. Partially hydrogenated oils are currently an issue in the US after it was determined they are not "generally regarded as safe" (Wayland, 2015). Creation of soybeans with high oleic acid (>80%) and low-linolenic acid (<3%) allow soybean oil to be trans-fat free. Finding QTL responsible for the variation in protein, oil and fatty acids would be beneficial to public soybean breeders, as they can incorporate these genomic regions into elite soybean lines.

CHAPTER ONE

LITERATURE REVIEW

Abstract

Seed protein, oil, linolenic acid, and meal protein are traits of soybeans that breeders are modifying. These are quantitative traits, controlled by multiple genes. To improve these traits, plant breeders can utilize quantitative trait loci (QTLs). QTLs are regions of a chromosome controlling the variation in a quantitative trait. Although QTLs already exist for these traits, an increased number of novel QTLs can help a breeder by providing new options for QTL selection. Using marker assisted selection (MAS), breeders can incorporate QTLs into their soybeans to create lines with improved profiles.

Protein

Soybean protein is an integral component of human and animal diets. In 2018, 44.5 million metric tons of soybean meal were produced by the United States, and of that, 35.1 million metric tons went directly to livestock production (soystats.com). With a projected exponential growth in human population over time, it is paramount that soybean protein content meets the needs of the livestock that we consume. According to a study conducted by Yaklich et al. (2002), soybean seed protein content averaged from 40.4% to 41.4% across maturity groups over a span of 51 years. Recently, it is estimated that soybean protein content is closer to 39.6% on a dry basis (Brzostowski et al., 2017). The problem with seed protein content is that it is generally negatively proportional to seed oil and seed yield (Burton et al., 1987). The yield discourages farmers from using high protein lines, because soybeans are sold on a basis of weight rather than quality (Yaklich et al. 2001). Although increased seed protein is generally negatively correlated

with yield, there are some cases where breeders have been able to develop high yielding, high protein lines. For example, in a 1995 study, Wilcox and Cavins found that protein content can be increased without sacrificing yield when high protein lines are backcrossed into high yielding lines. Pantalone and Smallwood (2018) describe the development of the cultivar TN11-5102 with high yields and 49% meal protein. Furthermore, Pantalone et al. (2020) describe the new cultivar TN15-5007 with high yields and 50.5% meal protein.

Seed protein is a quantitative trait, which means that many genes with small and large effects govern the trait. Molecular strategies for improvement can be helpful. QTL studies allow researchers to target and use genomic regions that account for some of the variation in seed protein. Several papers have successfully identified protein QTL across multiple chromosomes. For example, Zhang et al. (2015) identified 9 protein QTL, 5 of which were additive QTL, in a population consisting of 147 F_6 recombinant inbred lines. Three of these were considered major additive QTLs. These QTLs spanned across 7 chromosomes. In a 2004 study by Hyten et al., 4 protein QTL were found in a population of 131 F_6 RILs that spanned 4 chromosomes and explained up to 27.6% of the phenotypic variation averaged over multiple environments. These QTL were located on chromosome 6,7,9, and 13. Pro-1, the QTL which explained 27.6% of variation combined, was located on chromosome 6 at 119.8 cM however, the effect was likely due to the E_1 maturity gene at this position. Panthee et al. (2005) found a protein QTL associated with marker Satt570, located on chromosome 18 that explained 20.2% of phenotypic variation in a population of 101 F_6-derived RILs. The QTL near Satt570 was stable across

environments and was found to be related to seed nitrogen accumulation (Panthee et al., 2005). Soybean seed protein is often influenced by environmental factors (Cunicelli et al., 2019), so it is critical to validate QTL in geographically different areas.

Oil

Soybean oil is a useful product for industrial and food applications. 11.1 million metric tons of soybean oil were produced by the United States in 2018, and out of that, 10.3 million metric tons went to United States vegetable oil consumption (soystats.com). The average return on each metric ton came out to 661 USD. With approximately 7.34 billion USD stemming from United States soybean oil production alone, it is worth the effort to breed for increased seed oil. Seed oil content averaged across all maturity groups is between 19.8% and 21.2% (Yaklich et al. 2002). In a study by Li et al. (2018), it was noted that in a soybean population, a negative correlation (-0.66, $P < 0.01$) existed between seed oil and seed protein.

There have been many oil QTL discovered. Pantalone et al. (2004) stated that, at the time, 53 oil QTL were reported. Currently, Soybase reports there are 322 bi-parental QTL associated with seed oil. Chapman et al. (2003) discovered 2 oil QTL in a population of 208 F_2 plants and 177 $F_{4:6}$ lines. The first QTL was additive ($r^2 = 0.05$) and located near Satt14 on chromosome 17. The second QTL was additive ($r^2 = 0.04$ oil and 0.03 protein) and located near Satt251 on chromosome 11. QTLs such as the one linked to Satt251 are desirable because they offer the ability to increase protein and oil simultaneously, in an otherwise negative relationship. Although many oil QTLs have been discovered, oil QTLs need confirmation. Oil QTL are environmentally sensitive,

with oil biosynthesis being affected by factors including temperature, rain, etc. (Pantalone et al., 2004). For example, an oil QTL discovered in a southern US population may not be detected in the same population grown in the northern US. Pantalone et al. (2004) states that confirmed oil QTL would be useful to breeders in targeted geographical areas. Finding more oil QTLs will increase options available to breeders wanting to create superior soybean lines.

Linolenic Acid

On average, soybean oil contains 8% linolenic acid (Hoshino et al., 2014). Linolenic acid is a primary factor in the instability and oxidative properties of soybean oil (Warner and Fehr, 2008). Because of these properties, foods produced with unmodified soybean oil will have a short shelf life. Currently, the industry standard for low linolenic acid soybean oil is < 3% (Pham et al., 2012; Smallwood et al., 2017). With genetic methods, breeders have accomplished producing low-linolenic soybean lines (Pham et al., 2012, Hoshino et al., 2014, Bilyeu et al., 2005). Confirming QTL responsible for linolenic acid, or the lack thereof, would allow breeders to identify lines containing genes responsible for low linolenic acid.

Three genes are identified in soybean that control linolenic acid levels: FAD3A (Glyma.14g194300), FAD3B (Glyma.02g227200), and FAD3C (Glyma.18g06200) (Bilyeu et al., 2003, Held et al., 2019). In a 2005 study by Bilyeu et al., 107 F_2 progeny resulting from a cross between W82 and a mutant FAD3A / FAD3C donor (2721) showed that mutated alleles of the FAD3 genes significantly reduced seed linolenic acid. When FAD3A and FAD3C were mutated (denoted *aacc*), linolenic acid dropped over

66% compared to the wild type (*AACC*) (Bilyeu et al., 2005). When comparing FAD3A and FAD3C, mutations in FAD3A resulted in larger reductions in linolenic acid than the latter (Bilyeu et al., 2005). Furthermore, F_2 progeny were advanced to the F_4 generation and the fatty acid profile remained, suggesting that F_2 screening for FAD3A and FAD3C mutations is a reliable method for identifying low linolenic acid progeny. Mutant alleles of FAD3B can be incorporated along with mutant FAD3A and FAD3C alleles to produce ~1% linolenic acid in progeny (Bilyeu et al., 2011). Bilyeu et al. found that in F_2 progeny that contained triple homozygous FAD3 alleles (*aabbcc*), mean linolenic acid percentage fell below 1.5% (Bilyeu et al., 2011).

Hyten et al. (2004) identified 3 QTL in a population of 131 $F_{6:8}$ RILs spanning across chromosomes 13 (11.4 cM) and 19 (50.6 cM, 82.5 cM). The QTL located on chromosome 19 (82.5 cM) explained 24.8% of the phenotypic variation in linolenic acid. Panthee et al. (2006) found two QTL associated with linolenic acid in a population of 101 F_6 RILs. The first QTL was found on chromosome 15 near marker Satt263. This QTL explained 12.3% of variation in the RILs. The second QTL was located on chromosome 18 near marker Satt235. This QTL explained 22.5% of the variation in linolenic acid in the RILs.

In a study published in 2017, Smallwood et al. discovered multiple QTL explaining a combined 19% of the variation for linolenic acid in an F_5 derived RIL population. These QTL were discovered after compiling 3 years of data from various environments. 5 QTL were discovered, which were located on chromosomes 9, 13, 17, and 19 (Smallwood et al., 2017). QTLs were named Len9.1, Len 9.2, Len13, Len17, and Len19. These QTLs

explained 4%, 6%, 6%, 1% and 2% of variation in linolenic acid, respectively (Smallwood et al., 2017). These QTL were not associated with FAD3A, FAD3B, or FAD3C (Smallwood et al., 2017). QTL such as these point to the importance of identifying and utilizing modifier QTLs for linolenic acid manipulation.

With the industry standard for low linolenic soybeans being set at < 3%, breeders can identify and use major and minor QTLs to efficiently meet this objective.

Objectives

1. Determine various traits of the RILs including seed yield, plant height, lodging, and relative maturity.
2. Use near infrared reflectance to analyze seed protein and oil content.
3. Use gas chromatography to analyze seed fatty acid content.
4. Extract genomic DNA from RILs to genotype the population.
5. Detect quantitative trait loci (QTL) influencing seed protein, oil and linolenic acid content in a soybean population.

REFERENCES

Bilyeu, K. D., Palavalli, L., Sleper, D. A., & Beuselinck, P. R. (2003) Three Microsomal Omega-3 Fatty-acid Desaturase Genes Contribute to Soybean Linolenic Acid Levels. *Crop Science*, **43**:1833-1838.

Bilyeu, K., Palavalli, L., Sleper, D., & Beuselinck, P. (2005) Mutations in Soybean Microsomal Omega-3 Fatty Acid Desaturase Genes Reduce Linolenic Acid Concentration in Soybean Seeds. *Crop Science*, **45**: 1830-1836.

Bilyeu, K., Gillman, J. D., & LeRoy, A. R. (2011) Novel FAD3 Mutant Allele Combinations Produce Soybeans Containing 1% Linolenic Acid in the Seed Oil. *Crop Science*, **51**: 259-264.

Brzostowski, L. F., Pruski, T.I., Specht, J.E., & Diers, B. W. (2017) Impact of seed protein alleles from three soybean sources on seed composition and agronomic traits. *Theoretical and Applied Genetics*, **130**: 2315–2326.

Burton, J. (1987) Quantitative genetics: Results relevant to soybean breeding. 111 IR. Wilcox (ed) Soybeans: Improvement, Production, and Uses. *American Society of Agronomy*.

Chapman, A., Pantalone, V., Ustun, A., Allen, F., Landau-Ellis, D., Trigiano, R., & Gresshoff. P. (2003) Quantitative trait loci for agronomic and seed quality traits in an F2 and F4:6 soybean population. *Euphytica*, **129**: 387–393.

Cunicelli, M.J., Bhandari, H.S., Chen, P., Sams, C.E., Mian, M.A.R., Mozzoni, L.A., …Pantalone, V.R. (2019) Effect of a Mutant Danbaekkong Allele on Soybean Seed Yield, Protein, and Oil Concentration. *Journal of the American Oil Chemists Society*, **96**: 927-935.

Held, J. P., Carrero-Colón, M., & Hudson, K. A. (2019) Combination of Novel Mutation in FAD3C and FAD3A for Low Linolenic Acid Soybean. *Agrosystems, Geosciences & Environment*, **2:**1.

Hoshino, T., Watanabe, S., Takagi, Y., & Anai, T., (2014) A novel GmFAD3-2a mutant allele developed through TILLING reduces α-linolenic acid content in soybean seed oil. *Breeding Science*, **64**: 371–377.

Hyten, D.L., Pantalone, V. R., Sams, C.E., Saxton, A.M., Landau-Ellis, D., Stefaniak, T.R., & Schmidt, M.E. (2004) Seed quality QTL in a prominent soybean population. *Theoretical and Applied Genetics*, **109**: 552–561.

Kim, S.W., Less, J.F., Wang, L., Yan, T., Kiron, V., Kaushik, S.J., & Lei, X.G. (2019) Meeting Global Feed Protein Demand: Challenge, Opportunity, and Strategy. *Annual Review of Animal Biosciences*, **7**: 221–243.

Li, Y.H., Reif, J.C., Hong, H.L., Li, H.H., Liu, Z.X., Ma, Y.S., …Qiu, L.J. (2018) Genome-wide association mapping of QTL underlying seed oil and protein contents of a diverse panel of soybean accessions. *Plant Science*, **266**: 95–101.

Pantalone, V., Cunicelli, M., & Wyman, C. (2020) Registration of soybean cultivar 'TN15-5007'with high meal protein. *Journal of Plant Registrations*, **14**: 139-143.

Pantalone, V., & Smallwood, C. (2018) Registration of 'TN11-5102'soybean cultivar with high yield and high protein meal. *Journal of Plant Registrations*, **12**: 304-308.

Pantalone, V., Walker, D., Dewey, R., and Rajcan, I. (2004) DNA Marker-Assisted Selection for Improvement of Soybean Oil Concentration and Quality. *Legume Crop Genomics*, 283-311.

Panthee, D.R., Pantalone, V.R., & Saxton, A.M. (2006) Modifier QTL for fatty acid composition in soybean oil. *Euphytica*, **152**: 67–73.

Panthee, D.R., Pantalone, V.R., West, D.R., Saxton, A.M., & Sams, C.E. (2005) Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean. *Crop Science*, **45**: 2015–2022.

Pham, A. J., Shannon, G., & Bilyeu, K. D. (2012) Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. *Theoretical and Applied Genetics*, **125**: 503–515.

Qiu, L., & Chang, R. (2010) The origin and history of soybean. *The soybean: botany, production, and uses***:** 1-23.

Smallwood, C. J., Gillman, J. D., Saxton, A. M., Bhandari, H. S., Wadl, P. A., Fallen, B. D., ... & Pantalone, V. R. (2017) Identifying and exploring significant genomic regions associated with soybean yield, seed fatty acids, protein and oil. *Journal of crop science and biotechnology*, **20**: 243-253.

Soy Stats. (2019) *A Reference Guide to Important Soybean Facts and Figures*. The American Soybean Association. Retrieved from http://soystats.com/

USDA. (2019) *Crop Production 2018 Summary*. National Agricultural Statistics Service. Retrieved from https://www.nass.usda.gov/Publications/Todays_Reports/reports/cropan19.pdf

Warner, K., & Fehr, W. (2008) Mid-Oleic/Ultra Low Linolenic Acid Soybean Oil: A Healthful New Alternative to Hydrogenated Oil for Frying. *Journal of the American Oil Chemists Society* **85**: 945–951.

Wayland, M. M. (2018) Final determination regarding partially hydrogenated oils. *Fed. Regist*, 80, 116.

Wilcox, J.R., and Cavins, J.F. (1995) Backcrossing High Seed Protein to a Soybean Cultivar. *Crop Science*, **35**: 1036–1041.

Yaklich, R.W. (2001) β-Conglycinin and Glycinin in High-Protein Soybean Seeds. *Journal of Agricultural and Food Chemistry*, **49**: 729–735.

Yaklich, R.W., Vinyard, B., Camp, M., & Douglass, S. (2002) Analysis of Seed Protein and Oil from Soybean Northern and Southern Region Uniform Tests. *Crop Science*, **42**:1504-1515.

CHAPTER TWO

AGRONOMIC AND SEED QUALITY TRAITS

Abstract

200 recombinant inbred lines were created from a cross between TN12-4098 and TN13-4303. In 2018, field trials were conducted in Knoxville, TN and Springfield, TN. In 2019, field trials were conducted in Knoxville, TN, Springfield, TN, and Milan, TN. Height, lodging, pubescence, maturity, and flower color were recorded in the growing season, and yield at harvest. Seed samples from harvested plots were then subject to near infrared spectroscopy and gas chromatography. The RILs had an average yield of 3161.7 kg ha^{-1}, dry protein content of 40.1%, dry oil content of 22.2%, linolenic acid content of 4.1% (with 23% of lines falling below the 3% threshold), and meal protein of 46.2%. TN13-4303 (parent) had an average yield of 3426.9 kg ha^{-1}, protein content of 42.3%, oil content of 21.6%, linolenic acid of 7.1%, and meal protein of 49.2%. TN12-4098 (parent) had an average yield of 1756.1 kg ha^{-1}, protein content of 40.1%, oil content of 22.4%, linolenic acid content of 2.7%, and meal protein content of 47.1%. The checks averaged 3537.5 kg ha^{-1} yield, 41.0% protein, 21.9% oil, 7.3% linolenic acid, and 47.9% meal protein.

Introduction

The agronomic qualities of a soybean line and its seed quality characteristics are paramount to its success. Farm operations want to grow a soybean that is dependable and up to industry standards. Acceptable soybeans must be competitive in appearance as well as performance. Breeders will choose to drop lines lacking important characteristics such as a good lodging score and strong yield.

World soybean production is expected to reach 371.3 million tons by 2030 (Masuda et al., 2009). An increase in yield will be a driver for this expected jump. Soybean yield has been steadily increasing, with the United States average being 3187.7 kg ha^{-1} in 2019 compared to 1815.8 kg ha^{-1} in 1988 (soystats.com). Soybean yield is the most important factor considered by a farmer because soybean prices are determined by weight rather than quality (Yaklich et al., 2001). Yield can be challenging to breed for because it is a quantitative trait (Diers et al., 1992).

Aside from soybean yield, plant architecture is important to growers. The amount a soybean line lodges is critical to a successful harvest. Soybeans with a high lodging score will be difficult to harvest with a combine because the header can have trouble reaching bent plants. Height can also factor into the ease at which harvest is carried out. Soybeans that are too tall can become tangled in a combine header.

Seed quality is important to breeders, as global markets demand high quality soy profiles. Soybean quality must be higher in the United States than that of competitors to consistently win contracts from soybean importers such as China. In 2014 alone, China imported over 70 million tons of soybeans (Hairong et al., 2016). Important seed qualities include seed protein and seed oil. Soybean seed protein averages around 39.6% protein (Brzostowski et al., 2017) and between 19.8% and 21.2% oil content (Yaklich et al., 2002). An inverse relationship exists between seed protein and seed oil (Burton et al., 1987), which presents a challenge to combine both at a high rate in a soybean line. Soybean lines in this study are analyzed for various agronomic and seed quality traits.

Population Structure and Field Layout

A mapping population of 200 $F_{4:6}$ recombinant inbred lines (RILs) was created from a cross between TN13-4303 and TN12-4098 (See Table 2.1 in Appendix A). The RILs are segregating for protein, oil, and fatty acids. TN13-4303 is a line with high seed protein content and TN12-4098 is a line that is low in linolenic acid (<3%). The mapping population is MG-4L with white flowers. The mapping population was grown in two locations in 2018. The locations included Springfield, Tennessee (Highland Rim Research and Education Center, HRREC) and Knoxville, TN (East Tennessee Research and Education Center, ETREC). These locations were selected due to the difference in geographic location. At each location, the population of 200 RILs, two parents, and two checks was organized into a randomized complete block design with two replications. Lines that were selected for checks were top performers and were used for comparison of data. Each line was planted in a two-row plot, with seeding density set at 32.8 seeds per row meter. Plot length was planted at 6.1-meter rows and harvested in 4.9-meter rows. Harvest was done with an ALMACO SPC40 combine once soybeans reached maturity (~13% moisture content). This was repeated in 2019, with an additional location in Research and Education Center at Milan (RECM) and an additional replication to each location (totaling 3 replications for each location). The additional replication and location were not added to the first year of the study due to limited seed.

Phenotyping

Field Notes

Phenotyping was carried out on the mapping population at every location in 2018 and 2019. Measurements taken included plant height (cm), lodging (1-5), pubescence (grey or tawny), maturity (Julian calendar), and flower color. TN13-4303 had grey pubescence, and TN12-4098 had tawny pubescence. Plant height was measured with modified PVC rulers in inches and converted to cm, and the remaining notes (lodging, pubescence, maturity) were called with eyesight. Maturity was called up to three days in advance in the field. Any soybeans that displayed incorrect height, pubescence, maturity, or flower color within a plot were rouged. Any plot that was missing more than 0.3 m of soybeans from either row was adjusted for in the statistical analysis.

NIR Analysis (Protein and Oil)

Once harvest was completed, every plot from each location and year was subsampled and taken to the lab to be analyzed with near infrared reflectance spectroscopy (NIR). A whole bean analysis was carried out with a 30 g sample on a Perten DA-7250 NIR to obtain protein and oil. NIR analysis began by making sure that the machine was calibrated. This was done using a polystyrene sample that was placed under the light and read. If the machine had a sufficient calibration, subsamples were individually poured into a metal cup holding the seeds, and this cup was placed onto a tray under the instrument. Seeds were leveled before insertion under the machine, and each subsample was required to at least fill the bottom of the sample cup to eliminate misreads. Magnets

in the base of the instrument aligned the sample directly under the light for analysis. The sample was then be rotated under the light and analyzed for protein and oil. The data was relayed on a dry matter basis. The subsamples were then put back into their respective bags and placed in cold room storage for future use.

Fatty Acid Analysis

Although the NIR equations for fatty acids are improving, we analyze linolenic acid (18:3) though the primary chemistry method of gas chromatography. A Hewlett-Packard 6890 Gas Chromatograph was used to detect fatty acid levels in the seed. Seeds from all plots of all locations and years were subsampled into packaging envelopes. Five seeds were taken from each envelope and crushed with a hammer. The crushed seed was poured into its respective test tube. This process was repeated for 100 samples, which comprised a run. The samples were transported to a flow hood and received a 3 mL pump of extraction solvent. The extraction solvent is a solution of chloroform, hexanes, and methanol. To make the extraction solvent, 2000 mL chloroform, 1250 mL hexanes, and 500 mL methanol were mixed into a 4 L amber glass bottle fitted with a pipette pump. These tubes were then capped and sat for approximately 6-18 hours. The tubes were then uncapped, and 100 µL solution was pipetted into a 1.8 mL autosampler vial. The vial then received 0.75 mL of hexanes and 75 µL methylation reagent. Methylation reagent consisted of 5 mL 0.5M sodium methoxide solution in methanol, 10 mL ethyl ether, and 2 mL petroleum ether. Vials were then capped with an automatic crimping machine and placed onto a rack specially designed for the GC autosampler. A file on the GC was created for each run, producing a complete Excel sheet for each sample. The

Excel sheet showed a graph of the retention times for each of the five fatty acids. Underneath the graph, values were displayed for the percentages of each fatty acid. Any soybean sample that fell below 3% linolenic acid was considered ideal for low linolenic acid.

Statistical Analysis

SAS 9.4

SAS 9.4 (SAS Institute Inc., Cary, NC, USA) was used to analyze the 2018 and 2019 data individually, and then to look at combined data over all locations. PROC UNIVARIATE function was used to look at normality of data relating to each trait of interest. For each year, data was combined from locations (2018: HRREC and ETREC; 2019: HRREC, ETREC, and RECM). Normality was checked with the Shapiro-Wilks value at $p < 0.05$. After normality was analyzed, data was subject to PROC GLIMMIX to look at significant factors in the model. Two-year averages of each line were generated for all traits of interest using the %mmaov DANDA SAS macro developed by Dr. Arnold Saxton at the University of Tennessee. This macro also generated LSD values and correlation values between traits of interest. Correlations between traits of interest are included in Table 2.8 located in Appendix A. To calculate meal protein, the following formula was used: Meal Protein= [Protein 13% / (1-Oil 13%/100)]/.92. This formula was used by Pantalone and Smallwood (2018).

Microsoft Excel

Microsoft Excel was used to create frequency distributions for major traits of interest (found in figures 2.1-2.4 in Appendix A). Two-year averages were taken for each RIL and sorted in Excel from smallest to largest. Bin intervals were established for each trait. This resulted in approximately 7-10 bins per trait. Values falling into each bin were then counted and assigned appropriately. A histogram was generated for each trait of interest. Two-year averages were calculated for the parents and arrows above the bins in the histograms represented their values. The placement of the arrows allowed for visual representation of transgressive segregation for each trait.

Results and Discussion

Although the study started with 200 RILs, 20 lines had to be dropped from the study due to contamination or inaccurate genotyping. This brought the total RIL count to 180. It is also worth noting that 2018 was an abnormally wet field season in East Tennessee. Lodging was also a problem in this population. Typically, a taller soybean will lead to increased lodging. An example of this can be seen in a RIL population studied by Mansur et al. in 1996, where lodging and height had a correlation coefficient of 0.84. This study had a RIL population with an average height of 101.6 cm, and many plots with lodging values greater than 4, which led to soybeans lying flat on wet soil. This led to poor seed quality for the 2019 growing season. 2019 had poor germination related to seed fungus and dry, warm weather following periods of hard rain at planting. The germination issues led to re-planting and delayed emergence. In general, less favorable results came from the 2019 growing season. It is known that later planting dates can

affect agronomic traits such as seed yield, oil, and protein. An example of this phenomenon is shown in a study conducted by Beatty et al. in 1982, where soybeans dropped values in seed yield, oil, and protein when planted on a later date. In our study, protein content fell from a 2018 average of 405.9 g kg^{-1} to 397.7 g kg^{-1} in 2019. Oil fell from 234.4 g kg^{-1} to 213.6 g kg^{-1}. Meal protein dropped from 482.2 g kg^{-1} in 2018 to 461.9 g kg^{-1} in 2019. Yield dropped from 3339.1 kg ha^{-1} to 3062.0 kg ha^{-1}. Linolenic acid became more favorable with a decrease from 4.2% in 2018 to 4.1% in 2019. TNPL-123 was the highest performing RIL in terms of yield with a two-year average of 4037.8 kg ha^{-1}. This can be found in Table 2.2 located in Appendix A. TNPL-123 outperformed all checks in the study, including Ellis, which had a two-year average yield of 3854.0 kg ha^{-1}. Top performing RILs for protein, oil, linolenic acid, and meal protein can be found in Appendix A in Table 2.3, 2.4, 2.5, and 2.6, respectively. TNPL-111 was the top RIL for dry protein content with 42.1%. TNPL-146 was the top performing RIL for oil content, with dry matter content at 23.3%. TNPL-077 had the lowest linolenic acid content with an average of 2.3%, which is lower than the 3% standard (Pham et al., 2012; Smallwood et al., 2017) for low linolenic acid soybean lines. TNPL-146 had the highest meal protein value at 49.1%. TNPL-146 was in the top 10% of RILs when looking at protein, oil, and meal protein, ranking 9th, 1st, and 1st, respectively. TNPL-146 is an example of a line that was able to achieve high protein and oil concentrations, going against the typical inverse correlation as described by Burton et al. 1987. The seed yield of TNPL-146 was 2987.0 kg ha^{-1}. Although high protein and oil concentrations are not typically attractive to farmers when compared to yield (Yaklich et al. 2002), lines such as

TNPL-146 could be backcrossed into a high yielding line to produce an overall superior line, such as described in a 1995 study by Wilcox and Cavins. Only a small fraction of soybeans (~2%) are consumed by humans (Goldsmith 2008). This leaves most soybean production used directly for livestock feed applications. 98% of all soybean meal is used to feed animals (Hartman et al., 2011), which makes lines such as TNPL-146 (MP = 49.1%) valuable to producers. TNPL-146 did fail to surpass its parent, TN13-4303 and a check, TN15-5007 (Pantalone et al., 2020), which had meal protein values of 49.2% and 49.5% respectively.

The frequency distribution illustrating the two-year averages of linolenic acid content (Figure 2.4, Appendix B) is worth noting. Although the distribution is considered normal, two bins have noticeably higher peaks than others at “3.51-4.00” and “2.51-3.00”. This is most likely the result of a portion of the population containing the mutant alleles of the FAD3A or FAD3C gene at the “2.51-3.00” bin, which is known to lower the concentration of seed linolenic acid significantly (Bilyeu et al., 2003; Bilyeu et al., 2005; Held et al., 2019). To be positive that these lines contained one or more of the genes, SNPs would need to be screened using technology such as a light-cycler. This is an example of how marker-assisted selection can be a powerful tool in screening for traits that cannot otherwise be seen.

REFERENCES

Beatty, K. D., Eldridge, I. L., & Simpson Jr, A. M. (1982) Soybean Response to Different Planting Patterns and Dates. *Agronomy Journal*, **74**: 859-862.

Bilyeu, K. D., Palavalli, L., Sleper, D. A., & Beuselinck, P. R. (2003) Three Microsomal Omega-3 Fatty-acid Desaturase Genes Contribute to Soybean Linolenic Acid Levels. *Crop Science*, **43**:1833-1838.

Bilyeu, K., Palavalli, L., Sleper, D., & Beuselinck, P. (2005) Mutations in Soybean Microsomal Omega-3 Fatty Acid Desaturase Genes Reduce Linolenic Acid Concentration in Soybean Seeds. *Crop Science*, **45**: 1830-1836.

Brzostowski, L. F., Pruski, T.I., Specht, J.E., & Diers, B. W. (2017) Impact of seed protein alleles from three soybean sources on seed composition and agronomic traits. *Theoretical and Applied Genetics*, **130**: 2315–2326.

Burton, J. (1987) Quantitative genetics: Results relevant to soybean breeding. 111 IR. Wilcox (ed) Soybeans: Improvement, Production, and Uses. *American Society of Agronomy*.

Diers, B.W., Keim, P., Fehr, W.R., & Shoemaker, R.C. (1992) RFLP analysis of soybean seed protein and oil content. *Theoretical and Applied Genetics*, **83**: 608 612.

Goldsmith, P. D. (2008) Economics of soybean production, marketing, and utilization. In *Soybeans*, pp. 117-150.

Hairong, Y., Yiyuan, C., & Bun, K. H. (2016) China's soybean crisis: the logic of modernization and its discontents. *The Journal of Peasant Studies*, **43**: 373-395.

Hartman, G. L., West, E. D., & Herman, T. K. (2011) Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. *Food Security*, **3**: 5-17.

Held, J. P., Carrero-Colón, M., & Hudson, K. A. (2019) Combination of Novel Mutation in FAD3C and FAD3A for Low Linolenic Acid Soybean. *Agrosystems, Geosciences & Environment*, **2:** 1.

Masuda, T., & Goldsmith, P. D. (2009) World soybean production: area harvested, yield, and long-term projections. *International food and agribusiness management review*, **12**: 1-20.

Pantalone, V., Cunicelli, M., & Wyman, C. (2020) Registration of soybean cultivar 'TN15-5007'with high meal protein. *Journal of Plant Registrations*, **14**: 139-143.

Pantalone, V., & Smallwood, C. (2018) Registration of 'TN11-5102' soybean cultivar with high yield and high protein meal. *Journal of Plant Registrations*, **12:** 304-308.

Pham, A. J., Shannon, G., & Bilyeu, K. D. (2012) Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. *Theoretical and Applied Genetics*, **125**: 503–515.

Smallwood, C. J., Gillman, J. D., Saxton, A. M., Bhandari, H. S., Wadl, P. A., Fallen, B. D., ... & Pantalone, V. R. (2017) Identifying and exploring significant genomic regions associated with soybean yield, seed fatty acids, protein and oil. *Journal of crop science and biotechnology*, **20**: 243-253.

Soy Stats. (2020) *U.S. Yield and Production: Yield History*. The American Soybean Association. Retrieved from http://soystats.com/u-s-yield-production-yield history/

Wilcox, J.R., and Cavins, J.F. (1995) Backcrossing High Seed Protein to a Soybean Cultivar. *Crop Science*, **35**: 1036–1041.

Yaklich, R.W., Vinyard, B., Camp, M., & Douglass, S. (2002) Analysis of Seed Protein and Oil from Soybean Northern and Southern Region Uniform Tests. *Crop Science*, **42**: 1504-1515.

Yaklich, R.W. (2001) β-Conglycinin and Glycinin in High-Protein Soybean Seeds. *Journal of Agricultural and Food Chemistry*, **49**: 729–735.

APPENDIX A

Table 2.1: History of PRO-LIN Population Development (Personal communication: Dr. Vince Pantalone)

Year	Location	Generation	Activity	Rows
2019	ETREC, HRREC, RECM	$F_{4:7}$	Yield trials	14,001-16,606
2018	ETREC, HRREC	$F_{4:6}$	Yield trials	13,001-13,908
2017/2018 WN	Santa Isabel, PR	$F_{4:5}$	Seed increase	VM18-2241-VM18-2588
2017	ETREC	F_4	Pull Single Plants	41,125-41,170
2016/2017 WN	Santa Isabel, PR	F_3	Pod pick	VM17-052-073
2016	ETREC	F_2	Pod pick	20,070-20,090
2015/2016 WN	Isabela, PR	F_1	Grow F1 plants	VP056-VP065
2015	ETREC	P_1 x P_2	Make Cross	Cross 15-09 (TN12-4098 x TN13-4303)

Table 2.2: Two-year means of all RILs along with parents and checks for various traits of interest sorted by yield.

Line	Yield	DAP	Lodging	Height	Protein	Oil	18 : 3	Meal
	kg ha^{-1}		1-5	cm	g kg^{-1}	g kg^{-1}	%	g kg^{-1}
TNPL-123	4037.8	132.1	2.6	114.8	391.7	226.3	5.7	461.4
TNPL-122	3927.3	130.9	2.0	77.2	387.7	226.4	3.2	456.7
Ellis[b]	3854.0	129.9	1.7	68.4	403.7	218.5	7.0	471.3
TNPL-136	3816.6	134.4	2.5	122.5	410.4	226.4	5.8	483.2
TNPL-098	3793.2	131.3	2.0	77.6	402.6	224.2	3.0	473.0
TNPL-115	3729.1	133.5	2.1	76.6	388.5	217.8	2.6	453.4
TNPL-186	3706.5	133.6	2.3	116.2	393.3	225.3	4.5	462.5
TNPL-052	3695.1	131.5	2.3	84.3	386.6	221.4	2.6	452.9
TNPL-108	3677.3	132.6	2.2	87.0	402.8	217.6	2.8	469.3
TNPL-105	3639.3	133.7	2.1	79.2	410.8	218.8	4.7	479.9
TNPL-019	3620.4	132.2	2.0	80.0	409.5	211.5	4.2	474.6
TNPL-165	3610.9	134.4	2.4	120.3	409.0	227.3	5.9	482.2
TN15-5007[b]	3606.7	128.9	2.0	70.3	427.3	211.0	6.8	494.8
TNPL-167	3604.6	130.5	2.0	77.3	396.2	223.6	4.8	465.2
TNPL-151	3602.2	131.9	2.0	70.5	397.6	211.9	3.3	460.9
TNPL-059	3599.7	131.2	2.0	72.1	390.2	223.2	7.0	457.9
TNPL-043	3578.6	131.0	1.9	72.8	406.4	228.7	4.6	479.8
TNPL-154	3578.4	132.5	2.1	81.1	400.0	220.4	3.1	468.1
TNPL-124	3572.5	133.1	1.9	78.7	401.3	212.8	7.1	465.9
TNPL-006	3553.6	132.6	2.1	77.1	394.4	220.4	3.5	461.6
TNPL-028	3548.3	128.8	2.1	74.2	408.9	221.6	6.7	479.0
TNPL-175	3537.1	132.8	3.2	123.1	397.0	229.7	3.0	469.3
TNPL-109	3530.0	132.3	1.8	70.5	407.4	209.6	2.8	471.2
TNPL-101	3529.7	131.0	1.9	73.7	397.8	227.8	6.4	469.3
TNPL-081	3528.4	133.2	3.2	122.2	401.1	221.1	2.6	469.7
TNPL-014	3525.8	131.9	1.9	81.4	403.6	219.2	6.3	471.7
TNPL-097	3522.6	131.4	1.9	78.0	411.8	224.4	4.3	483.9
TNPL-155	3501.8	131.0	2.2	78.1	400.9	220.1	3.8	468.9
TNPL-005	3494.4	132.7	2.0	78.8	393.1	219.0	2.5	459.3
TNPL-064	3493.4	132.1	2.1	75.2	398.1	218.2	2.6	464.8
TNPL-118	3490.5	131.2	2.9	116.9	405.6	221.9	5.3	475.2
TNPL-070	3489.2	130.8	2.0	76.7	410.5	230.7	6.1	485.8
TNPL-022	3483.3	132.2	3.5	123.3	399.5	224.9	3.8	469.8
TNPL-023	3482.9	132.8	3.2	123.2	403.7	222.3	3.9	473.5
TNPL-148	3476.3	131.9	2.2	73.4	397.8	217.4	2.5	464.1
TNPL-080	3461.0	131.8	2.1	72.2	414.9	214.7	6.5	482.5
TNPL-017	3458.3	132.6	2.0	73.2	407.6	212.2	2.7	472.9

Table 2.2, continued

Line	Yield	DAP	Lodging	Height	Protein	Oil	18 : 3	Meal
	kg ha^{-1}		1-5	cm	g kg^{-1}	g kg^{-1}	%	g kg^{-1}
TNPL-078	3454.1	131.7	2.6	116.5	399.6	224.4	3.8	469.7
TNPL-048	3451.3	131.7	3.2	118.6	398.8	228.6	3.9	470.9
TNPL-119	3450.2	131.9	3.2	119.5	415.9	220.5	4.7	486.8
TNPL-012	3439.3	132.1	2.6	115.2	392.9	220.3	4.0	459.8
TNPL-082	3439.1	130.4	2.0	67.1	413.0	213.5	5.8	479.6
TNPL-053	3435.3	132.3	2.1	80.4	380.8	221.2	3.6	446.2
TNPL-152	3430.1	132.2	2.7	117.5	405.5	213.8	4.3	471.2
TN13-4303[a]	3426.9	130.3	1.8	72.6	422.5	215.7	7.1	491.8
TNPL-111	3423.9	132.7	2.0	73.3	420.5	213.4	6.0	488.4
TNPL-049	3416.9	134.7	2.9	126.0	399.8	219.8	2.4	467.6
TNPL-160	3407.3	132.1	2.5	115.8	389.9	219.4	3.2	455.9
TNPL-163	3405.7	133.2	2.4	110.1	403.2	223.0	6.1	473.1
TNPL-185	3405.2	133.3	2.3	116.4	399.9	222.1	4.4	468.7
TNPL-087	3405.0	130.4	2.0	71.1	405.2	213.5	2.7	470.6
TNPL-135	3404.1	132.5	2.9	116.5	395.9	230.2	3.0	468.2
TNPL-062	3402.5	128.3	2.9	114.5	405.5	227.5	3.5	478.4
TNPL-069	3395.0	133.0	2.7	123.4	400.6	226.4	4.8	471.8
TNPL-073	3386.0	134.5	2.9	120.1	390.9	218.4	3.8	456.5
TNPL-025	3383.9	131.9	2.8	122.5	406.0	224.4	4.7	477.1
TNPL-100	3367.4	133.0	2.6	116.1	397.3	218.4	2.7	464.0
TNPL-077	3359.6	134.9	3.6	121.4	392.6	223.6	2.3	460.9
TNPL-040	3349.5	132.1	2.0	79.6	394.0	219.9	3.6	461.0
TNPL-107	3347.9	132.4	2.1	76.9	412.7	224.7	3.1	485.2
TNPL-027	3342.0	129.4	2.0	69.9	410.7	225.2	6.6	483.1
TNPL-090	3340.6	131.4	2.0	72.8	395.0	226.3	4.1	467.4
TNPL-161	3339.5	132.0	2.0	79.6	399.4	223.1	3.9	468.8
TNPL-169	3328.3	134.3	2.8	116.5	398.9	218.5	2.5	465.8
TNPL-128	3322.1	131.0	3.0	118.8	405.4	219.0	5.3	473.6
TNPL-192	3316.4	133.5	2.4	112.7	401.5	220.1	3.6	471.7
TNPL-066	3315.4	132.2	3.3	120.1	397.0	226.5	4.0	467.6
TNPL-021	3308.6	130.5	2.6	111.0	409.9	225.6	5.6	482.4
TNPL-054	3308.4	131.5	3.0	107.0	386.8	219.2	2.6	452.1
TNPL-103	3307.3	131.5	2.4	110.4	397.0	227.3	4.8	468.0
TNPL-178	3305.3	134.3	2.7	120.4	408.5	221.1	4.6	478.4
TNPL-129	3299.9	131.6	2.4	104.7	398.7	226.4	5.4	469.7
TNPL-150	3299.0	132.3	2.1	69.8	390.7	221.3	2.4	457.7
TNPL-127	3284.1	132.2	2.0	76.3	396.9	225.9	4.2	467.2
TNPL-149	3284.0	129.6	3.2	116.1	393.5	223.6	3.9	461.1

Table 2.2, continued

Line	Yield	DAP	Lodging	Height	Protein	Oil	18 : 3	Meal
	kg ha^{-1}		1-5	cm	g kg^{-1}	g kg^{-1}	%	g kg^{-1}
TNPL-067	3277.4	132.0	2.0	80.7	400.1	216.4	5.7	466.2
TNPL-176	3276.8	130.3	3.4	120.1	401.9	230.7	4.9	475.6
TNPL-162	3275.2	132.2	2.8	113.2	400.7	220.9	5.6	469.1
TNPL-195	3271.8	132.6	3.1	119.3	398.9	228.7	4.9	471.0
TNPL-034	3265.4	134.1	2.5	111.5	395.5	215.6	2.7	460.5
TNPL-200	3263.1	130.8	2.0	71.7	407.4	224.4	3.7	479.4
TNPL-196	3262.0	131.2	2.0	68.7	392.6	218.3	4.0	458.4
TNPL-141	3254.1	133.2	2.5	115.4	395.3	219.8	2.7	462.3
TNPL-060	3252.7	132.9	3.5	116.6	400.1	225.5	2.7	470.7
TNPL-114	3245.5	134.2	2.6	109.9	409.9	218.2	3.2	478.5
TNPL-033	3242.2	133.9	2.5	115.4	398.8	216.4	2.7	464.7
TNPL-002	3237.1	131.8	3.2	112.5	404.0	227.5	4.6	476.4
TNPL-003	3236.2	133.2	2.0	77.0	402.3	216.7	3.6	468.9
TNPL-031	3233.7	134.3	3.0	116.2	385.3	222.0	2.5	451.7
TNPL-158	3229.6	132.2	2.0	76.6	411.3	216.9	5.3	479.5
TNPL-172	3228.7	134.8	2.5	114.2	411.0	219.1	3.8	479.7
TNPL-038	3225.6	131.0	3.1	121.1	401.0	225.0	5.9	471.6
TNPL-011	3224.9	133.9	2.9	114.3	403.6	215.8	3.2	470.0
TNPL-007	3219.7	134.6	2.5	105.5	398.9	222.5	2.8	467.8
TNPL-121	3219.2	133.5	2.3	109.9	402.0	222.1	5.1	470.6
TNPL-157	3218.8	134.2	3.3	125.2	391.7	220.1	3.1	456.1
TNPL-182	3213.5	130.9	1.9	67.9	396.7	220.2	3.5	464.1
TNPL-051	3205.8	133.2	3.3	118.4	394.3	219.7	3.7	461.1
TNPL-020	3203.3	132.1	3.1	118.6	399.3	225.3	3.7	469.8
TNPL-039	3201.5	133.1	1.9	69.5	401.2	222.4	2.9	470.5
TNPL-010	3199.6	130.4	3.0	117.5	403.9	223.4	5.7	475.9
TNPL-083	3187.3	132.1	2.6	105.5	400.5	228.3	4.6	472.7
TNPL-047	3187.0	131.1	2.2	75.8	406.4	220.7	5.0	475.8
TNPL-024	3167.7	132.7	3.4	117.0	404.3	224.8	3.4	475.3
TNPL-032	3163.1	130.6	2.1	74.6	387.0	219.3	4.0	452.5
TNPL-093	3157.2	134.7	3.1	120.8	387.4	220.2	2.4	453.3
TN12-4100[b]	3151.7	127.3	2.0	65.2	398.6	227.7	7.9	470.3
TNPL-188	3142.0	129.4	1.9	72.1	397.8	217.0	3.7	463.8
TNPL-164	3133.7	130.1	2.5	103.6	403.9	229.8	4.4	477.5
TNPL-191	3131.5	132.6	3.0	109.2	394.9	227.0	2.8	466.0
TNPL-001	3126.4	129.9	2.8	109.7	409.1	222.5	6.1	479.7
TNPL-193	3113.4	133.1	3.3	110.0	388.8	224.3	5.5	456.9
TNPL-086	3099.5	133.2	2.9	114.3	396.5	221.4	3.1	464.5
TNPL-153	3093.1	132.4	3.0	120.8	414.7	225.4	4.7	488.0

Table 2.2, continued

Line	Yield	DAP	Lodging	Height	Protein	Oil	18 : 3	Meal
	kg ha^{-1}		1-5	cm	g kg^{-1}	g kg^{-1}	%	g kg^{-1}
TNPL-179	3091.8	131.0	1.9	66.4	401.9	222.3	4.8	471.3
TNPL-046	3088.8	132.4	2.8	115.8	391.5	227.2	3.1	461.5
TNPL-091	3075.0	129.7	2.0	69.1	401.7	221.2	6.3	470.5
TNPL-138	3074.7	132.9	2.7	109.4	388.7	225.4	3.0	457.4
TNPL-089	3068.6	131.9	3.0	114.0	391.8	225.2	4.0	460.8
TNPL-189	3067.1	133.9	2.2	110.7	393.0	223.5	4.1	461.4
TNPL-104	3059.5	133.3	2.2	78.2	404.9	213.7	5.7	470.3
TNPL-197	3051.7	132.1	3.1	115.5	396.3	226.7	4.0	467.0
TNPL-058	3051.5	128.9	2.1	73.5	405.4	217.6	4.7	473.0
TNPL-068	3045.0	131.6	2.9	104.6	383.6	219.6	2.8	448.5
TNPL-018	3044.8	133.8	2.4	110.1	398.9	220.6	4.7	467.0
TNPL-139	3026.7	132.7	2.4	100.0	406.4	218.0	5.6	474.3
TNPL-088	3021.8	131.1	3.1	115.5	414.8	225.7	6.0	488.2
TNPL-110	3017.8	132.1	2.0	76.1	392.7	221.3	3.6	460.1
TNPL-143	3000.4	134.2	2.8	114.2	395.3	221.9	2.6	463.4
TNPL-187	2993.6	133.4	2.3	102.3	398.0	221.4	4.7	466.7
TNPL-044	2992.8	132.0	2.9	114.5	387.2	222.6	2.7	454.2
TNPL-140	2988.7	132.6	2.0	80.1	396.0	218.8	3.9	462.5
TNPL-117	2987.1	132.5	2.9	112.3	406.3	230.8	3.5	480.9
TNPL-029	2985.2	130.3	3.4	115.1	413.3	224.0	5.2	485.5
TNPL-016	2979.8	128.9	1.9	66.5	414.9	215.5	5.5	482.9
TNPL-035	2971.0	132.3	3.0	116.9	411.0	222.8	3.7	482.2
TNPL-084	2970.1	131.6	3.7	111.5	389.6	226.2	3.0	458.7
TNPL-131	2966.9	132.2	3.0	115.2	406.4	225.1	4.5	478.2
TNPL-171	2963.2	129.6	2.8	117.5	401.5	228.4	5.2	471.4
TNPL-147	2953.1	129.3	2.0	76.5	408.3	218.7	6.0	477.0
TNPL-146	2947.0	131.8	2.9	104.5	413.9	233.4	5.5	491.0
TNPL-145	2946.6	132.9	2.3	107.1	396.7	218.2	3.9	463.4
TNPL-092	2935.0	133.5	2.8	112.4	399.6	215.8	2.7	465.6
TNPL-042	2909.5	132.2	2.5	102.8	393.7	212.7	2.9	457.0
TNPL-065	2908.6	132.8	3.3	128.3	409.8	222.9	4.7	480.7
TNPL-166	2904.5	134.0	3.0	118.4	403.1	221.4	3.0	472.4
TNPL-063	2904.5	131.7	3.0	110.5	389.4	226.5	3.0	458.7
TNPL-096	2902.4	130.3	2.0	75.9	400.1	220.5	3.2	468.3
TNPL-057	2901.6	132.0	3.1	116.8	409.1	221.3	5.7	479.2
TNPL-095	2894.1	130.7	2.0	70.8	410.5	219.0	5.8	479.6
TNPL-071	2889.5	132.6	3.0	122.2	406.7	224.5	5.4	478.1
TNPL-112	2874.1	131.8	2.8	110.2	409.4	221.2	4.5	479.6
TNPL-076	2855.5	132.9	3.0	113.1	400.2	220.5	5.4	468.5

Table 2.2, continued

Line	Yield	DAP	Lodging	Height	Protein	Oil	18 : 3	Meal
	kg ha^{-1}		1-5	cm	g kg^{-1}	g kg^{-1}	%	g kg^{-1}
TNPL-198	2850.4	129.9	2.5	120.3	398.0	217.2	3.7	464.3
TNPL-113	2846.6	132.9	3.0	114.5	409.2	221.6	3.8	479.5
TNPL-116	2845.4	133.6	3.1	118.2	397.0	216.8	3.9	462.7
TNPL-170	2845.3	128.4	2.4	101.6	406.1	220.0	5.1	475.1
TNPL-156	2823.8	131.4	3.1	120.3	394.6	226.8	4.4	465.0
TNPL-130	2815.9	134.1	2.5	125.7	392.5	228.0	2.3	463.3
TNPL-055	2804.9	131.8	3.3	115.9	406.7	223.3	3.8	477.5
TNPL-045	2801.3	131.3	2.6	106.2	379.9	220.4	2.6	444.7
TNPL-030	2760.4	133.3	3.0	119.6	409.2	219.2	3.9	478.1
TNPL-168	2721.3	131.7	3.2	110.8	399.1	216.6	5.7	465.2
TNPL-126	2719.4	133.1	2.5	111.8	402.2	226.1	5.9	474.4
TNPL-159	2707.5	132.0	2.6	113.5	393.7	227.6	3.5	464.4
TNPL-106	2704.5	130.7	3.2	128.9	408.2	224.4	5.8	479.7
TNPL-072	2703.1	132.3	3.1	121.1	390.7	220.5	3.7	457.3
TNPL-174	2702.0	130.4	2.7	109.5	403.6	219.0	5.8	471.6
TNPL-015	2680.4	128.7	4.0	123.9	395.2	224.0	2.8	464.2
TNPL-004	2660.7	134.1	3.5	116.4	390.7	219.7	3.2	456.9
TNPL-061	2659.9	130.4	2.6	107.5	409.2	226.5	2.9	482.1
TNPL-074	2648.9	130.9	2.6	101.4	396.5	222.1	4.8	465.1
TNPL-134	2637.9	130.6	3.1	104.8	403.3	228.8	4.4	476.3
TNPL-075	2618.5	131.9	3.0	105.8	408.4	221.7	5.9	478.6
TNPL-177	2530.5	130.2	2.5	110.3	405.3	227.1	5.6	477.7
TNPL-102	2529.0	132.6	2.3	98.3	405.1	222.9	3.2	477.5
TNPL-133	2478.0	133.7	3.6	116.5	390.2	212.7	3.0	452.9
TNPL-008	2476.3	130.6	2.7	112.7	414.1	220.7	5.2	485.4
TNPL-099	2382.6	131.0	2.0	80.8	417.9	220.4	2.7	488.3
TNPL-142	2374.7	128.9	2.5	107.5	403.4	229.0	4.8	476.6
TNPL-199	2336.3	131.0	3.4	121.8	398.7	227.7	3.0	470.3
TNPL-184	2043.3	132.2	3.2	106.7	402.1	225.8	3.2	473.3
TNPL-183	1905.7	124.5	2.7	95.7	403.1	227.9	2.9	475.2
TN12-4098[a]	1756.1	124.1	2.5	91.6	400.6	224.0	2.7	470.8
TNPL-094	1737.0	127.6	2.2	75.7	406.1	222.0	3.5	476.2
LSD (0.05)	**549.3**	**1.8**	**0.5**	**11.5**	**7.1**	**4.4**	**0.5**	**7.6**

[a] Parent of the RIL population.
[b] Check in field trials. Ellis and TN12-4100 used in 2018. Ellis and TN15-5007 used in 2019.

Table 2.3: Top 10% of RILs sorted by two-year mean protein content on a dry matter basis.

Line	Protein	Oil	Meal	Yield
	%	%	%	kg ha^{-1}
TNPL-111	42.1	21.3	48.8	3423.9
TNPL-099	41.8	22.0	48.8	2382.6
TNPL-119	41.6	22.1	48.7	3450.2
TNPL-080	41.5	21.5	48.2	3461.0
TNPL-016	41.5	21.6	48.3	2979.8
TNPL-088	41.5	22.6	48.8	3021.8
TNPL-153	41.5	22.5	48.8	3093.1
TNPL-008	41.4	22.1	48.5	2476.3
TNPL-146	41.4	23.3	49.1	2947.0
TNPL-029	41.3	22.4	48.6	2985.2
TNPL-082	41.3	21.3	48.0	3439.1
TNPL-107	41.3	22.5	48.5	3347.9
TNPL-097	41.2	22.4	48.4	3522.6
TNPL-158	41.1	21.7	47.9	3229.6
TNPL-035	41.1	22.3	48.2	2971.0
TNPL-172	41.1	21.9	48.0	3228.7
TNPL-105	41.1	21.9	48.0	3639.3
TNPL-027	41.1	22.5	48.3	3342.0
LSD (0.05)	**0.7**	**0.4**	**0.8**	**549.3**

Table 2.4: Top 10% of RILs sorted by two-year mean oil content on a dry matter basis.

Line	Oil	Protein	Meal	Yield
	%	%	%	kg ha^{-1}
TNPL-146	23.3	41.4	49.1	2947.0
TNPL-117	23.1	40.6	48.1	2987.1
TNPL-176	23.1	40.2	47.6	3276.8
TNPL-070	23.1	41.1	48.6	3489.2
TNPL-135	23.0	39.6	46.8	3404.1
TNPL-164	23.0	40.4	47.8	3133.7
TNPL-175	23.0	39.7	46.9	3537.1
TNPL-142	22.9	40.3	47.7	2374.7
TNPL-134	22.9	40.3	47.6	2637.9
TNPL-043	22.9	40.6	48.0	3578.6
TNPL-195	22.9	39.9	47.1	3271.8
TNPL-048	22.9	39.9	47.1	3451.3
TNPL-171	22.8	40.1	47.1	2963.2
TNPL-083	22.8	40.0	47.3	3187.3
TNPL-130	22.8	39.3	46.3	2815.9
TNPL-183	22.8	40.3	47.5	1905.7
TNPL-101	22.8	39.8	46.9	3529.7
TNPL-199	22.8	39.9	47.0	2336.3
LSD (0.05)	**0.4**	**0.7**	**0.8**	**549.3**

Table 2.5: Top 10% of RILs for two-year mean linolenic acid content sorted from smallest to largest.

Line	18:3
	%
TNPL-077	2.3
TNPL-130	2.3
TNPL-093	2.4
TNPL-150	2.4
TNPL-049	2.4
TNPL-005	2.5
TNPL-169	2.5
TNPL-148	2.5
TNPL-031	2.5
TNPL-115	2.6
TNPL-081	2.6
TNPL-052	2.6
TNPL-054	2.6
TNPL-045	2.6
TNPL-064	2.6
TNPL-143	2.6
TNPL-099	2.7
TNPL-044	2.7
LSD (0.05)	**0.5**

Table 2.6: Top 10% of RILs sorted by two-year mean meal protein content.

Line	Meal Protein	Protein	Oil	Yield
	%	%	%	kg ha^{-1}
TNPL-146	49.1	41.4	23.3	2947.0
TNPL-111	48.8	42.1	21.3	3423.9
TNPL-099	48.8	41.8	22.0	2382.6
TNPL-088	48.8	41.5	22.6	3021.8
TNPL-153	48.8	41.5	22.5	3093.1
TNPL-119	48.7	41.6	22.1	3450.2
TNPL-070	48.6	41.1	23.1	3489.2
TNPL-029	48.6	41.3	22.4	2985.2
TNPL-008	48.5	41.4	22.1	2476.3
TNPL-107	48.5	41.3	22.5	3347.9
TNPL-097	48.4	41.2	22.4	3522.6
TNPL-136	48.3	41.0	22.6	3816.6
TNPL-027	48.3	41.1	22.5	3342.0
TNPL-016	48.3	41.5	21.6	2979.8
TNPL-080	48.2	41.5	21.5	3461.0
TNPL-021	48.2	41.0	22.6	3308.6
TNPL-165	48.2	40.9	22.7	3610.9
TNPL-035	48.2	41.1	22.3	2971.0
LSD (0.05)	**0.8**	**0.7**	**0.4**	**549.3**

Table 2.7: Top 10% of RILs sorted by two-year mean seed yield.

Line	Yield	Protein	Oil	Meal
	kg ha^{-1}	%	%	%
TNPL-123	4037.8	39.2	22.6	46.1
TNPL-122	3927.3	38.8	22.6	45.7
TNPL-136	3816.6	41.0	22.6	48.3
TNPL-098	3793.2	40.3	22.4	47.3
TNPL-115	3729.1	38.9	21.8	45.3
TNPL-186	3706.5	39.3	22.5	46.3
TNPL-052	3695.1	38.7	22.1	45.3
TNPL-108	3677.3	40.3	21.8	46.9
TNPL-105	3639.3	41.1	21.9	48.0
TNPL-019	3620.4	41.0	21.1	47.5
TNPL-165	3610.9	40.9	22.7	48.2
TNPL-167	3604.6	39.6	22.4	46.5
TNPL-151	3602.2	39.8	21.2	46.1
TNPL-059	3599.7	39.0	22.3	45.8
TNPL-043	3578.6	40.6	22.9	48.0
TNPL-154	3578.4	40.0	22.0	46.8
TNPL-124	3572.5	40.1	21.3	46.6
TNPL-006	3553.6	39.4	22.0	46.2
LSD (0.05)	**549.3**	**0.7**	**0.4**	**0.8**

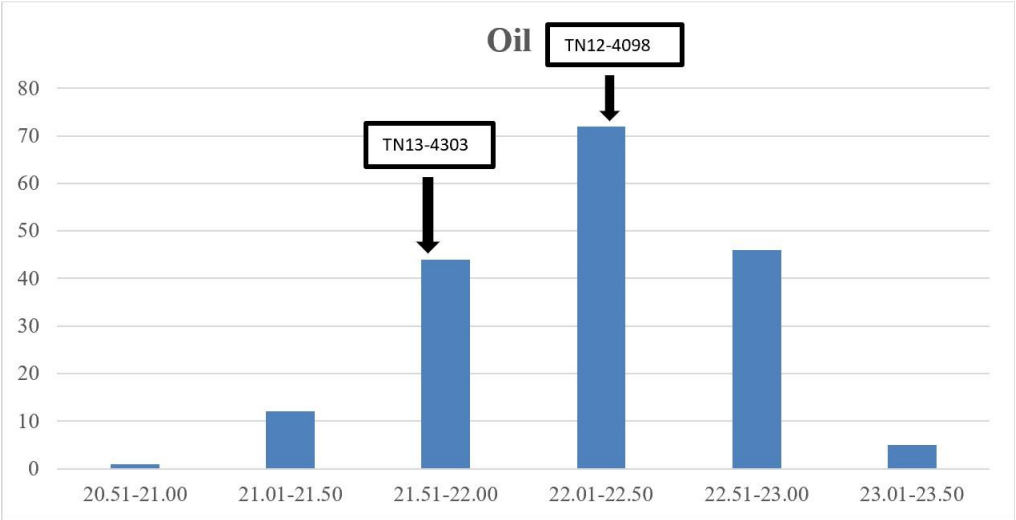

Figure 2.1: Frequency distribution of two-year average NIR oil data (whole bean analysis) on a dry weight basis.

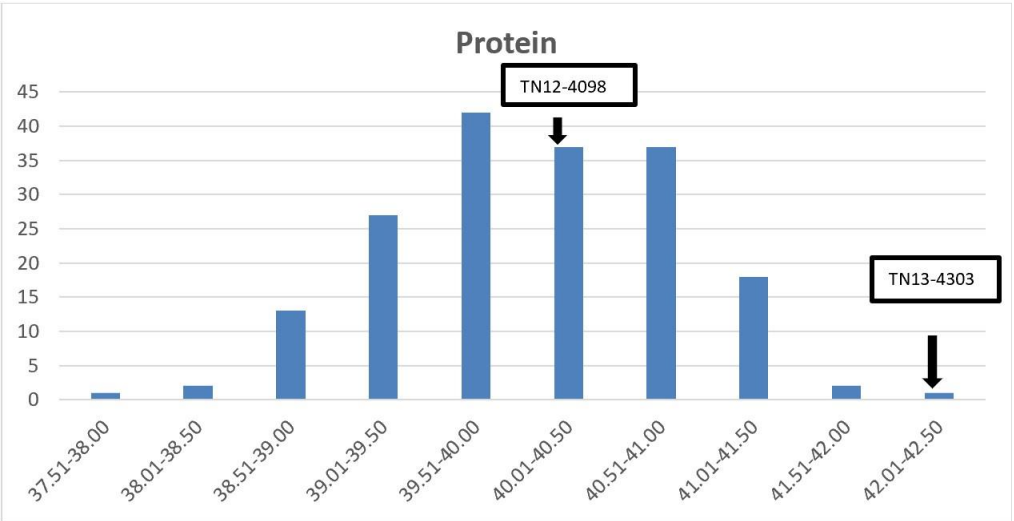

Figure 2.2: Frequency distribution of two-year average NIR protein data (whole bean analysis) on a dry weight basis.

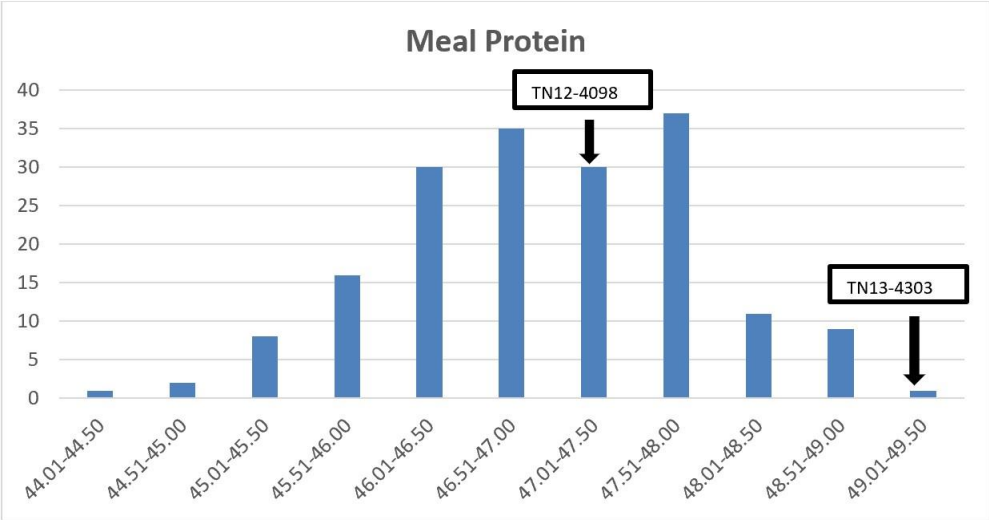

Figure 2.3: Frequency distribution of two-year average NIR meal protein data (whole bean analysis) on a dry weight basis.

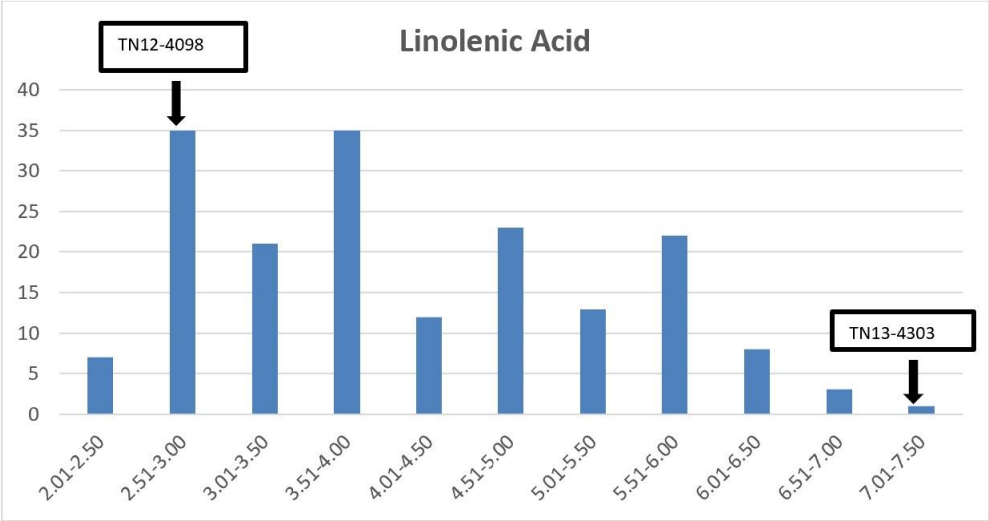

Figure 2.4: Frequency distribution of two-year average gas chromatography linolenic acid data. Bins represent percent of total fatty acids present.

Table 2.8: Correlation between yield, protein, oil, linolenic acid, and meal protein. Within each cell, correlation value is listed on the top and correlation significance is listed on the bottom. Correlation significance falling below 0.05 is shown in bold font.

Correlation Between Major Traits of Interest					
	Yield	Protein	Oil	Linolenic Acid	Meal Protein
Yield		-0.24 **<.0001**	0.23 **<.0001**	0.05 **0.03**	-0.14 **<.0001**
Protein			-0.01 0.60	0.29 **<.0001**	0.93 **<.0001**
Oil				0.03 0.17	0.37 **<.0001**
Linolenic Acid					0.28 **<.0001**

CHAPTER THREE

QTL ANALYSIS

Abstract

Quantitative traits such as seed protein, seed oil, yield, etc. are challenging to breed for because they are controlled by multiple genes. Quantitative trait loci (QTL) offer breeders a tool to target and incorporate quantitative traits into desired lines. QTL can be screened for by scanning a line for a SNP that flanks the locus of interest. After screening the population of 180 RILs, 21 QTL were identified using WinQTLCart2.5. 1 QTL for DAP, 2 for lodging, 1 for plant height, 1 for seed yield, 6 for seed protein, 3 for seed oil, 3 for seed linolenic acid, and 4 for meal protein were detected. Of the 21 QTL, 9 loci were considered novel after scanning SoyBase.

Introduction

Quantitative traits such as yield, seed oil, and seed protein content are not easily bred for because they are controlled by multiple genes throughout the soybean genome. Classical methods of breeding such as phenotypic selection can be extremely slow and unpredictable for these traits of interest, considering the large genetic and environmental uncertainty surrounding them. Identifying quantitative trait loci allows researchers to use marker assisted selection to target and utilize genomic regions of a chromosome controlling much of the variability in a trait. Quantitative trait loci are regions of a genome that control quantitative traits and are commonly identified by positioning of molecular markers adjacent to or within their reach (Collard et al., 2005). A QTL can either be “minor” or “major” depending on how much of the variability the loci explains for a trait of interest in a certain population. For example, in a 2005 study, Panthee et al.

found a protein QTL responsible for over 20% of the variation in the mapping population. This would be categorized as a major QTL. QTLs are commonly identified by crossing parents that differ in traits of interest, creating linkage maps after genotyping the resulting population, and then running analyses on computer programs such as QTL Cartographer (Collard et al., 2005). Composite interval mapping is the analysis method most used in research, as it gives the most accurate results and accounts for linked QTLs (Collard et al., 2005). Many QTLs exist for a range of soybean characteristics. As of 2020, there are 248 bi-parental QTL identified for seed protein content alone (soybase.org). Although many QTLs exist, most have yet to be confirmed (Collard et al., 2005). QTLs benefit from confirmation because they can be unstable across populations and environmental conditions. Although, environmental uncertainty can be minimized if the QTL study is conducted with multiple replications across multiple environments. Breeders can effectively target traits of interest by selecting for lines containing markers associated with QTLs. This study outlines the process of QTL analysis for multiple agronomic and seed quality traits of interest in a recombinant inbred line soybean population.

Materials and Methods

Genotyping

Genotyping was conducted by extracting genomic DNA from young trifoliate leaves using a Qiagen Plant Mini Kit (Qiagen Crawley, UK). Leaves from each RIL and both parents were collected in the first growing season prior to phenotyping. The samples

were collected in the field, placed in 2 ml Eppendorf tubes, and placed into a liquid nitrogen container. These tubes were then placed in a freezer set at -80^0 C. Extractions were done using the protocol included in the kit, apart from including vigorous vortexing after adding AP1 buffer and RNase to the samples. The genomic DNA that was extracted was analyzed with a Fisher Scientific Nanodrop using the dsDNA function to find its concentration. Concentrations over 100 microliters were kept and imaged on an electrophoresis gel. Genomic DNA that looked clean in the gel imaging was sent to the USDA Soybean Genomics and Improvement lab located in Beltsville, MD for genotyping with a 6k BARCSoy SNP chip (Illumina).

GenomeStudio 2.0

GenomeStudio 2.0 software was used to call SNPs and exclude any markers when necessary. After genotyping was completed using the 6k BARCSoy SNP chip, data was compiled by Dr. Qijian Song and sent back to Knoxville in the form of a GenomeStudio 2.0 ZIP file. From here, the GenomeStudio 2.0 project was opened and the auto-called SNPs were revealed. Although the software is mostly accurate, each of the 6000 markers needed to be checked for accuracy. An example of a bad SNP call is shown in Figure 3.1 (Appendix B). Each black dot represents a recombinant inbred line that could not be accurately called for that marker. Grey dots represent lines that failed at all 6000 SNPs on the SNP chip. Depending on how badly the SNP was called, it could either be adjusted or completely excluded. A SNP was considered salvageable if most of the the black dots could be fit into homozygous dominant (red region), heterozygous (pink region), or homozygous recessive (blue region). If more than a couple black dots remained

unaccounted for, then the SNP was "excluded" or removed from further use in the experiment. An example of a correctly called SNP is shown in Figure 3.2 (Appendix B). Out of 6000 SNPs, 63 needed to be excluded, leaving 5937 markers for linkage mapping.

Linkage Mapping

Linkage mapping was done using JoinMap 4.1. The population was entered into JoinMap as an RI6, due to the population being at the $F_{4:6}$ stage of development during genotyping. Genotypic data was arranged in Microsoft Excel and then converted to a .txt file that was entered into the program. 5937 individual loci were retained from the original 6000. Loci were then doubled using a method called "dummy coding" which takes each marker and converts it to the exact opposite allele call and then denotes the marker as "marker***x***_1" compared to the original "marker_1". The dummy coding was done using the program EMeditor. After dummy coding, the JoinMap project had 11,874 loci to be used. Genotype frequencies were then calculated for each locus. Any locus that was insignificant (> *** or 0.01) was excluded from the mapping program. Similarity of loci and individuals was then calculated. Any loci or individuals that were an exact match to one another (1.0) were excluded from the program. Finally, grouping trees were made for each of the 20 chromosomes, adjusting for LOD threshold values. The correct value was reached when most of the loci in a tree corresponded to the same chromosome. 20 trees were made in total. Maps were then created that revealed where each marker's position was on a chromosome. This position was given in centimorgans.

Microsoft Excel

After determining marker positions on each of the 20 chromosomes, the data was exported to Microsoft Excel for creation of heatmaps. The heatmap code was created by Dr. Bode Olukolu and uses recombination frequencies and LOD values for each locus in the map. The heatmaps would create a mosaic ranging from an excellent fit (red) to a moderate fit (blue). The heatmap is designed to have the colors group with one another, flowing from red to orange, to yellow, to green, and finally to blue. Any colors that are out of place are likely caused by incorrect marker positioning on a linkage map. If a marker was out of place, it was excluded from JoinMap, and a new linkage map was calculated. An example of an unedited heatmap for chromosome 18 is included with Figure 3.3 (Appendix B). A modified heatmap after the removal of incorrect markers is included in Figure 3.4 (Appendix B). After the 20 heatmaps were finished, 638 of the original 11,874 markers remained across the 20 chromosomes.

WinQTLCart2.5

Once accurate linkage maps were generated for each chromosome, five files were generated for input into WinQTLCart2.5. The files included a chromosome, label, position, phenotype, and genotype file. The "chromosome" file listed how many markers were in each chromosome's linkage map. The "label" file listed the marker names in each chromosome. The "position" file listed the centimorgan positions of each of the markers on each chromosome. The "phenotype" file listed values for traits of interest for each of the 180 RILs in sequential order. Finally, the "genotype" file listed SNP calls for each of the markers present in the linkage maps. All the files were generated in Microsoft Excel

and then converted to .txt format for input into WinQTLCart2.5. These files comprised the project named "PROLINR1". Inside of this program, composite interval mapping (CIM) was selected. LOD threshold was determined by 1000 permutations at 0.05 significance threshold and a walking speed of 1cM. The permutations automatically updated to the program as it finished running. Controls for CIM were Model 6 "Standard", 5 control markers, forward regression, and a window size of 10 cM.

Results and Discussion

The 21 QTL that were found span 7 different chromosomes, including chromosome 1, 3, 5, 7, 14, 18, and 19 (Table 3.1 Appendix B). A DAP QTL, dap-1, was found on chromosome 19 at 0.0 cM. This QTL is associated with marker Gm19_3343257_G_T and has a LOD score of 5.0. Its flanking marker was Gm19_34840388_C_A. Dap-1 had an r^2 value of 11.7, and its additive effect is 0.6. TN13-4303 was associated with positive additive affects, and TN12-4098 with the negative effects. Dap-1 is in the same area as previously discovered QTL "Pod Maturity 24-3," which is associated with marker Satt495 (Bachlava et al., 2009). Two lodging QTL were discovered, both located on chromosome 19. The first lodging QTL, lodg-1, (marker Gm19_36641660_G_A) is located at 12.4 cM and has a LOD score of 5.5 with an additive effect of -0.1. This QTL has flanking markers of Gm19_35744912_C_T and Gm19_36780878_G_T. Lodg-1 is in the same cM window as "Lodging 5-11" associated with marker EV2_1 (Lee et al., 1996). The second lodging QTL, lodg-2, on chromosome 19 is located at 28.5 cM and is associated with marker Gm19_37631304_T_G. The flanking markers for this QTL include Gm19_36780878_G_T and Gmx19_39433067_C_T. Lodg-2 has an additive

effect of -0.3 and is in the same cM region as “Lodging 28-5,” which is associated with marker BARC-041643-08051 (Lee et al., 2015). One plant height QTL, hgt-1, associated with marker Gm19_40219547_C_T, was found at 41.1 cM. The flanking markers for this QTL include Gm19_40053178_G_A and Gmx19_40508288_C_T. Hgt-1 had an r^2 value of 31.5 and an additive effect of -11.1. No previously discovered QTL was located at this region, making this a “novel” locus. One yield QTL, yld-1, associated with marker Gm14_2597934_A_G with flanking markers at Gm14_2480875_A_C and Gm14_2762103_T_C, was found at 15.9 cM on chromosome 14. This QTL has an additive effect of 104.7 and is novel. Two protein QTL were found on chromosome 1. The first QTL, pro-1, at marker Gm01_1887205_G_A was found at 20.1 cM, with flanking markers of Gm01_1744951_C_A and Gm01_1936523_T_C. This QTL has an additive effect of 2.0 and is novel. The second QTL, pro-2, associated with Gmx01_1428598_T_G was found at 24.2 cM with flanking markers at Gmx01_1344976_A_G and Gm01_1502816_G_A. Pro-2 has an additive effect of 1.9 and is novel. Two protein QTL were found on chromosome 3. The first QTL, pro-3, associated with marker Gm03_44118764_C_T was found at 17.4 cM. This QTL was flanked by Gm03_44019995_G_A and Gmx03_44171693_A_C. Pro-3 has an additive effect of -2.1 and is found in the same region as “cqSeed protein-010” associated with marker BARC-055149-13089 (Pathan et al., 2013). The second protein QTL, pro-4, on chromosome 3 is associated with marker Gmx03_43599557_T_C and is found at 22.2 cM. Flanking markers for this QTL include Gmx03_43355787_C_T and Gmx03_43707104_A_G. Pro-4 has an additive effect of -1.9 and is also found in the

same area as “cqSeed protein-010”. Chromosome 7 had a protein QTL associated with marker Gm07_35194991_A_G at 0.0 cM, and had one flanking marker, Gm07_35142318_A_G. This QTL, pro-5, had an additive effect of -1.9 and was novel. Chromosome 14 contained a protein QTL, pro-6, near marker Gm14_41187024_T_C at 71.7 cM. Flanking markers for this QTL included Gm14_31348204_T_C and Gm14_46289371_C_T. Pro-6 has an additive effect of 3.3 and is in the same cM range as “Seed protein 39-1,” which is located at BARC-056587-14511 (Warrington et al., 2015). An oil QTL on chromosome 1 associated with marker Gmx01_3016694_A_G was found at 13.8 cM, and had flanking markers of Gm01_2708722_C_T and Gmx01_3063603_T_G. This QTL is oil-1 and had a -1.9 additive effect and was considered novel. Chromosome 3 had an oil QTL, oil-2, at 0.0 cM associated with marker Gm03_47320906_C_T and had one flanking marker at Gmx03_47039930_T_C. This QTL was novel and had an additive effect of -1.3. A third oil QTL, oil-3, was located on chromosome 14 at 33.5 cM. Oil-3 was closest to marker Gmx14_5347242_G_A and had an additive effect of 1.3. This QTL was flanked by markers Gmx14_4889916_T_C and Gm14_6829154_T_C. Oil-3 aligned with previously discovered QTL “Seed oil 45-6” flanked by ss248275088 and ss248293401 (Akond et al., 2014). A linolenic acid QTL, lin-1, was found on chromosome 5 at 32.2 cM. This QTL was associated with marker Gm05_34939267_A_G and had an additive effect of -0.3. Lin-1 was flanked by markers Gm05_33576968_A_G and Gmx05_35039076_C_T. This was a novel locus. A major QTL for linolenic acid, lin-2, was found on chromosome 14 at 78.7 cM. This QTL was associated with marker

Gmx14_5347242_G_A and had an R^2 value of 56.3. Lin-2 had flanking markers of Gmx14_4889916_T_C and Gm14_6829154_T_C. This locus has an additive effect of 0.9 and aligns with "Seed linolenic 7-6 at Satt066 (Bachlava et al., 2009). Chromosome 18 contained a linolenic acid QTL, lin-3, at 53.3 cM associated with marker Gmx18_8777288_A_G. Flanking markers for this QTL included Gm18_5522831_A_C and Gmx18_9678773_T_C. This is a novel QTL with an additive effect of 0.5. LOD thresholds were 3.1, 3.0, 3.0, 3.0, 3.0, 2.9, 3.1, and 3.1 for protein, oil, meal protein, linolenic acid, yield, DAP, lodging, and height, respectively. A mix of major and minor QTL were found in this study. One QTL of interest is the linolenic acid QTL on chromosome 14 associated with marker Gm14_46289371_C_T, which had an r^2 value of 56.3. This means that 56.3% of the variation in the linolenic acid content of the population could be explained by this locus. This QTL could contain the known gene "FAD3A", which is responsible for controlling linolenic acid content (Bilyeu et al., 2003). The QTL for linolenic acid located on chromosome 5 could be extremely useful to breeders. The presence of this QTL resulted in a lower linolenic acid level in the mapping population. Breeders could target soybeans containing this QTL and pair them with lines containing mutant FAD genes (Bilyeu et al., 2003; Bilyeu et al., 2005., Held et al., 2019) for extremely low linolenic acid lines that maintain stable linolenic acid levels across multiple environments.

As previously mentioned, some of the QTL found in this study overlap with QTL recorded in SoyBase. QTL can be volatile across populations and environments, so much so that confirmation of a QTL must be deemed valid by the Soybean Genetics Committee

(Fox et al., 2015). Per Soybean Genetics Committee guidelines, a QTL must be confirmed in a population at an error rate of 0.01 containing meiotic events separate from the original population (Fox et al., 2015). Although the QTL in this study have not received an official confirmation from the Soybean Genetics Committee, they can instill confidence in breeders to use the overlapped loci for MAS.

REFERENCES

Akond, M., Liu, S., Boney, M., Kantartzi, S. K., Meksem, K., Bellaloui, N., ... & Kassem, M. A. (2014) Identification of quantitative trait loci (QTL) underlying protein, oil, and five major fatty acids' contents in soybean. *American Journal of Plant Sciences*, **5**: 2014.

Bachlava, E., Dewey, R. E., Burton, J. W., & Cardinal, A. J. (2009) Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. *Crop Science*, **49**: 433-442.

Bilyeu, K. D., Palavalli, L., Sleper, D. A., & Beuselinck, P. R. (2003) Three Microsomal Omega-3 Fatty-acid Desaturase Genes Contribute to Soybean Linolenic Acid Levels. *Crop Science*, **43**:1833-1838.

Bilyeu, K., Palavalli, L., Sleper, D., & Beuselinck, P. (2005) Mutations in Soybean Microsomal Omega-3 Fatty Acid Desaturase Genes Reduce Linolenic Acid Concentration in Soybean Seeds. *Crop Science*, **45**: 1830-1836.

Collard, B. C., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica*, **142**: 169-196.

Fox, C. M., Cary, T. R., Nelson, R. L., & Diers, B. W. (2015) Confirmation of a seed yield QTL in soybean. *Crop Science*, **55**: 992-998.

Held, J. P., Carrero-Colón, M., & Hudson, K. A. (2019) Combination of Novel Mutation in FAD3C and FAD3A for Low Linolenic Acid Soybean. *Agrosystems, Geosciences & Environment*, **2:**1.

Lee, S. H., Bailey, M. A., Mian, M. A. R., Carter Jr, T. E., Ashley, D. A., Hussey, R. S., ... & Boerma, H. R. (1996) Molecular markers associated with soybean plant height, lodging, and maturity across locations. *Crop Science*, **36**: 728-735.

Lee, S., Jun, T. H., Michel, A. P., & Mian, M. R. (2015) SNP markers linked to QTL conditioning plant height, lodging, and maturity in soybean. *Euphytica*, **203**: 521-532.

Panthee, D.R., Pantalone, V.R., West, D.R., Saxton, A.M., & Sams, C.E. (2005) Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean. *Crop Science*, **45**: 2015–2022.

Pathan, S. M., Vuong, T., Clark, K., Lee, J. D., Shannon, J. G., Roberts, C. A., ... & Nguyen, H. T. (2013) Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. *Crop Science*, **53**: 765-774.

Soybase (2020) *List of QTL*. Retrieved from https://soybase.org/search/index.php?searchterm=Seed+protein&search=true

Warrington, C. V., Abdel-Haleem, H., Hyten, D. L., Cregan, P. B., Orf, J. H., Killam, A. S., ... & Boerma, H. R. (2015) QTL for seed protein and amino acids in the Benning× Danbaekkong soybean population. *Theoretical and Applied Genetics*, **128**: 839-850.

APPENDIX B

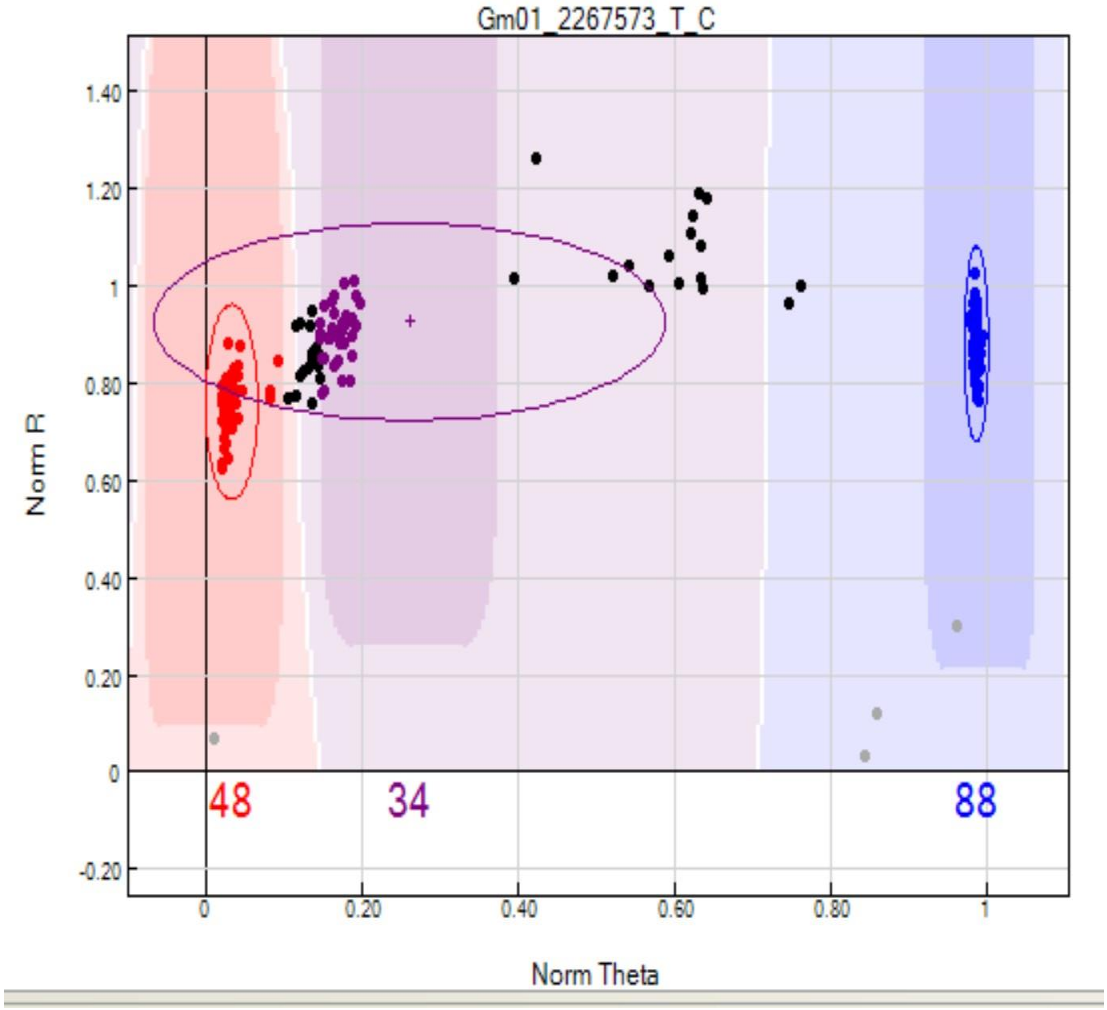

Figure 3.1: GenomeStudio 2.0 SNP called incorrectly. Black dots represent individual RILs without a call at the SNP location. Grey dots represent failed samples.

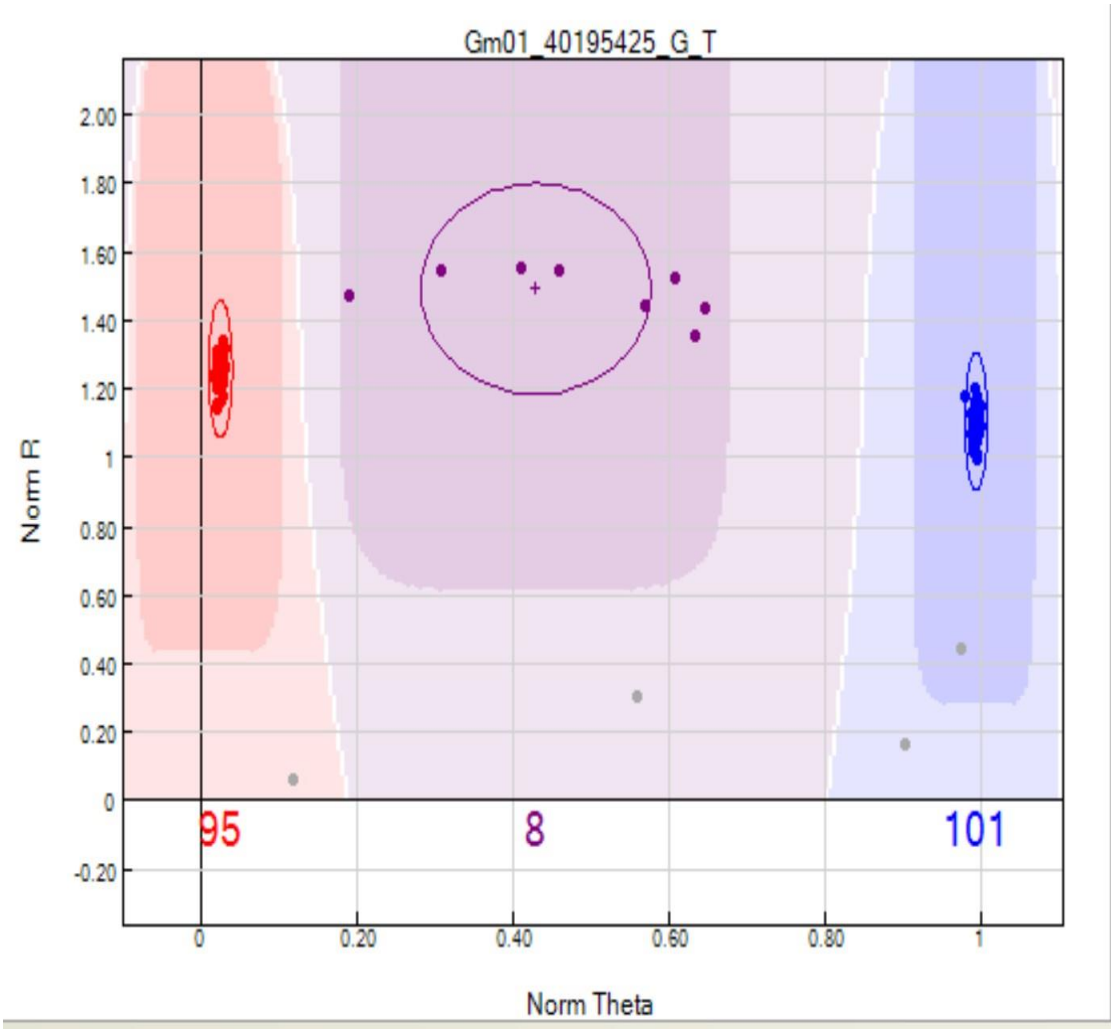

Figure 3.2: GenomeStudio 2.0 SNP called correctly. Red shading represents homozygous dominant, purple represents heterozygous, and blue represents homozygous recessive. Grey dots represent failed samples.

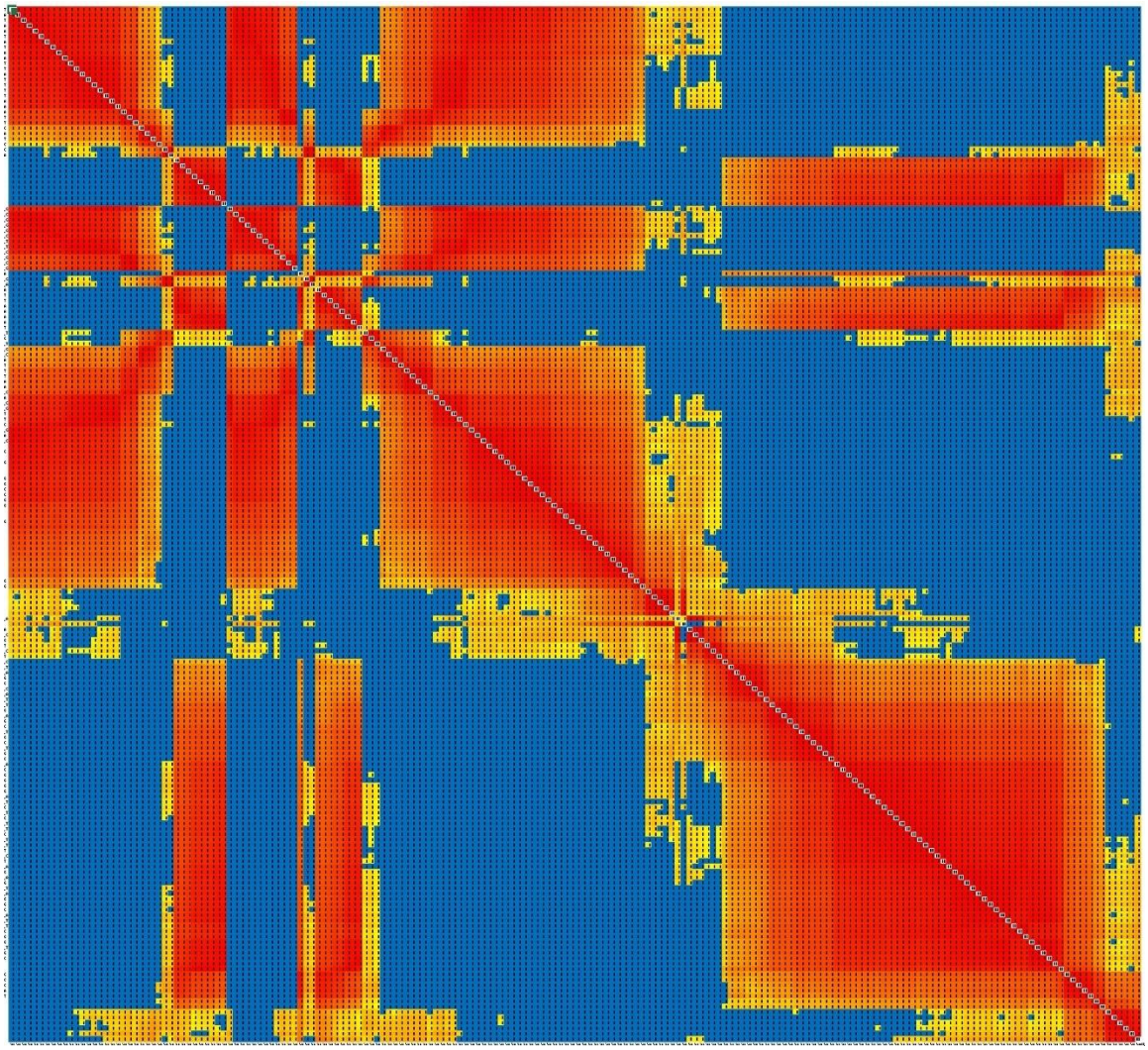
Figure 3.3: Chromosome 18 heatmap prior to revisions.

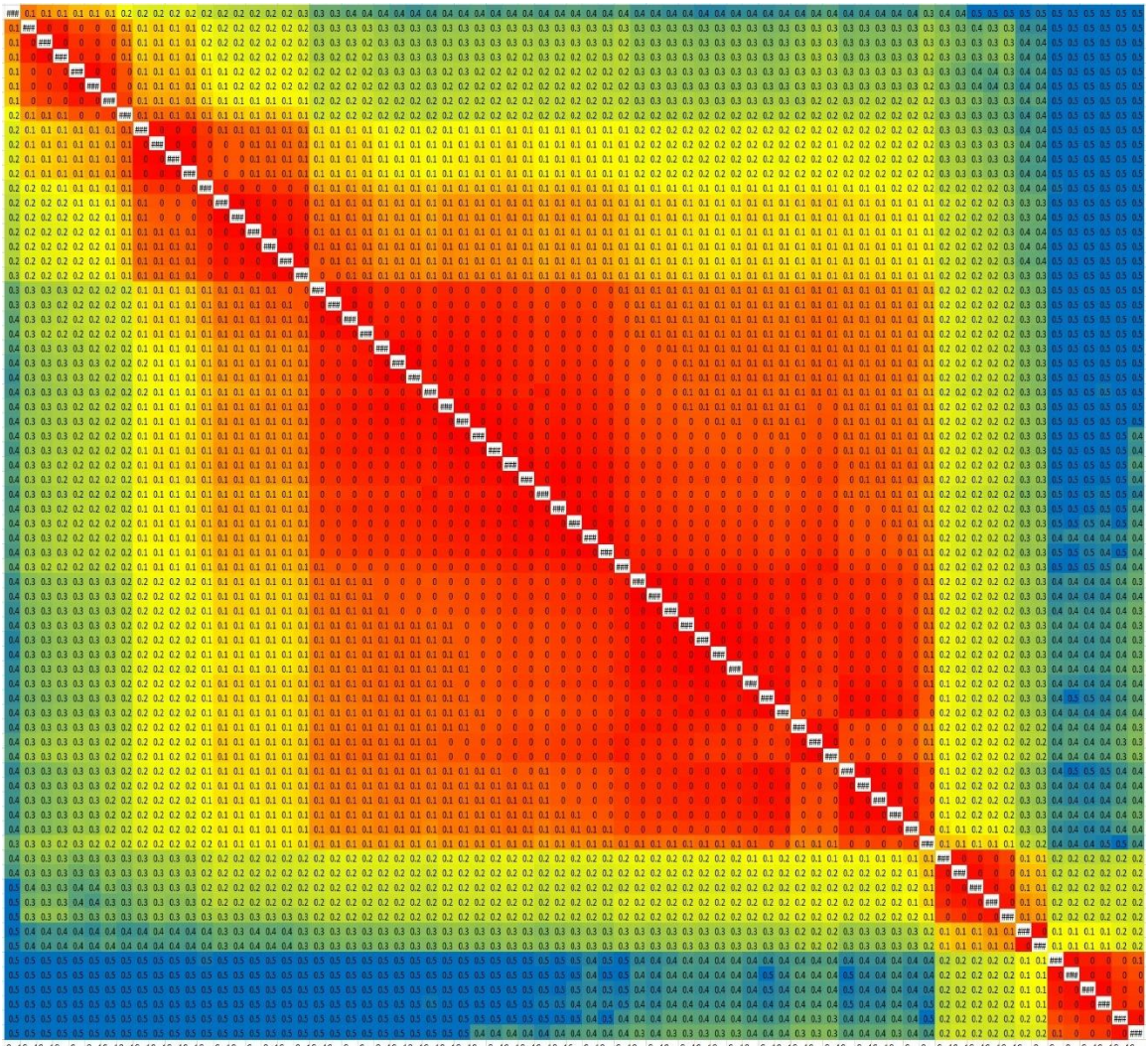

Figure 3.4: Chromosome 18 heatmap after 5 rounds of revisions.

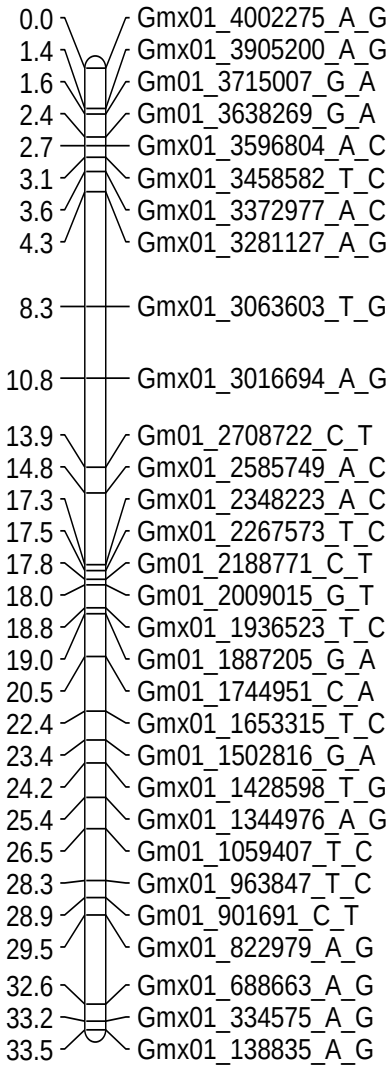

Figure 3.5: Linkage map of chromosome 1.

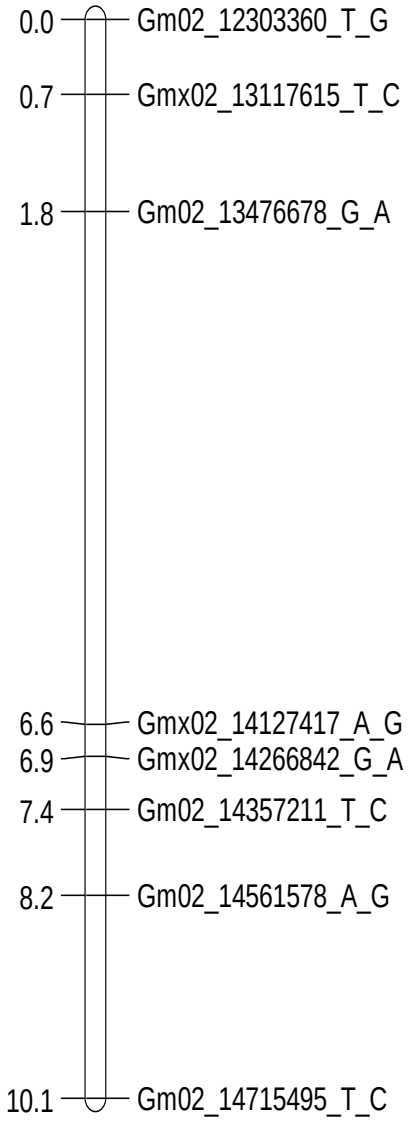

Figure 3.6: Linkage map of chromosome 2.

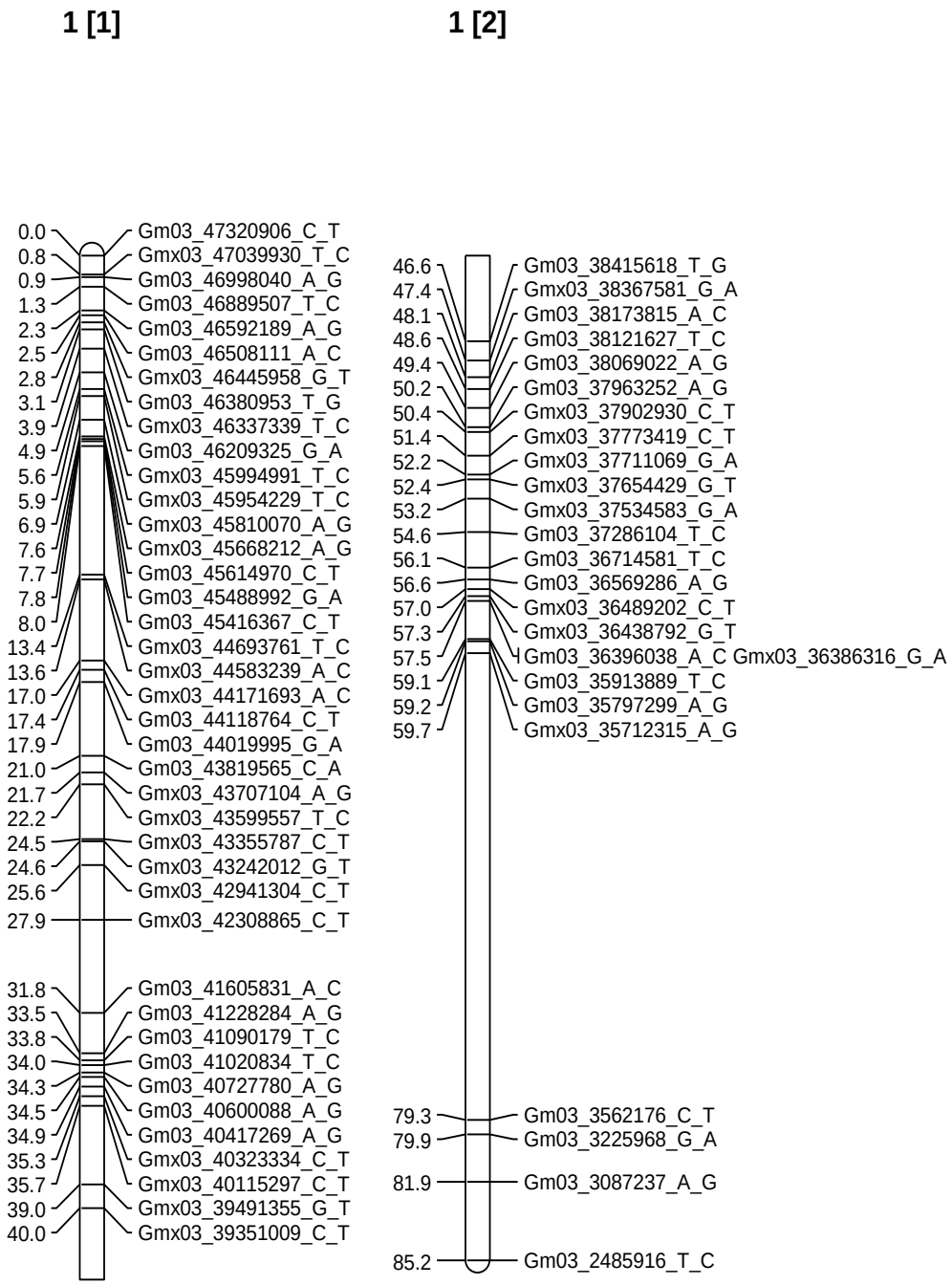

Figure 3.7: Linkage map of chromosome 3.

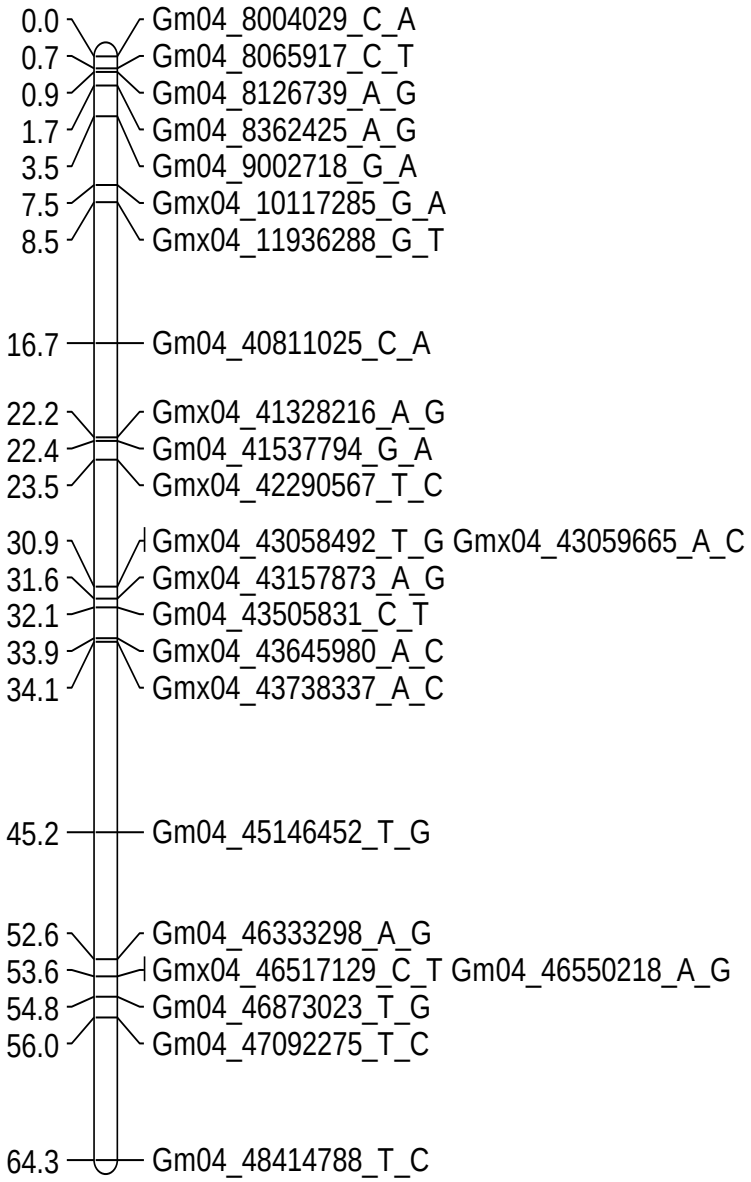

Figure 3.8: Linkage map of chromosome 4.

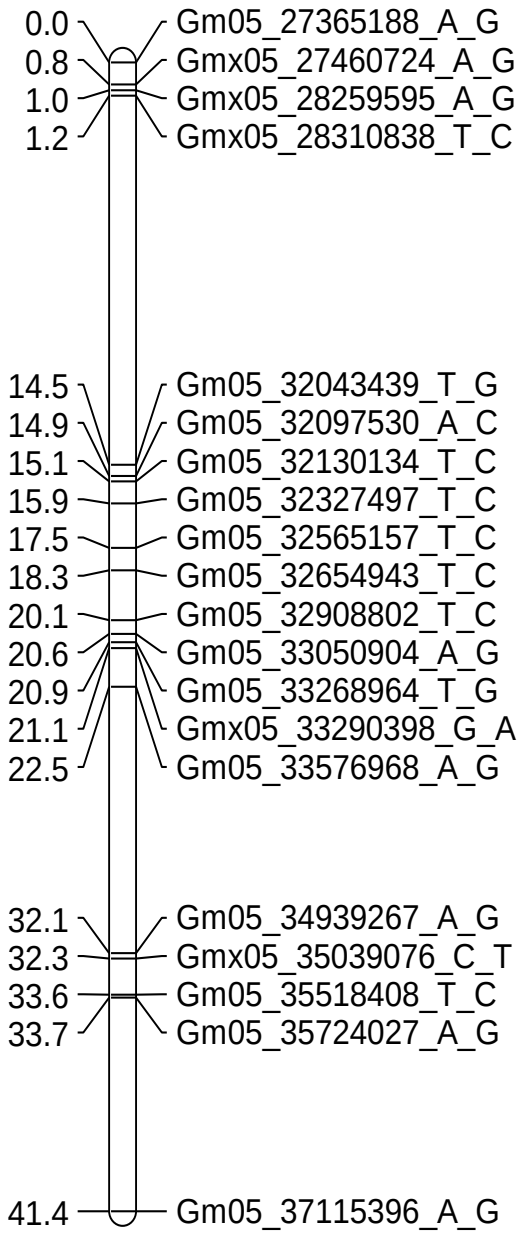

Figure 3.9: Linkage map of chromosome 5.

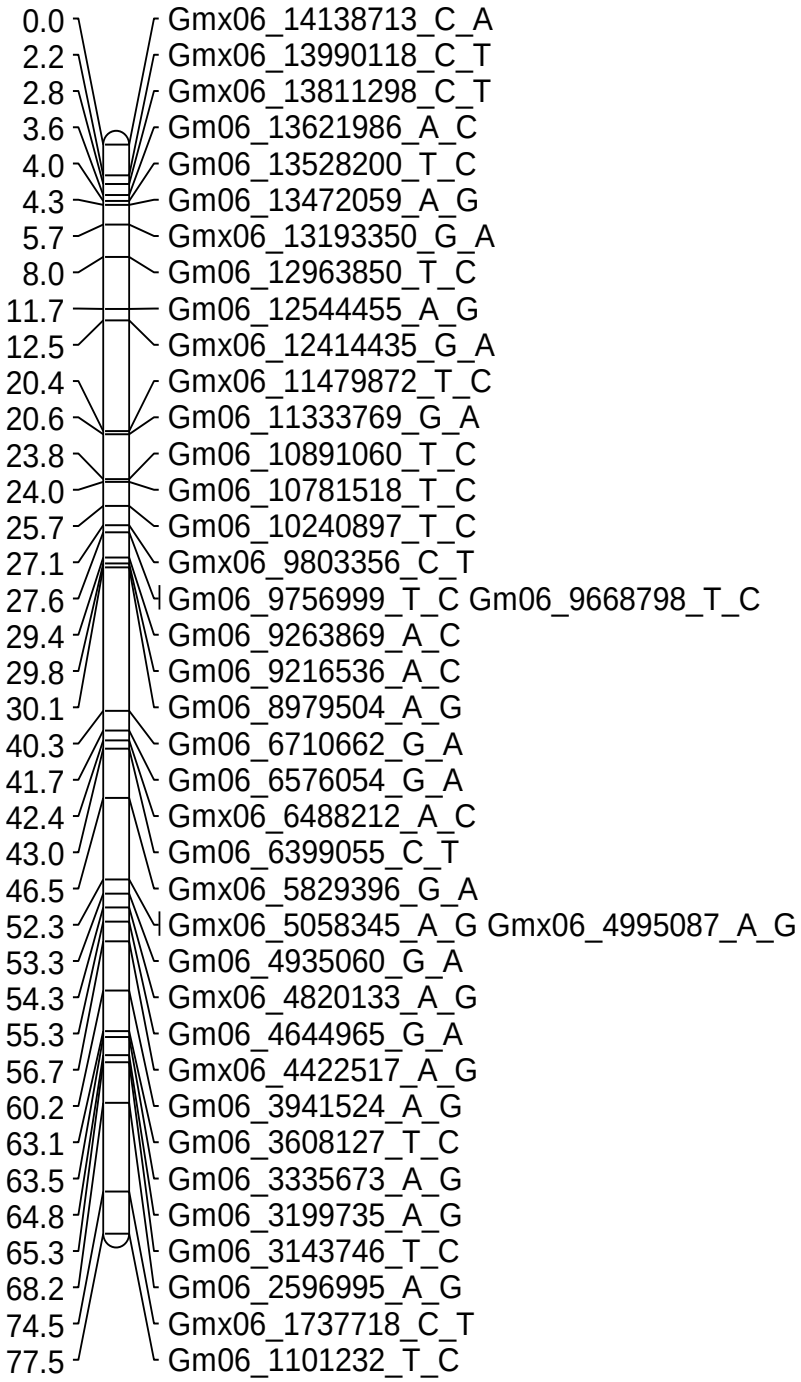

Figure 3.10: Linkage map of chromosome 6.

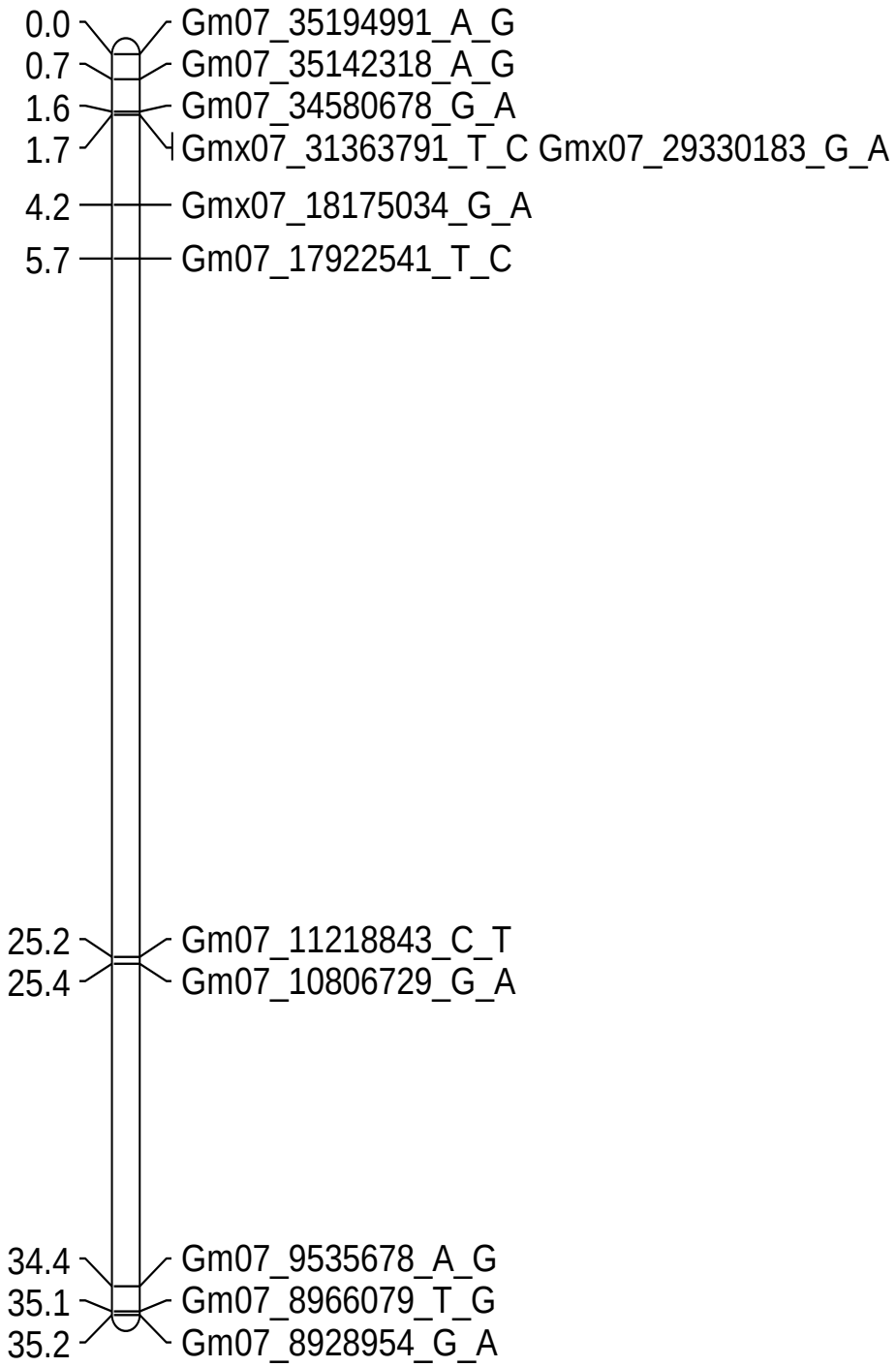

Figure 3.11: Linkage map of chromosome 7.

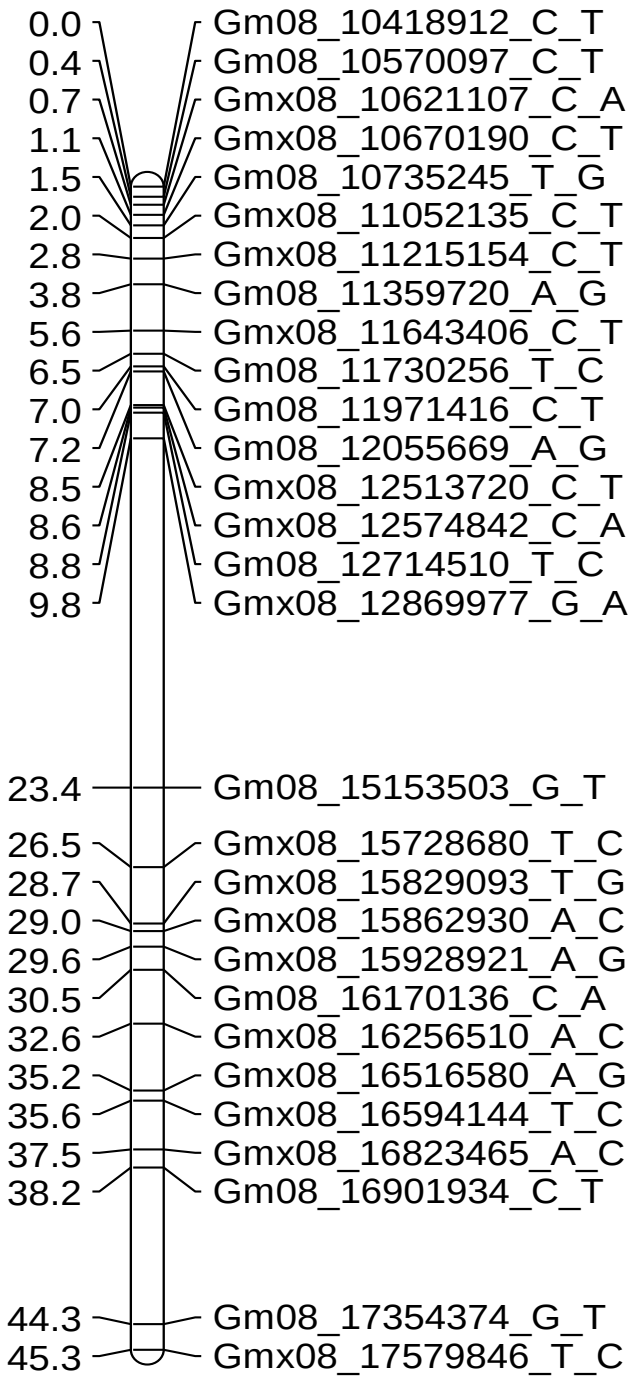

Figure 3.12: Linkage map of chromosome 8.

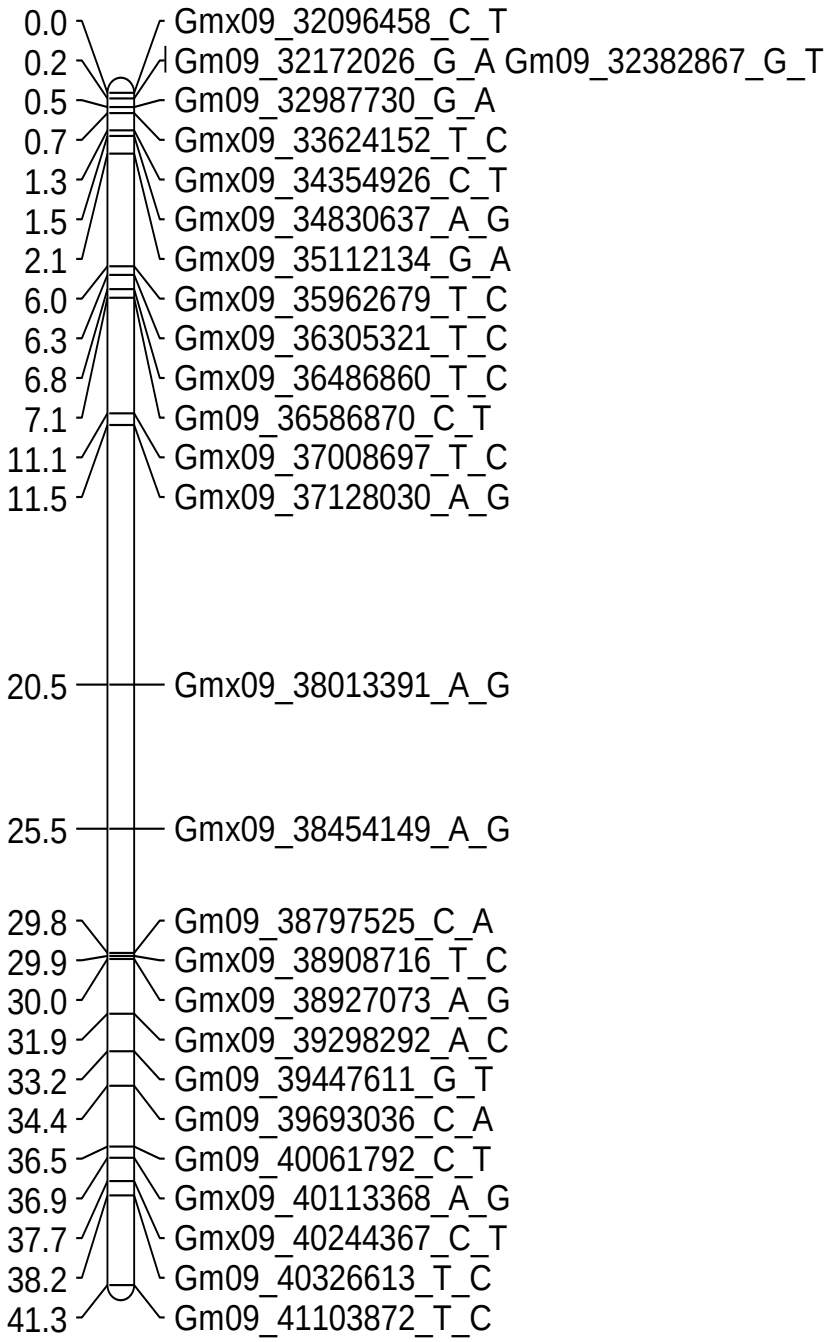

Figure 3.13: Linkage map of chromosome 9.

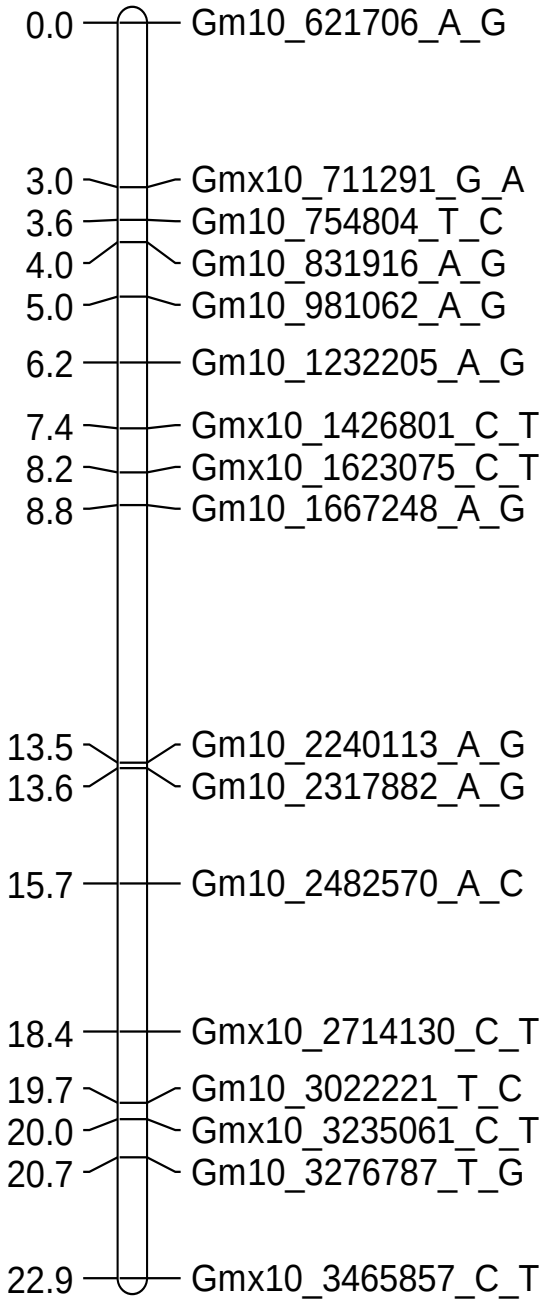

Figure 3.14: Linkage map of chromosome 10.

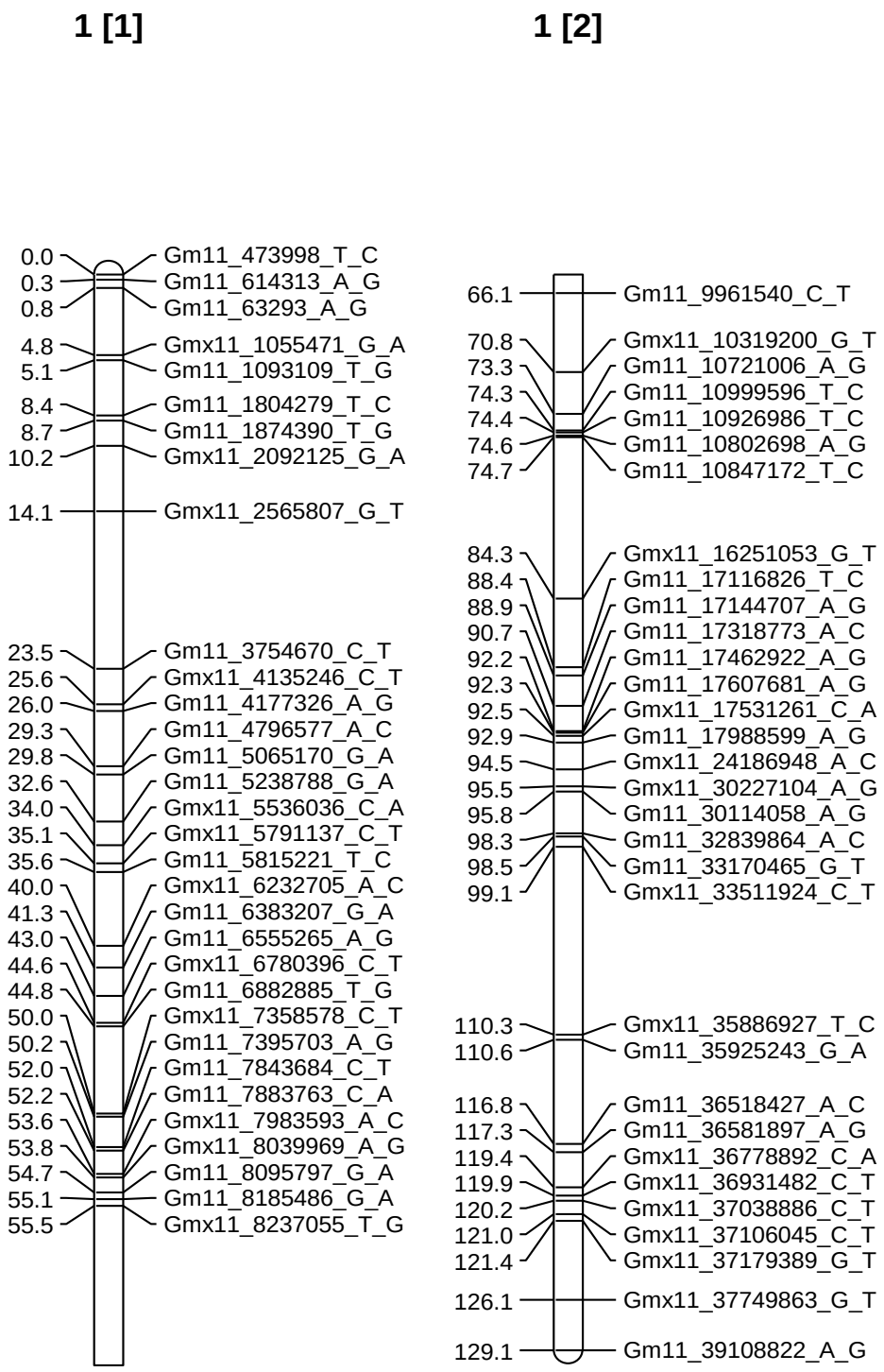

Figure 3.15: Linkage map of chromosome 11.

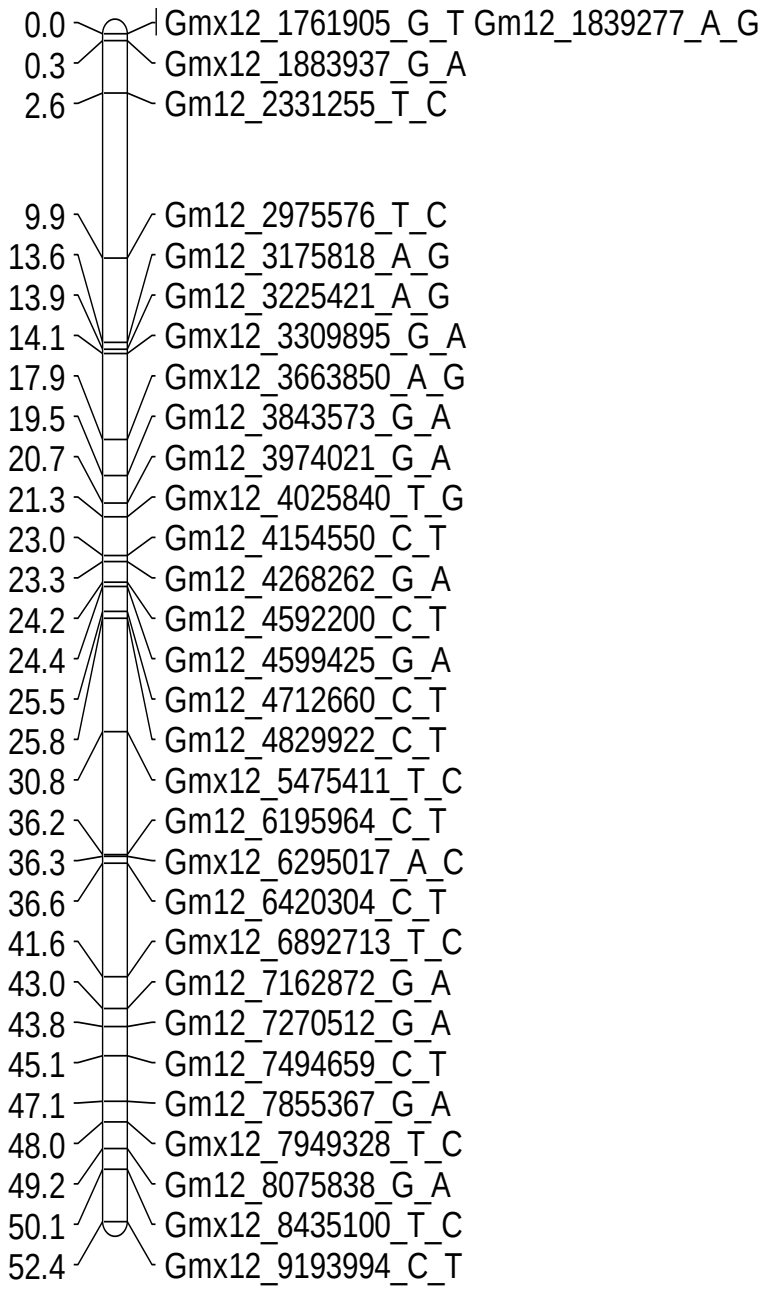

Figure 3.16: Linkage map of chromosome 12.

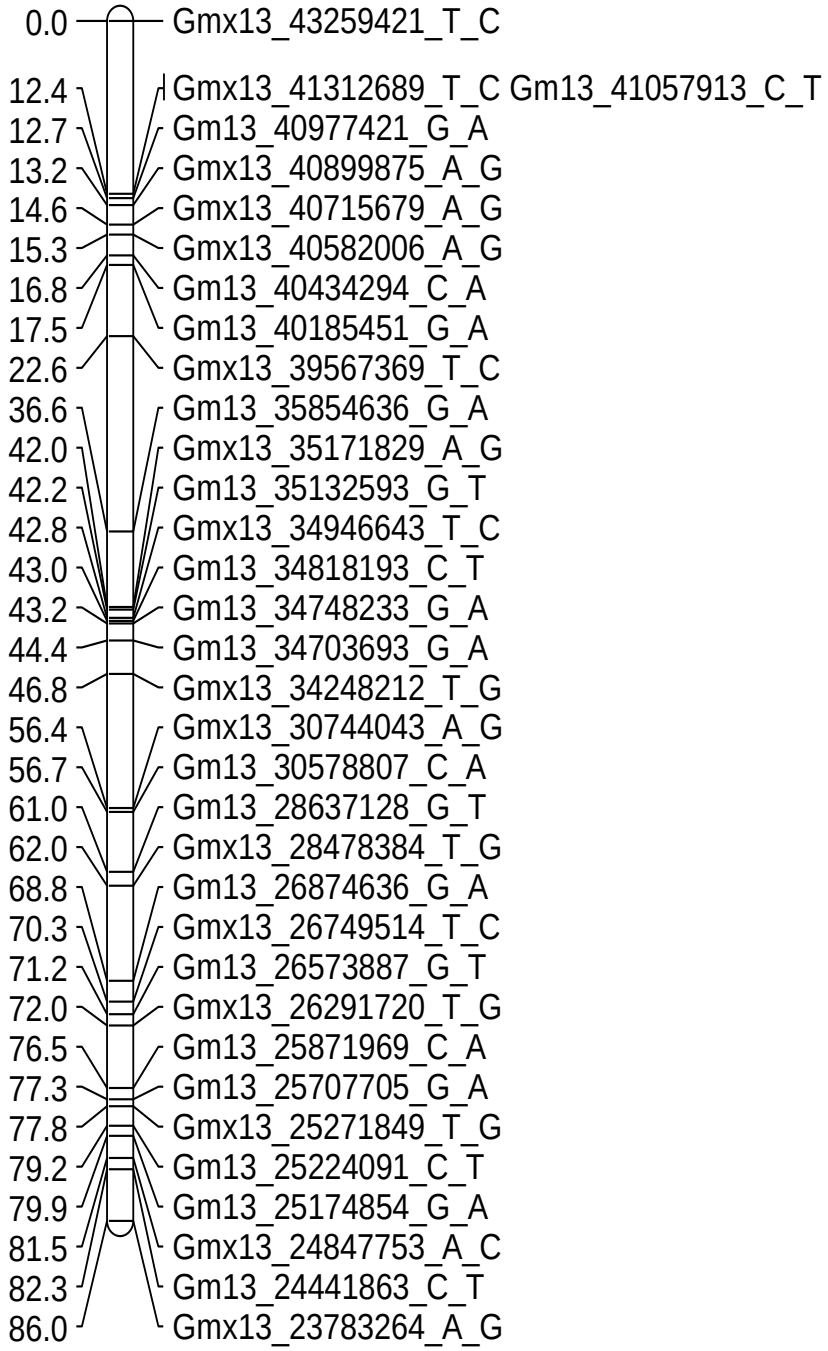

Figure 3.17: Linkage map of chromosome 13.

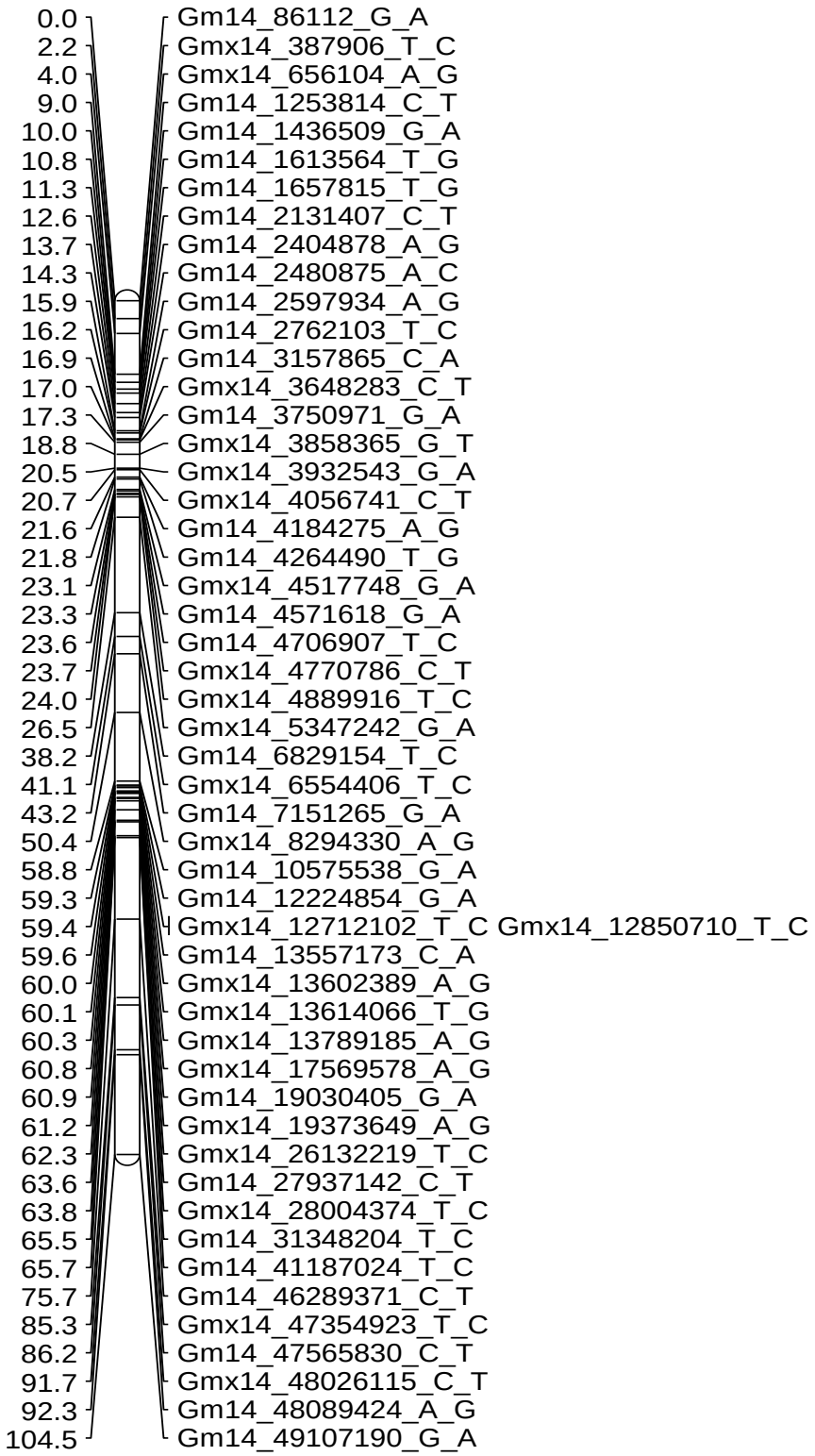

Figure 3.18: Linkage map of chromosome 14.

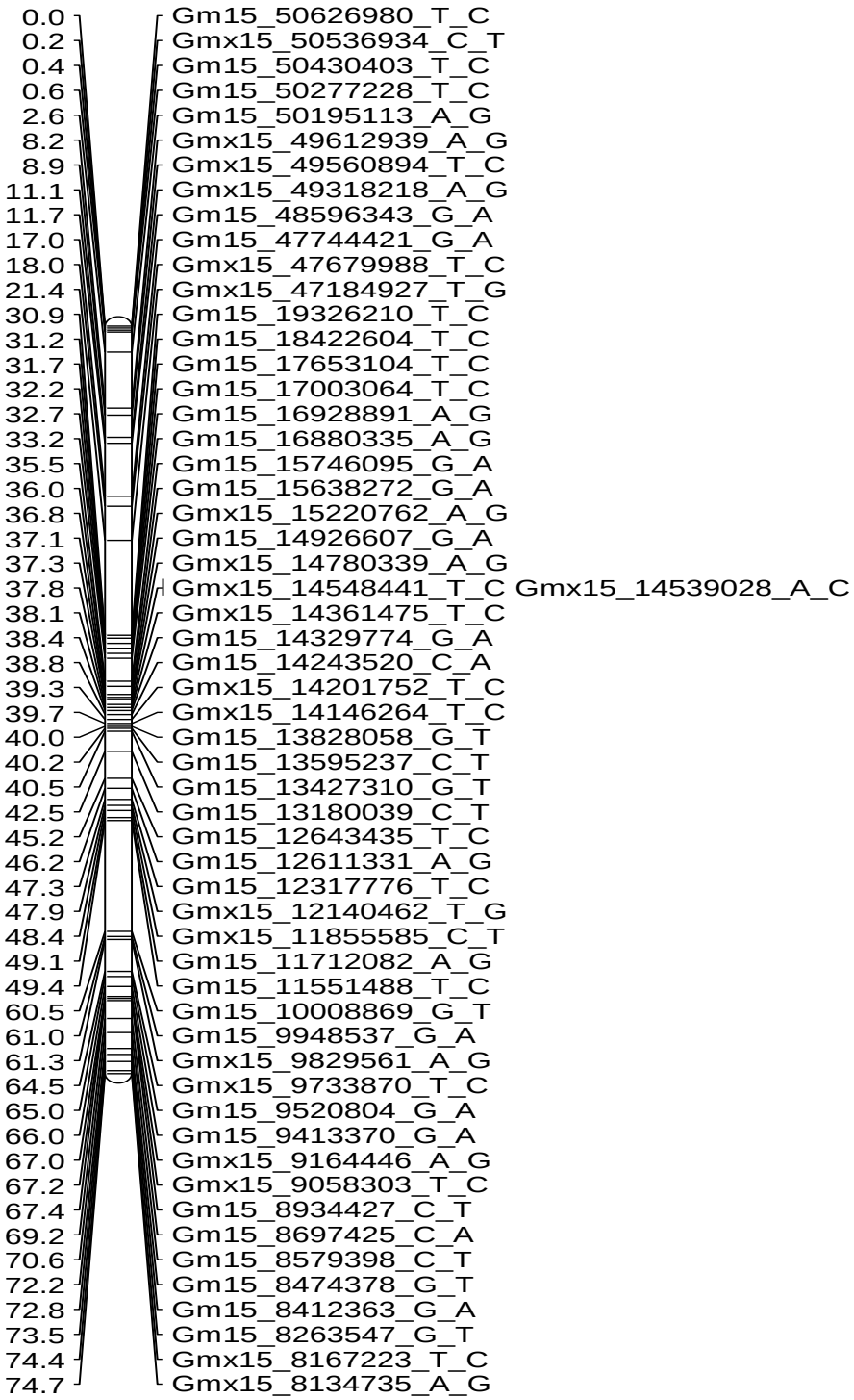

Figure 3.19: Linkage map of chromosome 15.

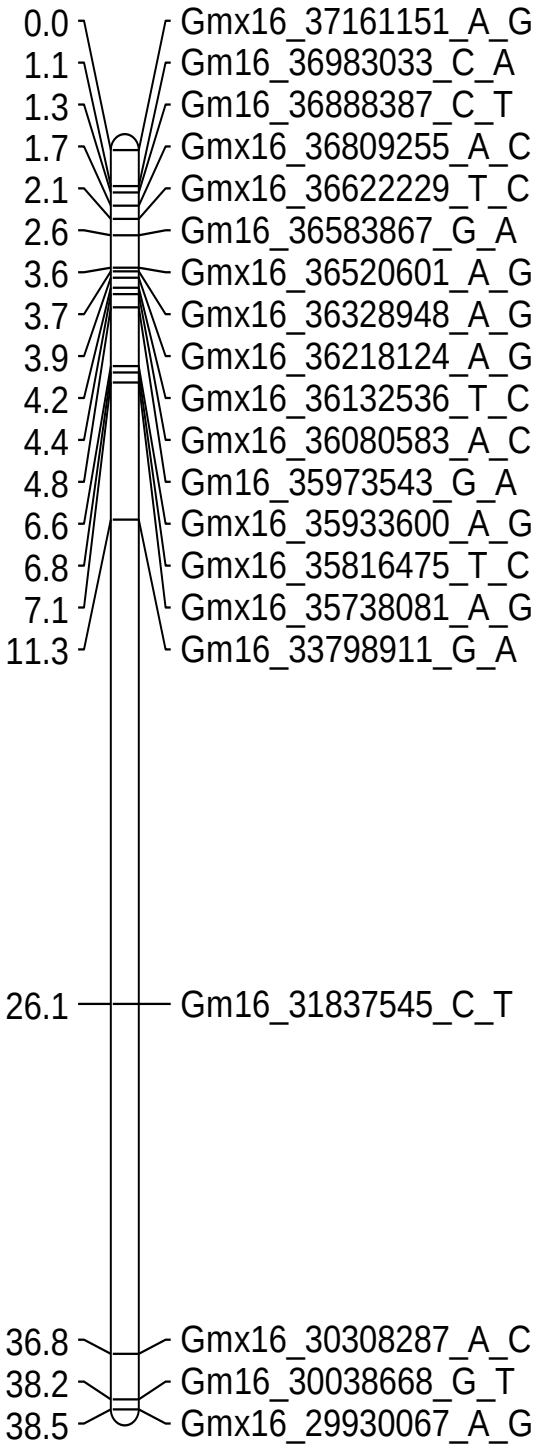

Figure 3.20: Linkage map of chromosome 16.

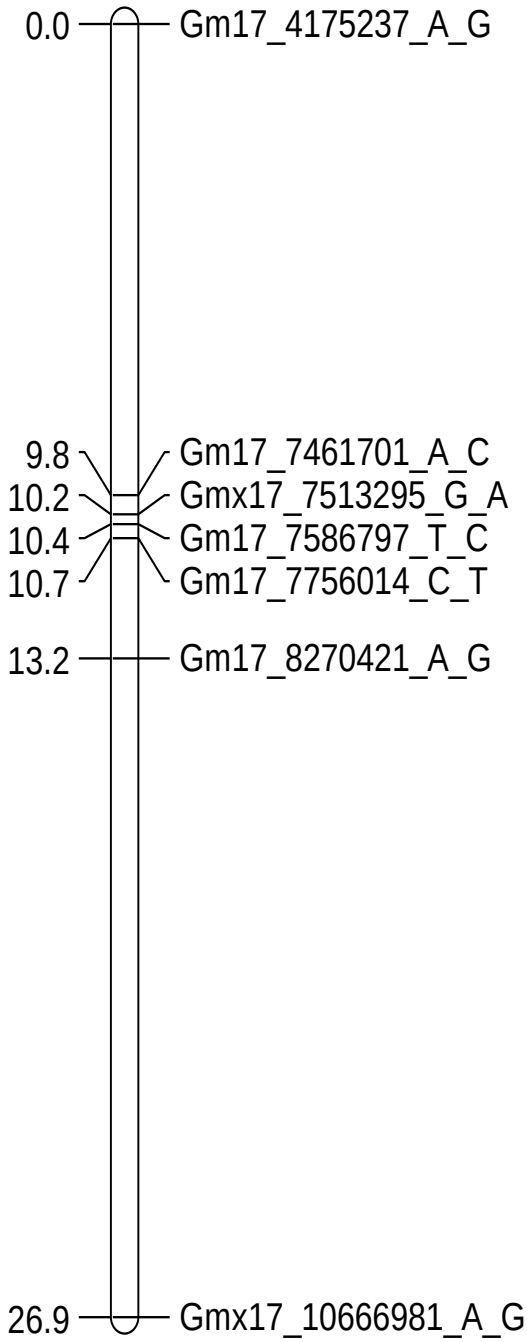

Figure 3.21: Linkage map of chromosome 17.

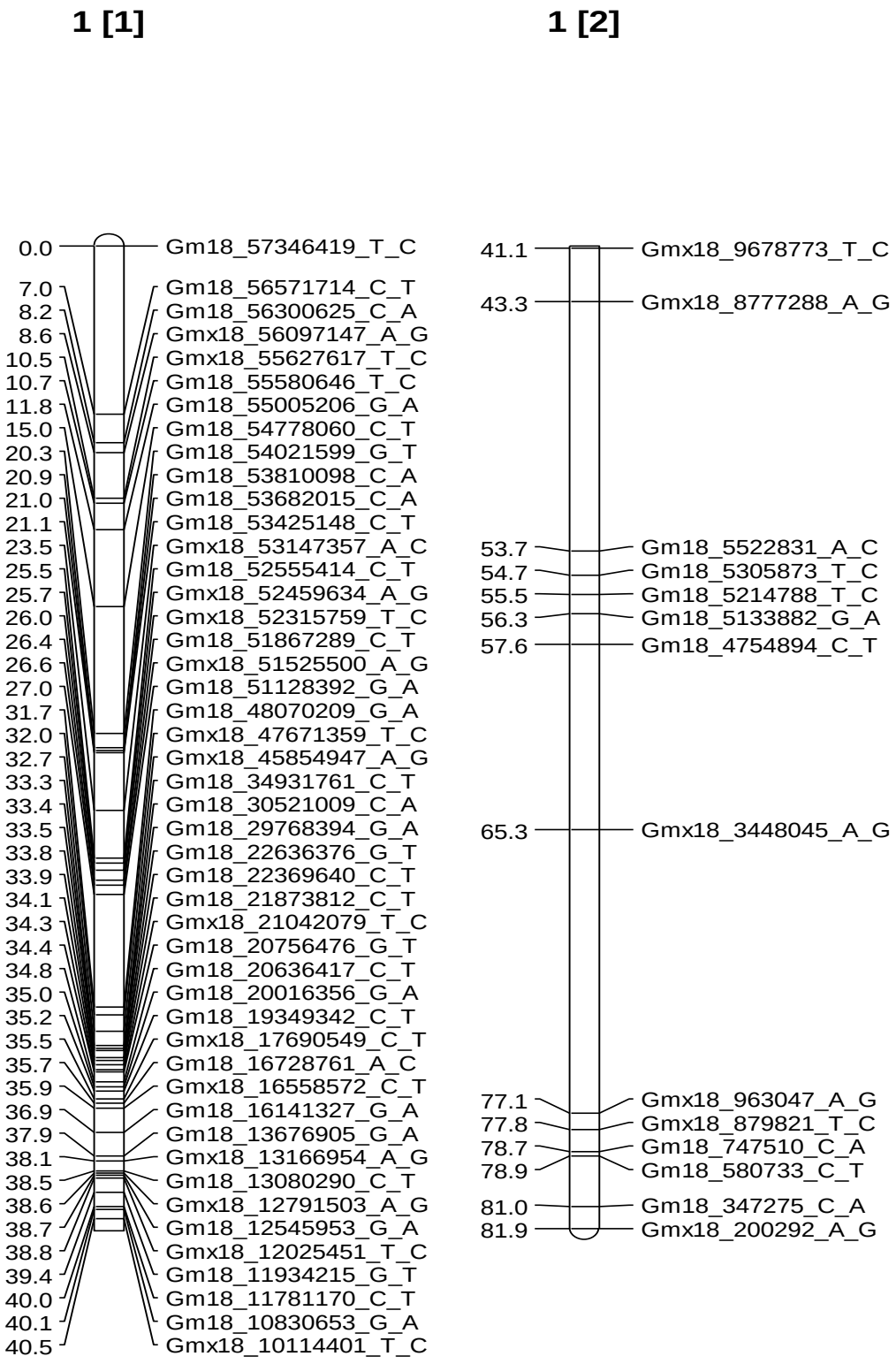

Figure 3.22: Linkage map of chromosome 18.

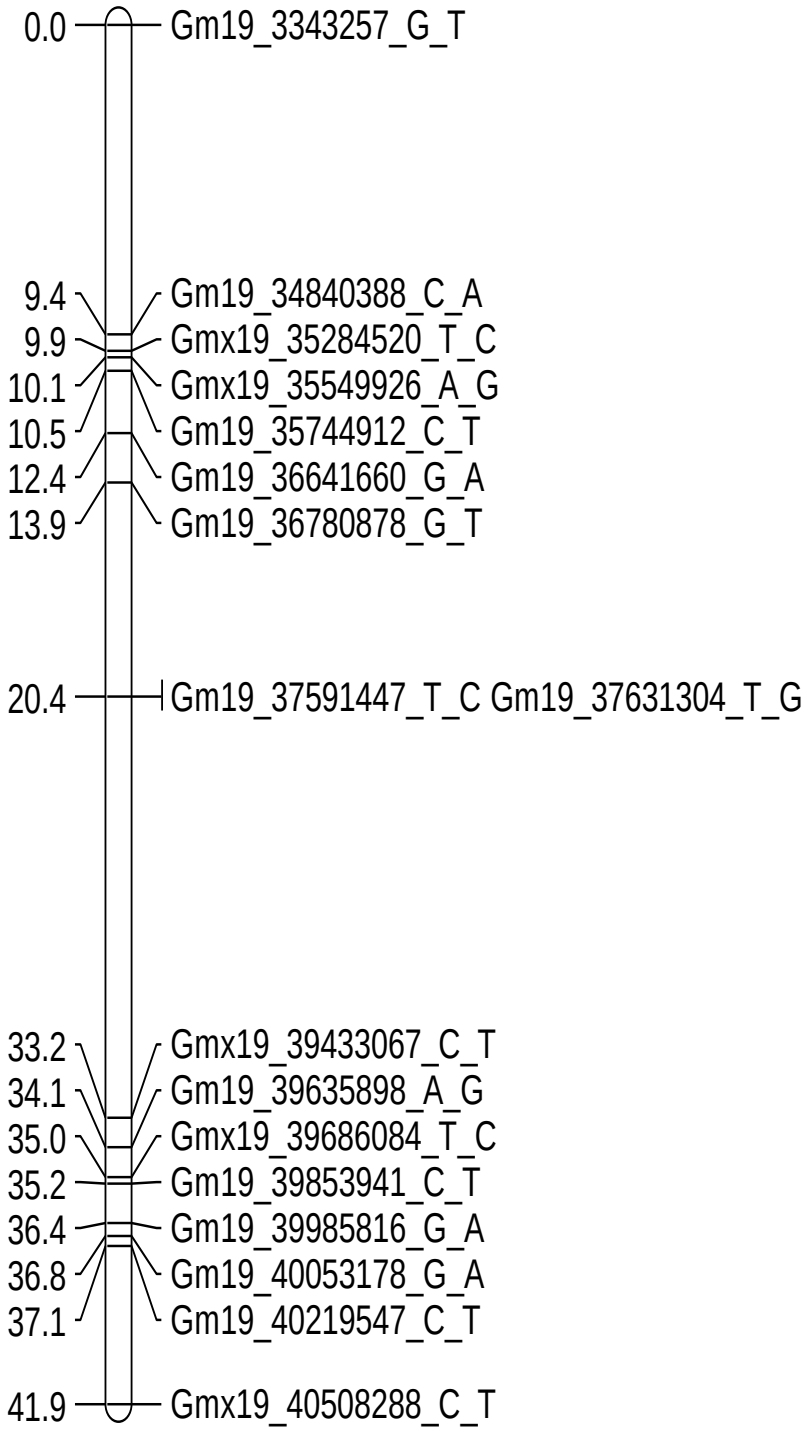

Figure 3.23: Linkage map of chromosome 19.

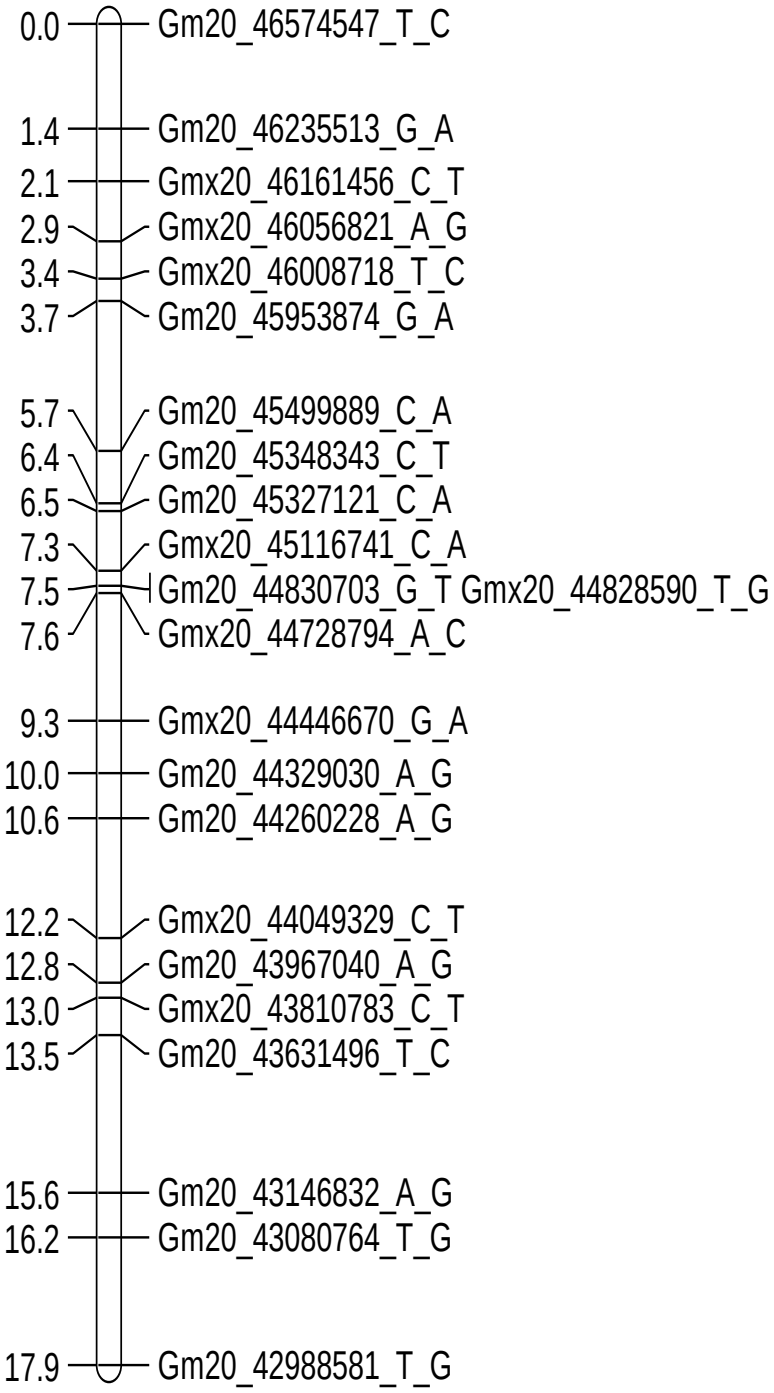

Figure 3.24: Linkage map of chromosome 20.

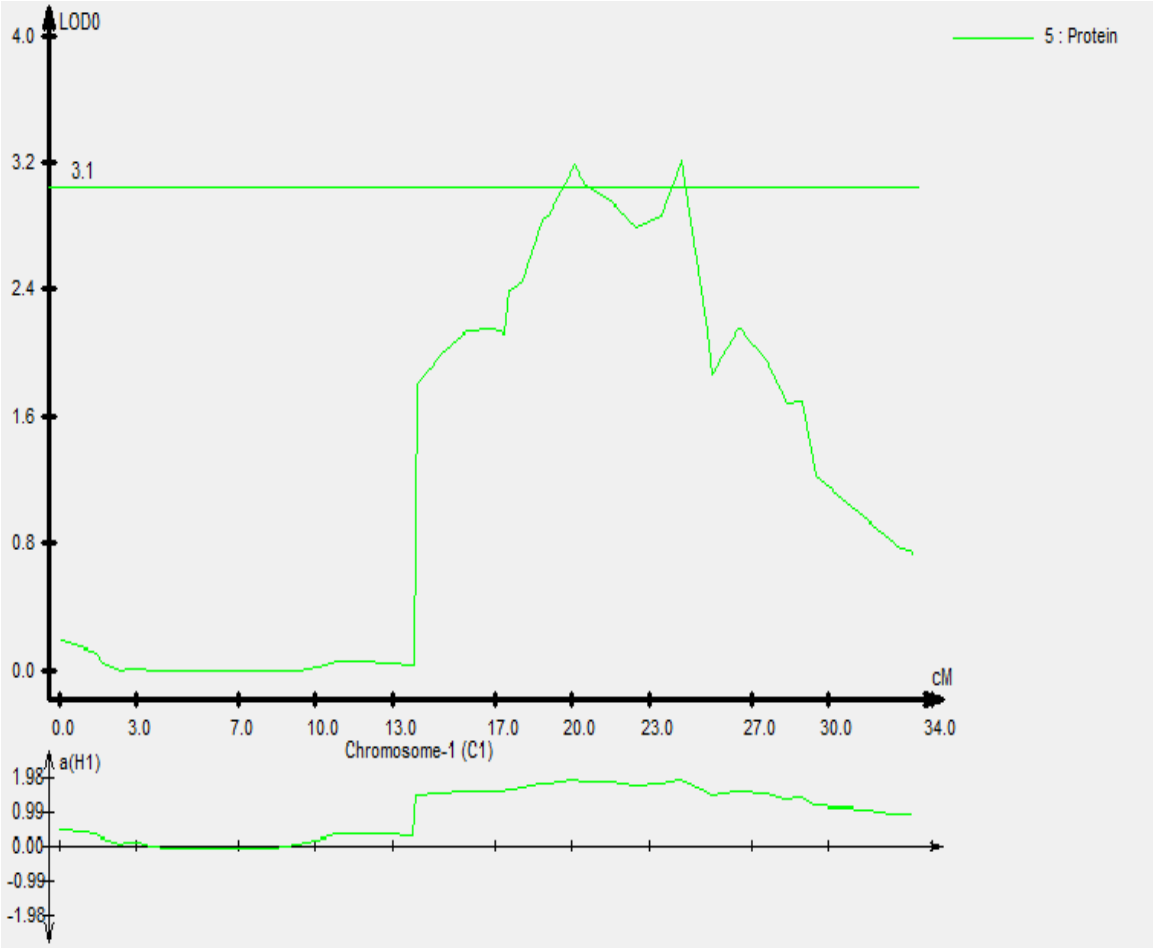

Figure 3.25: Two protein QTL located on chromosome 1. Pro-1 is located at 20.1 cM and pro-2 is located at 24.2 cM.

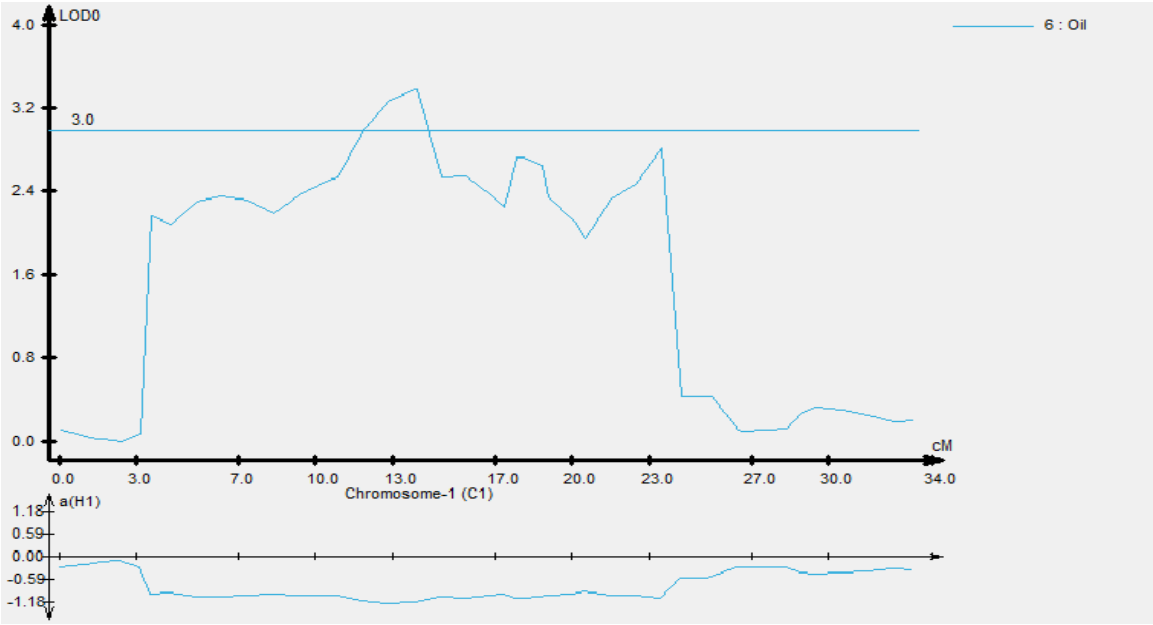

Figure 3.26: One oil QTL on chromosome 1. Oil-1 is located at 13.8 cM.

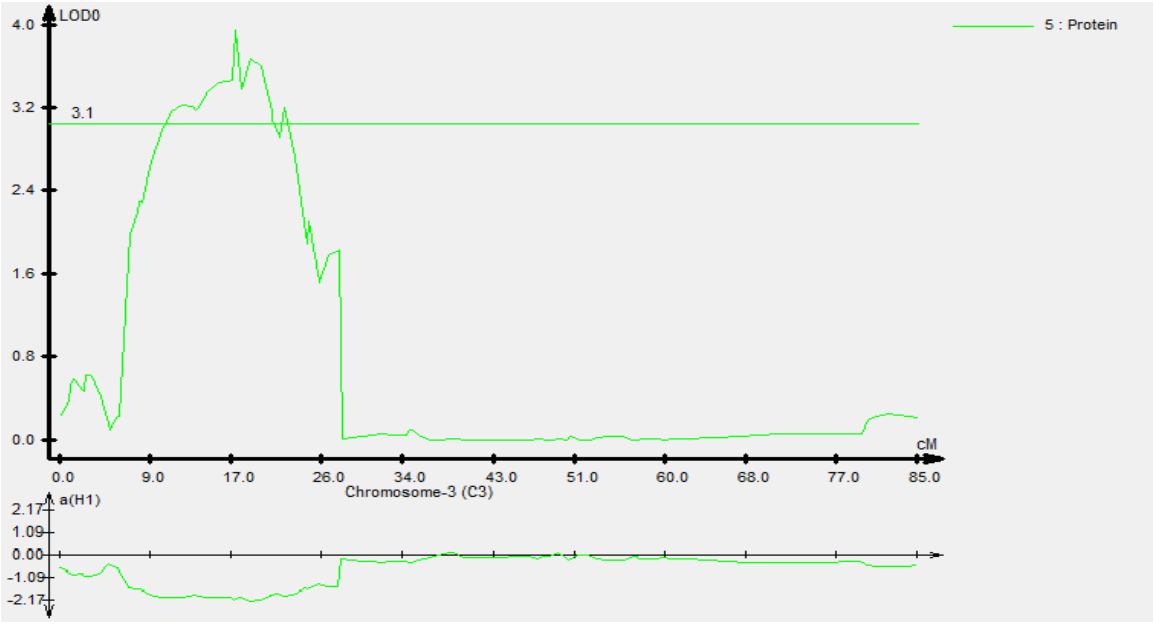

Figure 3.27: Two protein QTL on chromosome 3. Pro-3 is located at 17.4 cM and pro-4 is located at 22.2 cM.

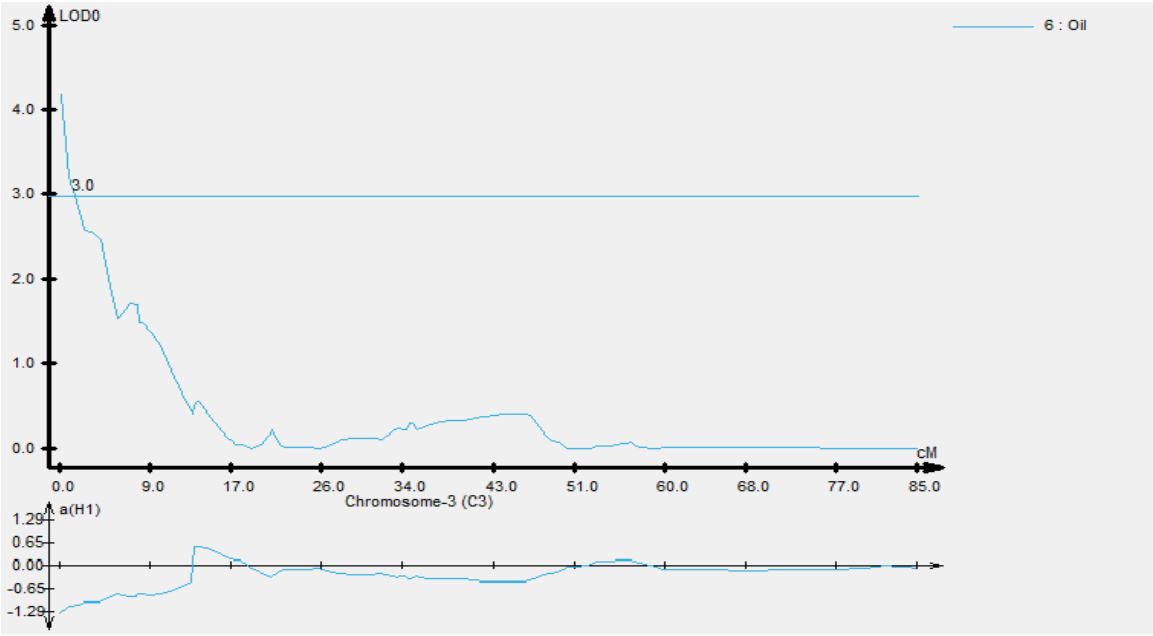

Figure 3.28: One oil QTL on chromosome 3. Oil-2 is located at 0.0 cM.

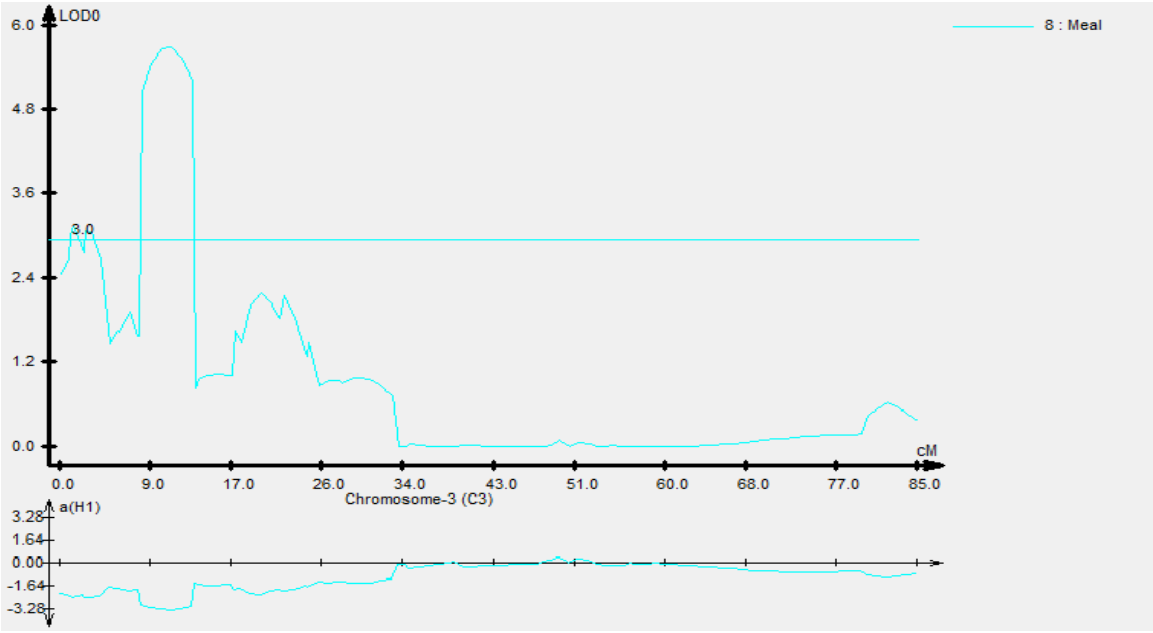

Figure 3.29: Three meal protein QTL located on chromosome 3. Mpro-1 is located at 1.4 cM, mpro-2 at 2.6 cM, and mpro-3 at 11.0 cM.

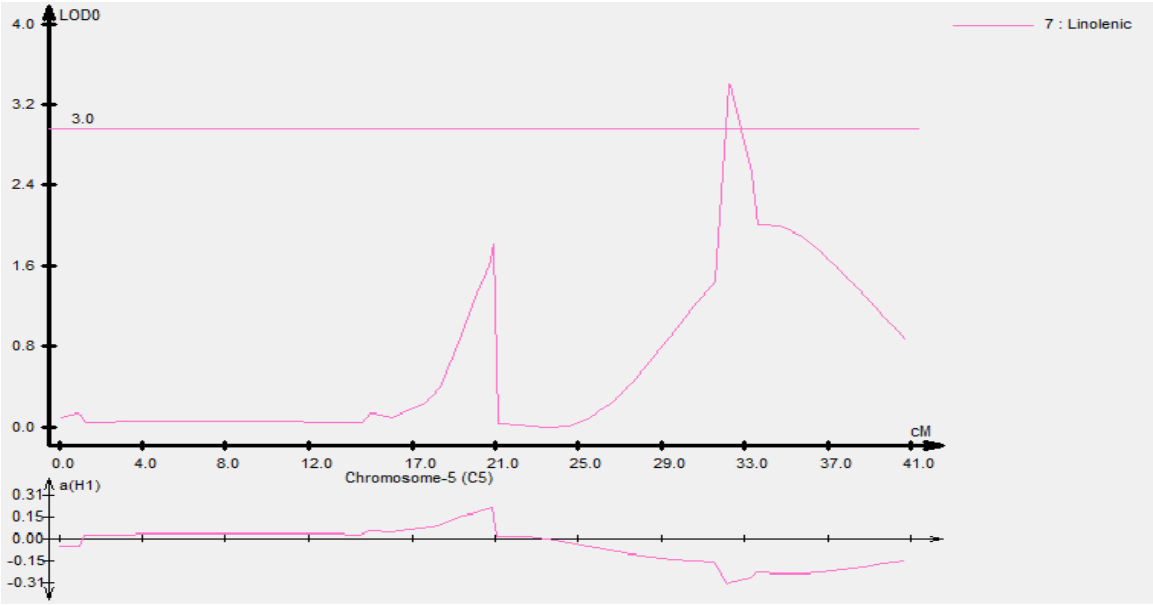

Figure 3.30: One linolenic acid QTL on chromosome 5. Lin-1 is located at 32.2 cM.

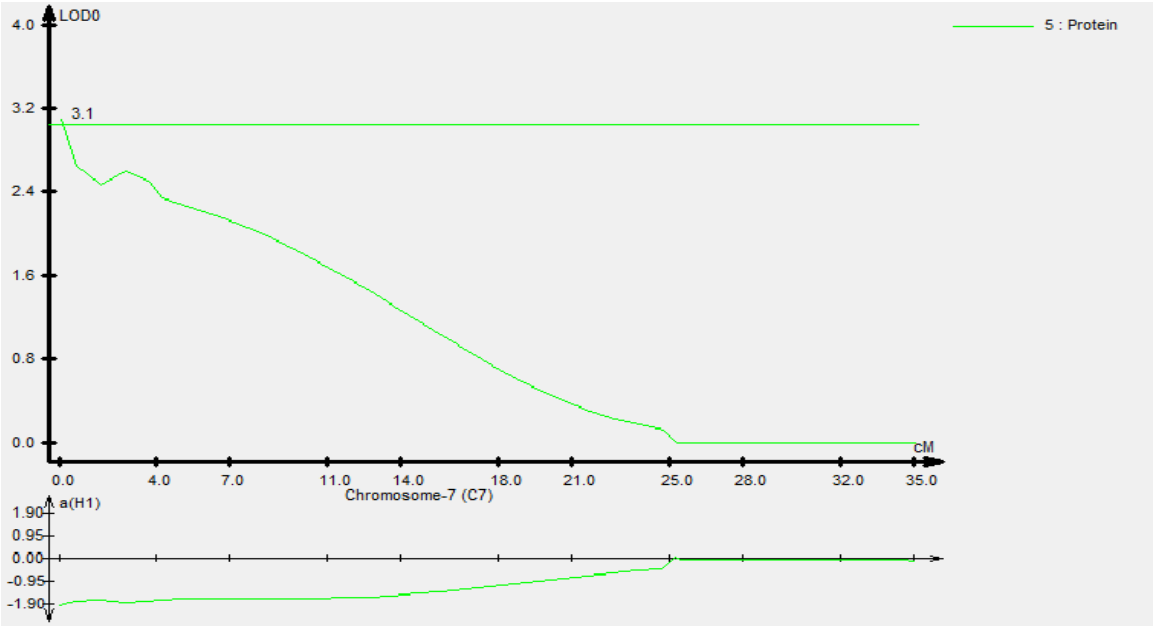

Figure 3.31: One protein QTL on chromosome 7. Pro-5 is located at 0.0 cM.

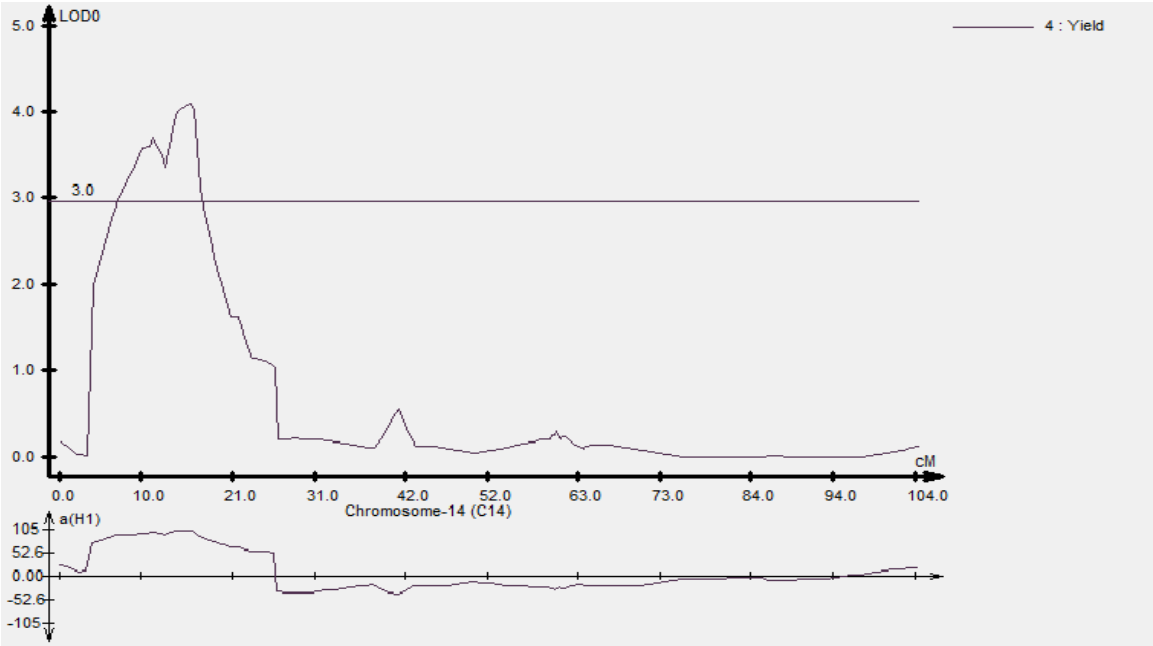

Figure 3.32: One yield QTL on chromosome 14. Yld-1 is located at 15.9 cM.

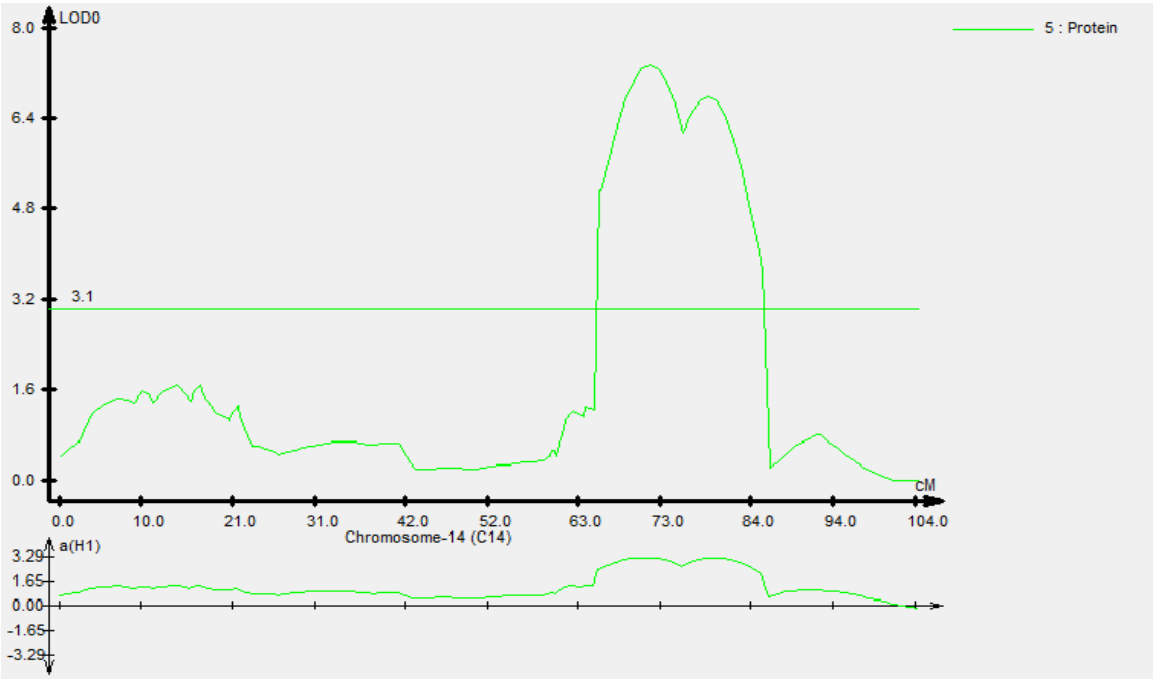

Figure 3.33: One protein QTL on chromosome 14. Pro-6 is located at 71.7 cM.

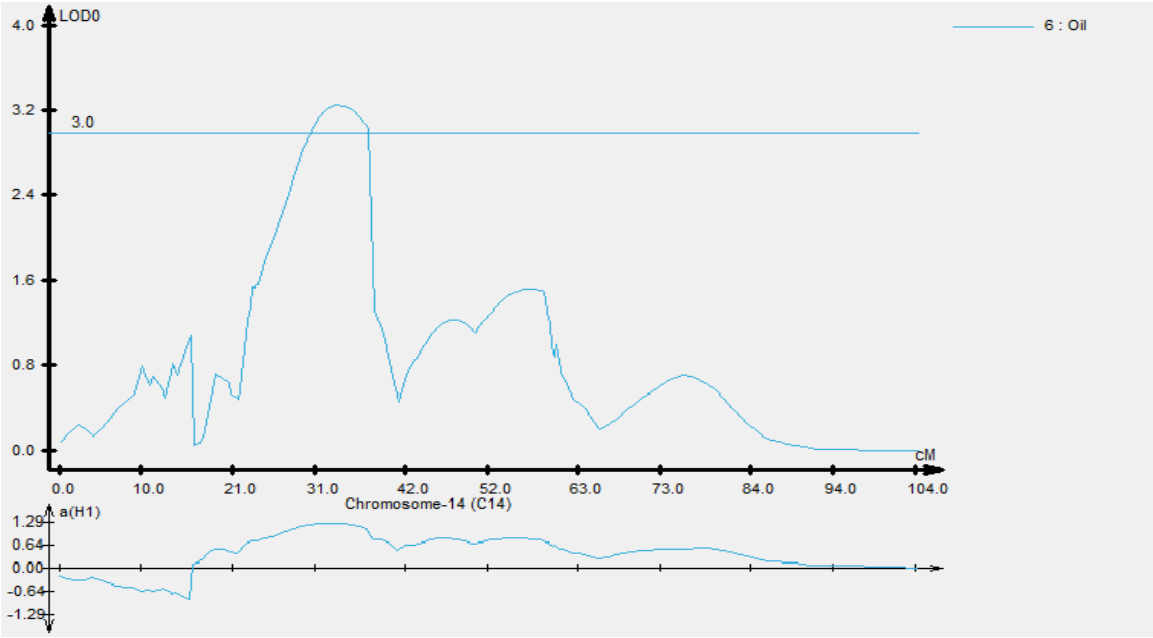

Figure 3.34: One oil QTL on chromosome 14. Oil-3 is located at 33.5 cM.

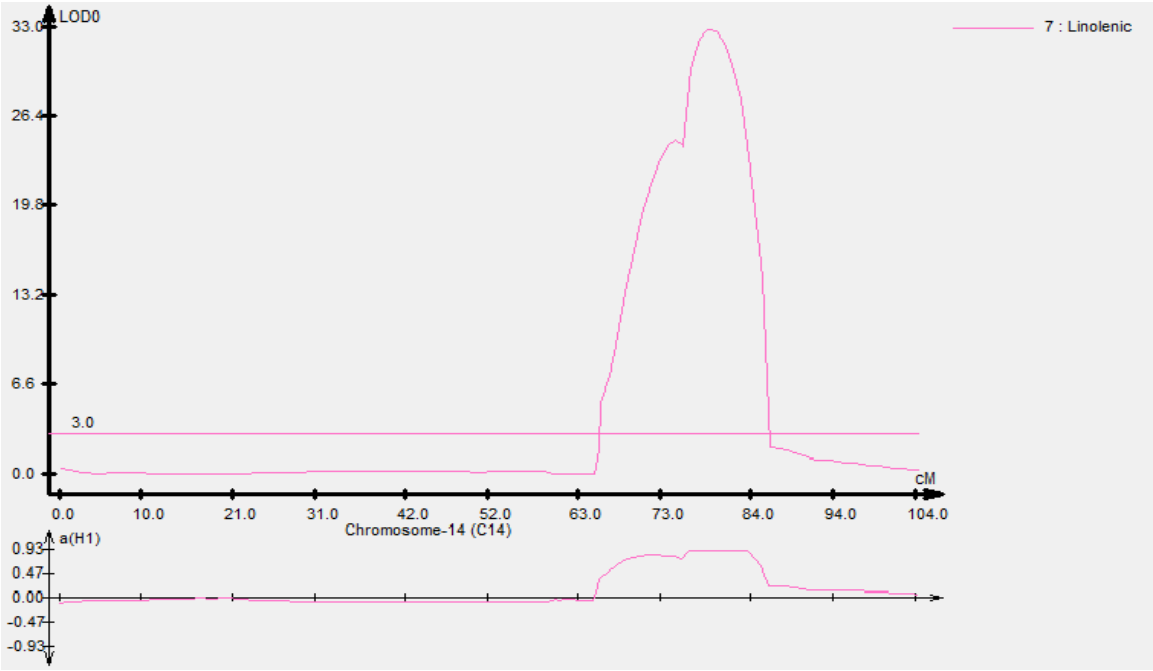

Figure 3.35: One linolenic acid QTL on chromosome 14. Lin-2 is located at 78.7 cM.

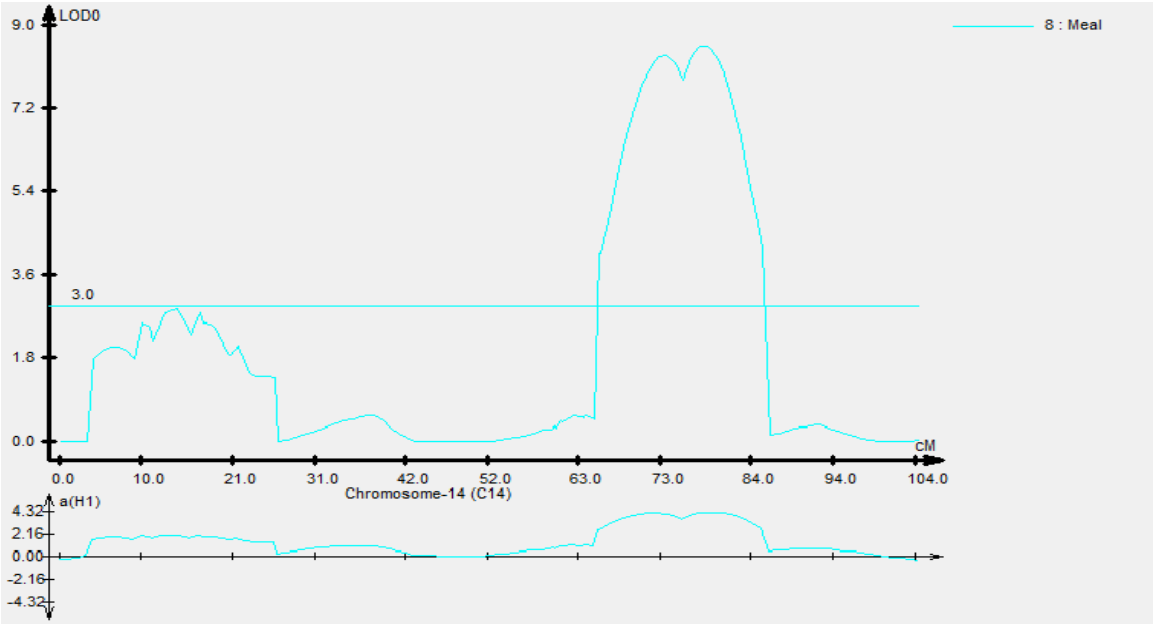

Figure 3.36: One meal protein QTL on chromosome 14. Mpro-4 is located at 78.7 cM.

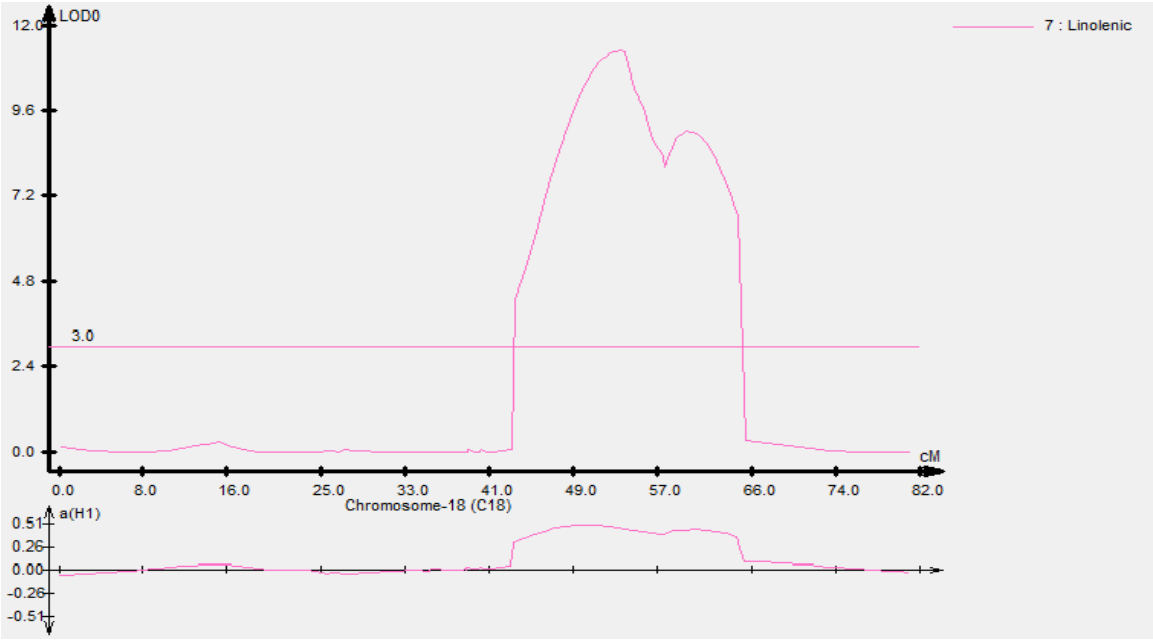

Figure 3.37: One linolenic acid QTL on chromosome 18. Lin-3 is located at 53.3 cM.

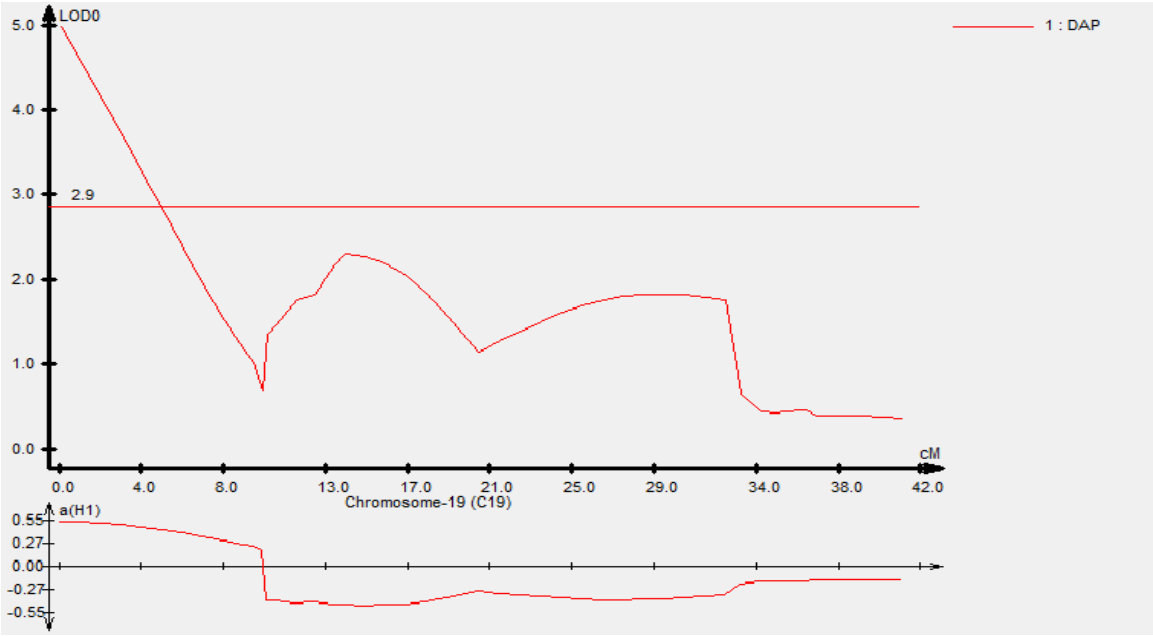

Figure 3.38: One days after planting (DAP) QTL on chromosome 19. Dap-1 is located at 0.0 cM.

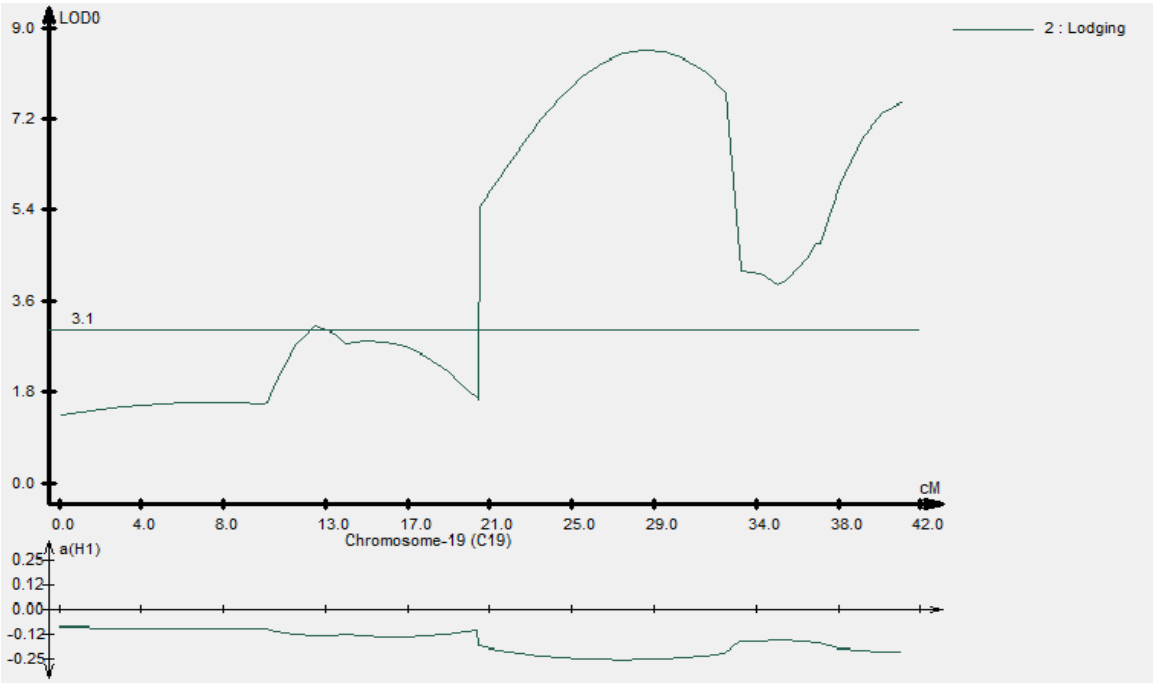

Figure 3.39: Two lodging QTL on chromosome 19. Lodg-1 is located at 12.4 cM and lodg-2 is at 28.5 cM.

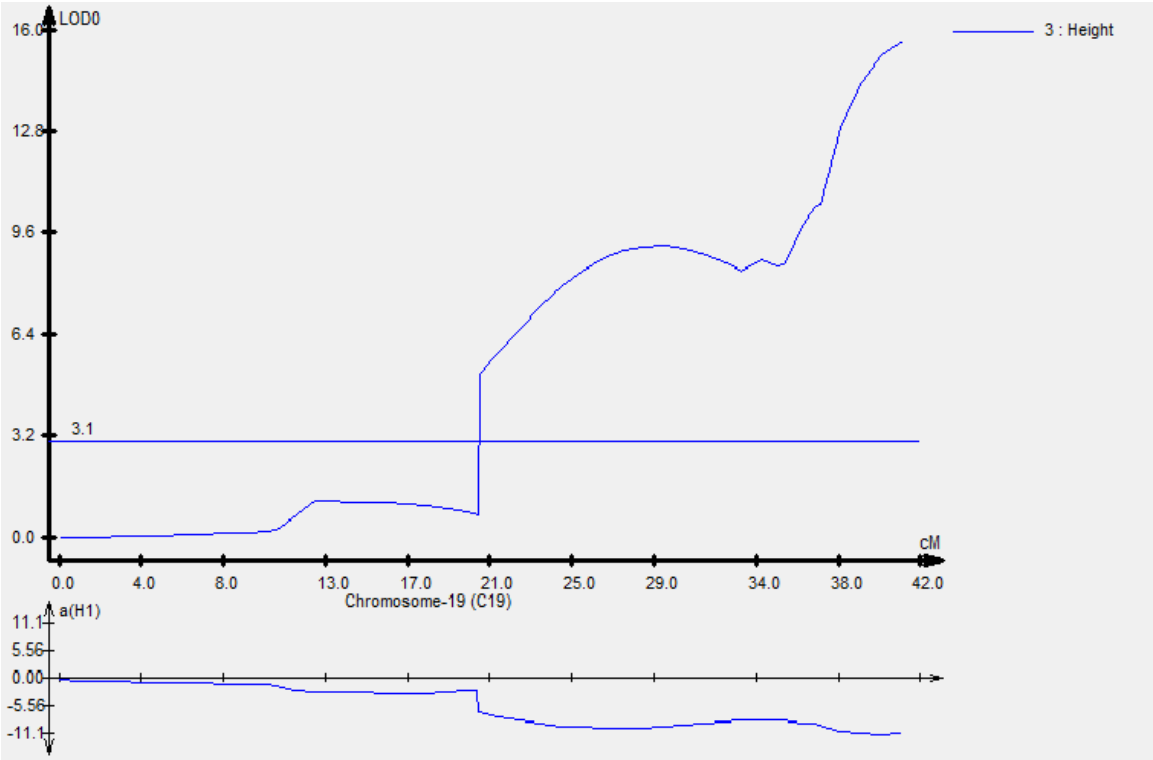

Figure 3.40: Hgt-1 on chromosome 19. This QTL is located at 41.1 cM.

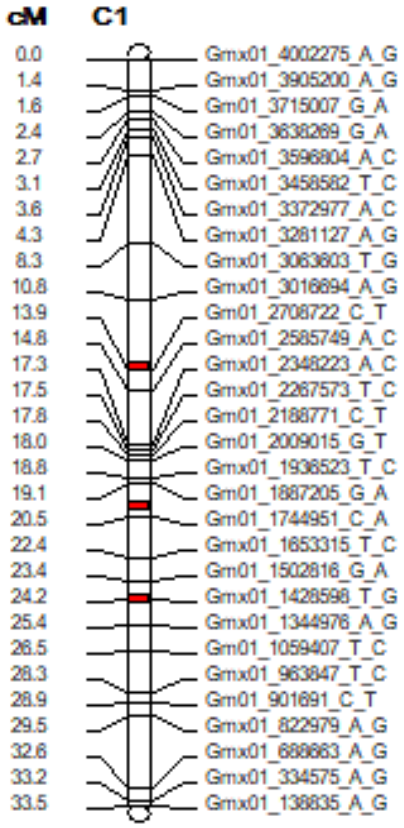

Figure 3.41: Chromosome 1 linkage map with QTL positions in red. Oil-1 is located at 13.8 cM, pro-1 is at 20.1 cM, and pro-2 at 24.2 cM.

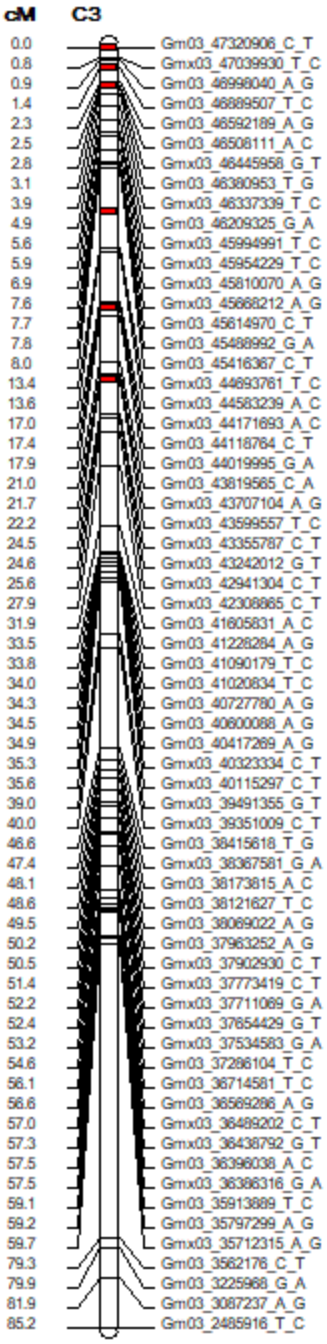

Figure 3.42: Chromosome 3 linkage map with QTL positions in red. Oil-2 is located at 0.0 cM, mpro-1 is at 1.4 cM, mpro-2 at 2.6 cM, mpro-3 at 11.0 cM, pro-3 at 17.4 cM, and pro-4 at 22.2 cM.

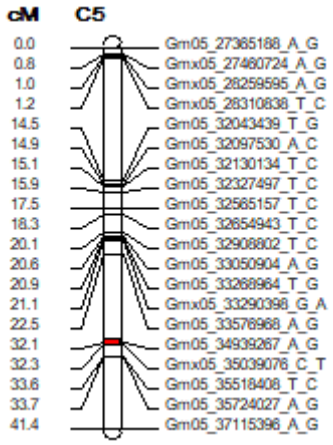

Figure 3.43: Chromosome 5 linkage map with QTL positions in red. Lin-1 is located at 32.2 cM.

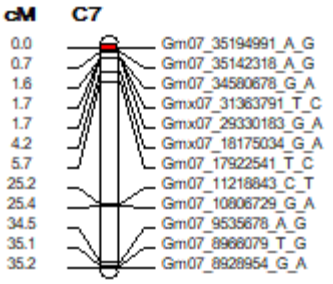

Figure 3.44: Chromosome 7 linkage map with QTL positions in red. Pro-5 is located at 0.0 cM.

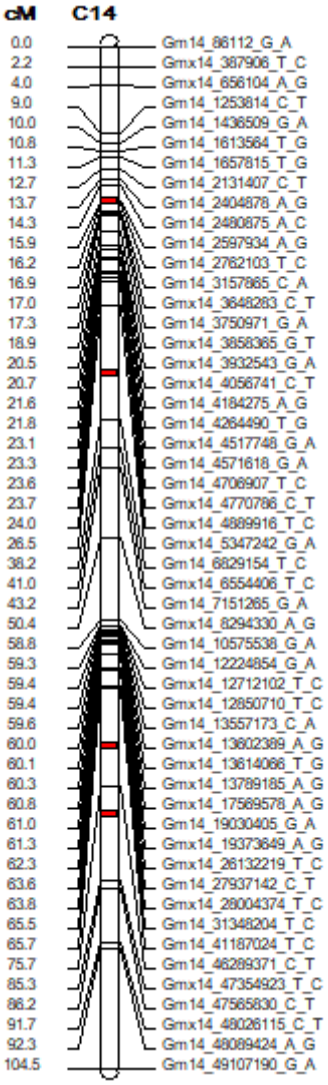

Figure 3.45: Chromosome 14 linkage map with QTL positions in red. Yld-1 is located at 15.9 cM, oil-3 at 33.5 cM, pro-6 at 71.7 cM, lin-2 at 78.7 cM, and mpro-4 at 78.7 cM.

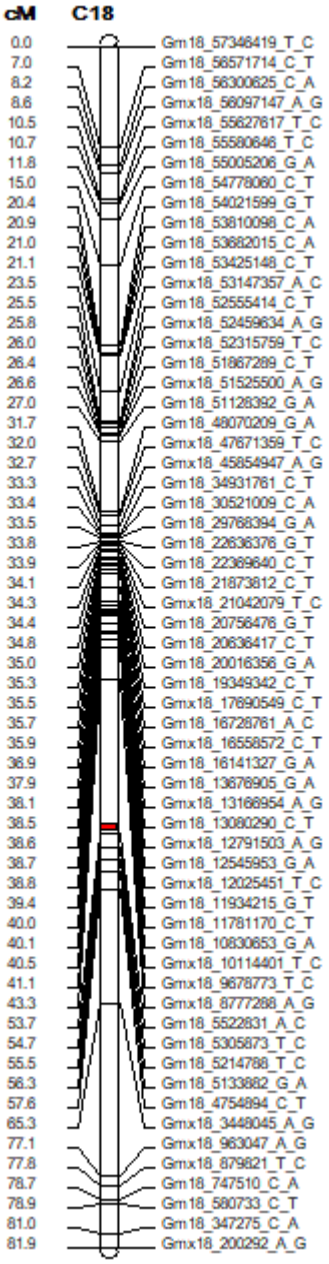

Figure 3.46: Chromosome 18 linkage map with QTL positions in red. Lin-3 is located at 53.3 cM.

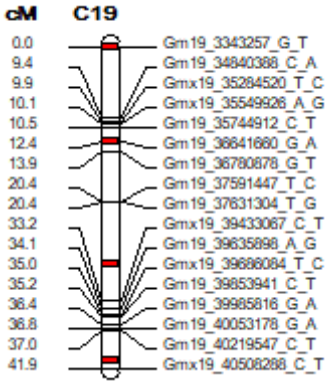

Figure 3.47: Chromosome 19 linkage map with QTL positions in red. Dap-1 is located at 0.0 cM, lodg-1 at 12.4 cM, lodg-2 at 28.5 cM, and hgt-1 at 41.1 cM.

Table 3.1: QTLs with respective traits and attributes.

Name	Trait	Chromosome	Marker Name	ssID	cM	LOD	R2	Effect[a]
dap-1	DAP	19	Gm19_3343257_G_T	715634003	0.0	5.0	11.7	0.5
lodg-1	Lodging	19	Gm19_36641660_G_A	715634509	12.4	3.1	5.5	-0.1
lodg-2	Lodging	19	Gm19_37631304_T_G	715634639	28.5	8.6	23.6	-0.2
hgt-1	Height	19	Gm19_40219547_C_T	715634990	41.1	15.7	31.5	-11.1
yld-1	Yield	14	Gm14_2597934_A_G	715618125	15.9	4.1	8.2	104.7
pro-1	Protein	1	Gm01_1887205_G_A	715578642	20.1	3.2	6.1	2.0
pro-2	Protein	1	Gmx01_1428598_T_G	715578509	24.2	3.2	5.9	1.9
pro-3	Protein	3	Gm03_44118764_C_T	715586280	17.4	4.0	7.2	-2.1
pro-4	Protein	3	Gmx03_43599557_T_C	715586248	22.2	3.2	5.8	-1.9
pro-5	Protein	7	Gm07_35194991_A_G	715597298	0.0	3.1	5.7	-1.9
pro-6	Protein	14	Gm14_41187024_T_C	715618716	71.7	7.4	17.6	3.3
oil-1	Oil	1	Gmx01_3016694_A_G	715578942	13.8	3.4	6.6	-1.2
oil-2	Oil	3	Gm03_47320906_C_T	715586637	0.0	4.2	8.2	-1.3
oil-3	Oil	14	Gmx14_5347242_G_A	715619662	33.5	3.3	8.0	1.3
lin-1	Linolenic	5	Gm05_34939267_A_G	715591126	32.2	3.4	3.9	-0.3
lin-2	Linolenic	14	Gm14_46289371_C_T	715619121	78.7	32.9	56.3	0.9
lin-3	Linolenic	18	Gmx18_8777288_A_G	715632812	53.3	11.3	14.5	0.5
mpro-1	Meal Protein	3	Gm03_46889507_T_C	715586578	1.4	3.2	5.5	-2.4
mpro-2	Meal Protein	3	Gm03_46508111_A_C	715586541	2.6	3.1	5.3	-2.4
mpro-3	Meal Protein	3	Gm03_45416367_C_T	715586421	11.0	5.7	12.2	-3.3
mpro-4	Meal Protein	14	Gm14_46289371_C_T	715619121	78.7	8.6	20.2	4.3

[a] TN13-4303 represents positive additive effects, and TN12-4098 represents the negative additive effects

CONCLUSION

180 recombinant inbred lines were created from a cross between TN12-4098 and TN13-4303 to create the project named "PRO-LIN". These lines were studied for two years in three locations (ETREC, HRREC, and RECM). The aim for the project was to find quantitative trait loci responsible for traits of interest including yield, seed protein, seed oil, meal protein, and seed linolenic acid content. In total, 21 QTL were found, and of those, 17 were for major traits of interest. One QTL was found for yield, six for seed protein, three for seed oil, four for meal protein, and three for seed linolenic acid content. QTLs discovered in this project need to be entered into SoyBase for future use by research teams. With the decreasing price of genomic screening, researchers will be able to utilize these loci in their elite soybean lines. Also, it would be wise to set up future experiments to confirm these QTLs, especially the QTLs considered to be novel. Although most lines showed little hope for future use in a breeding program, TNPL-146 did stand out from the rest. Of the RILs, this line ranked 9th in seed protein content, 1st in seed oil content, and 1st in meal protein content. This is a line that could be considered ideal for crossing into a high yielding line in the future. A modifier QTL located on chromosome 5 could be used in the future to help lower linolenic acid levels. This QTL has an additive effect of -0.3, associated with TN12-4098. Only 638 of the 11,874 markers were retained in the linkage maps used for the QTL detection. It would be interesting to see this project employ higher resolution linkage maps for a more accurate QTL detection. Many more minor QTL may be discovered with this method.

Nevertheless, QTLs discovered in this project should prove useful to researchers for years to come.

VITA

Ronald E. Moore was born February 27, 1996. Ronald grew up in Old Hickory, Tennessee. Ronald later attended the University of Tennessee, Knoxville where he received a Bachelor of Science degree in plant sciences and a Master of Science degree in plant sciences with a concentration in plant breeding.