ABSTRACT

NOBLE, AMANDA DAVINE. Improvement of Soybean Meal through Identification of High Protein QTL in Lines with Exotic Pedigree. (Under the direction of Dr. Lilian Miranda.)

Soybean [*Glycine max* (L.) Merr.] meal is a major source of protein globally for animal feeds, predominately in poultry and swine production. Elevating the protein concentration of this expensive and important feed component, would benefit the livestock feed industry. Therefore, increasing the protein concentration along with yield is a desirable, but complex goal due to the negative effects these traits tend to have on one another. The objectives of the present study were to investigate the genetic and phenotypic correlations between seed protein concentration and selected agronomic traits, to evaluate the yield potential of these high protein lines, to identify high protein QTL in a population derived from a cross between moderately high protein soybean genotypes G03-3385 and N06-10035, determine their relationship with oil QTL, and to study the interaction of high protein QTL from different exotic sources. A population of 132 $F_{4:6}$ recombinant inbred lines, derived from a cross between moderately high protein soybean genotypes G03-3385 and N06-10035, was evaluated at two locations in 2013. The population was genotyped with 392 polymorphic markers and was analyzed by near-infrared reflectance spectroscopy to determine protein and oil concentration. Putative QTL were detected using multiple interval mapping (MIM). The across-locations analyses revealed 12 and 8 QTL significantly associated with protein and oil concentration, respectively. Although several of these were co-localized, eight of the protein QTL were independent of oil QTL and one had the same positive allele for both traits. The QTL and the epistatic interactions identified in this study explain 93.99% and 97.26% of the total phenotypic variation observed for oil and protein, respectively. Dunnett's test to

compare multiple lines to a control line revealed several lines that were either not significantly different or even higher yielding than the highest yielding check and also had significantly higher protein concentration. Genotypic and phenotypic correlations were estimated to determine the effects of selected agronomic traits on soybean seed protein concentration. A significant negative genotypic correlation was observed between protein concentration and yield (-0.3355) but no significant phenotypic correlation was found. Protein concentration and oil concentration had significant negative phenotypic (-0.7123) and genotypic (-0.7233) correlations. Oil concentration and yield had significant positive genotypic (0.4409) and phenotypic (0.1175) correlations.

These results suggest that the negative associations between yield and protein could be overcome so that protein can be increased without significantly affecting yield; however, large changes in protein concentration will most likely have a negative impact on oil concentration.

Improvement of Soybean Meal through Identification of High Protein QTL in Lines with Exotic Pedigree

by
Amanda Davine Noble

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

Crop Science

Raleigh, North Carolina

2015

APPROVED BY:

_____	_____
Dr. Earl Taliercio	Dr. Consuelo Arellano

_____
Dr. Lilian Miranda
Committee Chair

BIOGRAPHY

Amanda Noble is a native of the mountains of North Carolina, born and raised in Henderson County. She began her education and participation in the natural sciences early in life, participating in FFA as well as working at a small animal veterinary hospital- all before high school graduation. As an undergraduate at North Carolina State University, Amanda majored in Plant and Soil Sciences with a concentration in Soil Science. While at NCSU she was an active member of Alpha Zeta, also participating and coaching the NCSU Soil Judging team. While working towards her undergraduate degree she worked as a part-time employee in Dr. Lilian Miranda and Dr. Tommy Carter's soybean breeding programs, during which time she decided to continue her education and pursue a career in Plant Breeding. Amanda remained with the Crop Science department at North Carolina State University under the guidance of Dr. Lilian Miranda and began her work towards a Master's Degree in Crop Science and completed an internship with Dr. Albert Santos of Bayer CropScience. Shortly before graduation with her M.S. in Crop Science, Amanda joined the Bayer CropScience Global Soy Molecular Breeding team as Assistant Soybean Discovery Breeder located in Pikeville, NC and will continue her career there.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Lilian Miranda, for her constant support and encouragement as I worked through this thesis. I would also like to thank my committee members: Dr. Consuelo Arellano and Dr. Earl Taliercio for all of their guidance, feedback and advice.

I also want to recognize the USDA-ARS Soybean and Nitrogen Fixation Unit staff for their positivity, assistance, and constant contribution to a wonderful work and learning environment: Rebecca Whetten, Earl Huie, Dr. Margarita Villagarcia, Connie Bryant, Cory Callahan, William "Bill" Novitzky, Dr. Thomas Carter, Robert Masson, Phil Rzewnicki, Dr. Andrew Scaboo, Hailey Huie, Katherine Novitzky, Brad Rush, John Graber, Christina Markunas, Kyle Harlos, Cameron Davis, Hannah Russell, Whitney Phillips, Jennifer Amick, and Ling Yuan. Additionally, my fellow past and present graduate students at the Soybean and Nitrogen Fixation Unit, Leah Ruff, Paul Ruddle, and Rochelle Ruddle, deserve mention and thanks for their unique knowledge, experience, and companionship.

Finally, I would like to thank the United Soybean Board for funding my graduate research and research assistantship. Also, Cathy Herring, Travis Lassiter, and the station workers at the Central Crops Research Station for their help with my research plots in Clayton, NC.

TABLE OF CONTENTS

LIST OF TABLES vi
LIST OF FIGURES viii
CHAPTER 1: Genetic Correlations of Soybean Seed Protein Concentration and Agronomic Traits 1
ABSTRACT 2
INTRODUCTION 3
MATERIALS AND METHODS 5
Plant Materials 5
Field Evaluation 5
Seed Composition Analysis 6
Statistical Analysis 7
RESULTS 8
Comparisons of Genotypic and Phenotypic Correlations 8
Field Evaluation 8
DISCUSSION 10
REFERENCES 13
TABLES 16
CHAPTER 2: Identifying Soybean Protein QTL in Lines with Exotic Pedigree **19**
ABSTRACT 20
INTRODUCTION 21
MATERIALS AND METHODS 22
Plant Materials 22
Field Evaluation 23
Seed Composition Analysis 24
Molecular Marker Analysis 25
Selective Genotyping 25
Statistical Analysis 27
Linkage and QTL Mapping 28
RESULTS 29
Seed Composition Analysis 29

Linkage Mapping....29
QTL Mapping of Protein Concentration....30
QTL Mapping of Oil Concentration....31
Selective Genotyping....31
Heritability....32
DISCUSSION....32
REFERENCES....36
TABLES....40
FIGURES....49

LIST OF TABLES

CHAPTER 1: Agronomic Evaluation and Genetic Correlations of Soybean Seed Protein Concentration and Agronomic Traits

Table 1.1. Early maturity group top 20 highest yielding lines for soybean [Glycine max (L.) Merr.] with values for protein concentration and oil concentration from population Pro 5 (G03-3385 x N06-10035) combined over locations, Clayton and Plymouth, NC 2013, compared to the highest yielding early maturity checks 16

Table 1.2. Late maturity group top 20 highest yielding lines for soybean [Glycine max (L.) Merr.] with values for protein concentration and oil concentration from population Pro 5 (G03-3385 x N06-10035) grown at Clayton and Plymouth, NC 2013, and combined over locations compared to the highest yielding late maturity checks 17

Table 1.3. Genotypic (above diagonal) and phenotypic (below diagonal) correlation coefficients, and their standard errors (in parentheses), between all pairs of traits (yield, protein, oil, and maturity date [R8]) for 132 soybean [Glycine max (L.) Merr.] lines from population Pro 5 grown at Clayton and Plymouth, NC, in 2013 18

CHAPTER 2: Genetic Mapping of Quantitative Trait Loci Associated with Seed Protein Content in a Soybean Population Derived from Germplasm Lines with Exotic Pedigree

Table 2.1. Protein concentration means for N06-10035, G03-3385, and for the 132 G03-3385 × N06-10035 RILs in all years and locations 40

Table 2.2. Protein concentrations for N06-10011, G03-3621, and for the 133 G03-3621 × N06-10011 RILs in all years 40

Table 2.3. Protein concentration correlation coefficients for G03-3385 x N06-10035 RIL (Pro 5) population across all years 41

Table 2.4. Protein concentration correlation coefficients for G03-3621 x N06-10011 RIL (Pro 16) population across all years 41

Table 2.5. Estimates of increased protein QTL positions, effects and interactions identified by multiple interval mapping from Clayton 2013 location for the G03-3385 x N06-10035 RIL (Pro 5) population 42

Table 2.6. Estimates of increased protein QTL positions, effects and interactions identified by multiple interval mapping from Plymouth 2013 location for the G03-3385 x N06-10035 RIL (Pro 5) population 43

Table 2.7. Estimates of increased protein QTL positions, effects and interactions identified by multiple interval mapping from across locations (Clayton 2013 and Plymouth 2013) location for the G03-3385 x N06-10035 RIL (Pro 5) population 44

Table 2.8. Estimates of increased oil QTL positions, effects and interactions identified by multiple interval mapping from Clayton 2013 location for the G03-3385 x N06-10035 RIL (Pro 5) population 45

Table 2.9. Estimates of increased oil QTL positions, effects and interactions identified by multiple interval mapping from Plymouth 2013 location for the G03-3385 x N06-10035 RIL (Pro 5) population 46

Table 2.10. Estimates of increased oil QTL positions, effects and interactions identified by multiple interval mapping from across locations (Clayton 2013 and Plymouth 2013) location for the G03-3385 x N06-10035 RIL (Pro 5) population..... 47

Table 2.11. Putative QTL associated with soybean seed protein concentration confirmed using a 'trait-based' bidirectional selective genotyping analysis (Navabi et al, 2009) of Pro 16..... 48

Table 2.12. Broad-sense heritability (H2) estimates for both entry mean and plot basis with their standard errors for yield, seed protein concentration, and seed oil concentration for G03-3385 x N06-10035 RIL (Pro 5) population across locations (Clayton, NC 2013 and Plymouth, NC 2013) 48

LIST OF FIGURES

FIGURE 1. Linkage map of the G03-3385 x N06-10035 RIL population (Pro 5)..... 49

CHAPTER 1: Agronomic evaluation and Genetic Correlations of Soybean Seed Protein Concentration and Agronomic Traits

ABSTRACT

Soybean [*Glycine max* (L.) Merr.] meal is a major source of protein for animal feeds worldwide. Therefore, increasing seed protein concentration along with yield is a desirable, but complex goal due to the negative effects these traits tend to have on one another. The objectives of the present study were to investigate the genetic and phenotypic correlations between seed protein concentration and selected agronomic traits and to evaluate the yield potential of these high protein lines. A population of 132 $F_{4:6}$ recombinant inbred lines, derived from a cross between moderately high protein soybean genotypes G03-3385 and N06-10035, was evaluated at two locations in 2013. Dunnett's test (Steel et al., 1997) for pairwise mean comparison revealed several lines that were either not significantly different or even higher yielding than the highest yielding check and also had significantly higher protein concentration. Genotypic and phenotypic correlations were estimated to determine the effects of selected agronomic traits on soybean seed protein concentration. A significant negative genotypic correlation was observed between protein concentration and yield (-0.3355) but no significant phenotypic correlation was found. Protein concentration and oil concentration had significant negative phenotypic (-0.7123) and genotypic (-0.7233) correlations. Oil concentration and yield had significant positive genotypic (0.4409) and phenotypic (0.1175) correlations. These results suggest that the negative associations between yield and protein could be overcome so that protein can be increased without significantly affecting yield; however, large changes in protein concentration will most likely have a negative impact on oil concentration.

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is a high demand crop that, in the form of soybean meal, is a major source of protein for animal feeds globally. The U.S. produced 36.3 million metric tons (Tg) of soybean meal in 2013 with 26.5 Tg consumed by livestock, poultry is the highest consumer at 13.3 Tg (American Soybean Association, 2014). Poultry and swine production combined is estimated to grow at about 5 Tg/year thus soybean meal supply is expected to grow at the same rate, globally (Wilson, 2010).

Soybean seed protein content is a quantitative trait, controlled by multiple genes with relatively small effects easily influenced by the environment. Along with the environmental factors affecting soybean protein concentration, significant positive and negative correlations between protein concentration and other agronomically important traits have been reported. Increasing seed protein content in soybean seed while also maintaining or increasing yield, has proven difficult due to the frequently observed negative correlation between these two traits (Brim and Burton, 1979; Wilcox and Shibles, 2001; Cardinal et al., 2008); however, more recent studies indicate that, with the appropriate breeding approach, it should be possible to improve yield and seed composition traits simultaneously. Recker et al. (2014) observed that the genetic correlation between protein and yield was negligible in populations that were randomly mated for many generations and had linkage phase disequilibrium (LD) likely close to zero. This research concluded that pleiotropic effects rather than genetic linkage likely caused the correlation between protein and oil to remain negative. Through the use of molecular markers, seed composition traits can be dissected into QTL and the effect of

individual loci on other agronomically important traits could be determined before they are incorporated into elite germplasm using marker assisted selection (MAS). Eskandari et al. (2013) identified QTL alleles associated with increased seed oil concentration that were also positively associated with increased protein concentration and yield in a population of recombinant inbred lines. They observed that the oil-beneficial allele of a QTL, linked to marker Sat_020, was positively associated with seed protein content and another, linked to markers Satt001 and GmDGAT2B, was positively correlated with seed yield. Significant two-way epistatic interactions, with one of the markers only associated with seed oil concentration, were observed for both seed protein concentration and seed yield. With the increasing market demand for soybean meal, identifying populations that can overcome the commonly found negative correlations between yield and protein; and protein and oil would help increase profitability of this high demand crop to meet future needs.

In this study, a recombinant inbred line (RIL) population, Pro 5, derived from a biparental cross of moderately high protein and genetically unrelated germplasm lines, was used to address the following objectives: (1) to elucidate the genetic and phenotypic correlations between seed protein concentration and selected agronomic traits and (2) to evaluate the yield potential of these high protein lines.

MATERIALS AND METHODS

Plant Materials

A high protein population (Pro 5) was developed by crossing the germplasm lines G03-3385 with N06-10035 in 2009 in Clayton, NC. F_1 plants were grown in Isabella, Puerto Rico in winter of 2009-2010. Plants were advanced by single seed descent (Brim, 1966) to the F_4 generation. One hundred thirty-two $F_{4:5}$ individual plants were harvested in Clayton, NC in the fall of 2011.

The pedigree of the parents is as follows: N06-10035 is a high protein selection from the cross 'Young' (Burton et al., 1987) x 'N6202' (Carter et al., 2010). G03-3385 is a high protein selection from the cross 'Benning' (Boerma et al., 1997) x Korean PI 'Danbaekkong' (PI 619083) (Kim, 1996).

Field Evaluation

In 2011, F_4 plants were grown in Clayton, NC and harvested individually. In 2012, $F_{4:5}$ lines were grown in un-replicated 3.7 m long one-row plots in Clayton, NC in Varina Loamy Sand soil. In 2012, flower color at the R2 reproductive stage, pubescence color, and maturity date at the R8 reproductive stage were recorded (Fehr and Caviness, 1977).

In 2013, 132 $F_{4:6}$ lines from Pro 5 were randomly divided into four sets for field evaluation. Within each set the lines were arranged in a randomized complete block design with three replications. The Pro 5 lines' parents G03-3385 and N06-10035, and 'N6202', a

parent of N06-10035, were included as check lines. N08-145, 'NC-Roy', 'Dillon', and '5601T' were included as check lines for yield and maturity. N08-145 is a high yielding MG VI experimental line that was the highest yielding line within its MG in the NC Official Variety Testing in 2012 (http://www.ncovt.com/files/Soybeans2012_Table22.pdf). NC-Roy (Burton et al., 2005) and Dillon (Shipe et al., 1997) are late maturity group VI cultivars, and 5601T is a maturity group V cultivar (Pantalone et al., 2003).

The tests were grown in 2013 in Plymouth, NC (planted 21 May) and Clayton, NC (planted 30 May). Soils were Portsmouth fine sandy loam and Norfolk loamy sand, respectively. Plots were three rows wide (0.97 m between rows) and 4.3 m long. Four hundred seed were planted in each plot. Plots were end-trimmed to 3.0 m and only the center row was harvested for yield and 100-seed weight determination. Flower color at the R2 reproductive stage, pubescence color, and maturity date at the R8 reproductive stage were recorded (Fehr and Caviness, 1977).

Seed Composition Analysis

In all years, protein and oil concentration on a dry weight basis were determined by sending approximately 25 grams of seed samples from each plant/line to the USDA-ARS National Center for Agricultural Utilization Research in Peoria, IL. Analysis of the whole bean samples was completed by near infrared reflectance spectroscopy (NIR) in an Infratec 1241 Grain Analyzer (Foss North America, Inc., Eden Prairie, MN).

Statistical Analysis

For yield evaluation, the MIXED (Littell et al., 2006) procedure in SAS 9.4 (SAS Institute Inc, 2013) was used. Lines were split into two groups (early and late) based on their maturity and tested against the highest yielding maturity check lines.

Initially, to identify outliers, studentized residuals for yield were calculated for each location separately and observations with studentized residuals greater in absolute values than 3.5 were excluded from the data for the remainder of the analysis. A mixed model was used with replicate (set) considered random, and set and lines (set) were considered fixed.

In the second analysis, genotypic and phenotypic correlations for only combined locations were estimated according to Holland (2006). For this analysis the Pro 5 lines (set), replicate (location*set), and the appropriate interactions with fixed effects were considered random effects. Locations, sets, and location*set were considered fixed effects.

A third analysis was conducted to evaluate the yield potential of the lines in the Pro 5 population. For this purpose the lines were divided based on their maturity group into an early and late group. The “early” group was tested against the yield check lines: 5601T and N08-145. The “late” group was tested against the yield check lines: Dillon and Roy. A mixed model was used to analyze the field experiments data with each location considered separately. Line (set), replicate, and the appropriate interactions with fixed effects were considered random. Set was considered a fixed effect. For the combined analysis of locations, with early and late groups still separate, an analysis of variance was performed

with location, line, and location*line being fixed effects. Set, set*location, and replicate (set*location) were considered random effects. Genotypic class LS means were compared using a Dunnett's test across locations and within locations at a 0.05 significance level (α=0.05).

RESULTS

Comparisons of Genotypic and Phenotypic Correlations

A significant negative genotypic correlation (-0.3355) was observed between protein concentration and yield, but no significant phenotypic correlation was observed. Significant and relatively large negative phenotypic (-0.7123) and genotypic (-0.7233) correlations were observed between protein concentration and oil concentration. The phenotypic and genotypic correlations between seed oil concentration and maturity were both significant and negative (-0.2208 and -0.2346, respectively). A very similar relationship was observed between protein concentration and maturity date. In this case the phenotypic correlation was -0.1742 and the genotypic correlation was -0.2156, both significant. In contrast, oil and yield were positively correlated genotypically (0.4409) and phenotypically (0.1175). No significant phenotypic or genotypic correlations were observed between maturity and yield (Table 1.7).

Field Evaluation

Forty-four lines were considered within the early maturing group. In the analysis of variance (ANOVA) for yield for the early maturing group across locations, lines were a significant source of variation ($p<0.01$) while locations or the interaction between location and lines

were not. Therefore, only the across locations results are presented (Table 1.1). Ten lines were not significantly different from the highest yielding check ($p<0.05$) based on the Dunnett's pairwise mean comparison test, with no lines being higher yielding than the highest yielding check (Table 1.1). These 10 lines were significantly higher in protein content ($p<0.01$) but all ten were significantly lower in oil content ($p<0.05$).

Eighty-six lines were considered within the late maturity group. The ANOVA for yield for the late maturing group across locations showed that locations ($p<0.05$), lines ($p<0.01$) and the interaction between location and lines ($p<0.01$) were all significant sources of variation (Table 1.2). In Clayton 2013, 69 lines from the late maturity group were not significantly different from the highest yielding check ($p<0.05$) based on the Dunnett's test, 22 of those lines were higher yielding than the highest yielding check and for one of these, the higher yielding value was statistically significant ($p<0.01$) based on Dunnett's test. Among the 21 highest yielding lines, all had significantly higher protein content, five of them were higher in oil compared to the highest yielding check but this difference was not statistically significant) . In Plymouth 2013, the yield of 42 lines were not significantly different from the highest yielding check ($p<0.05$) based on the Dunnett's pairwise comparison test but no lines were higher yielding than the highest yielding check. Across 2013 locations for the late group, 39 lines were not significantly different from the highest yielding check ($p<0.05$) based on the Dunnett's pairwise comparison and two of those lines were higher yielding than the highest yielding check. Among those two lines, both were

significantly higher in protein but were lower in oil content compared to the highest yielding check.

DISCUSSION

The lines evaluated in this study showed great potential as germplasm for breeding programs that aim to improve seed composition. Both the early and the late maturing groups had several lines that were either not statistically significantly different or even higher yielding than the highest yielding check, and also had significantly higher protein content. Because of their exotic pedigree, the breeding lines also have the added advantage of increasing genetic diversity.

The significant negative relationship between seed protein and oil concentrations has been observed in many studies (Wilcox and Shibles, 2001; Hyten et al., 2004; Kumar et al., 2006; Eskandari et al., 2013). The correlation values for protein and oil concentrations reported in these other studies are comparable to those found in our study.

The research by Recker et al. (2014) indicated that this relationship is most likely due to pleiotropic effects rather than genetic linkage; however, it also suggested that it is possible to improve yield and protein concentration without significantly affecting seed oil content given that the estimated value for the correlated response of protein and oil to direct selection on yield in a long term random mated population, was a slight gain in oil of 2.9 g kg^{-1} and a very slight gain in protein of 1.4 g kg^{-1}". A negative relationship between seed protein and yield has also been previously reported in studies (Brim and Burton, 1979; Wilcox and Shibles, 2001) but in this research the genetic correlation was only slightly significant and

the phenotypic correlation was not significant. These results point to the potential of this population for developing high protein cultivars without significantly negatively affecting yield. The positive relationship between oil and yield has been reported in previous studies (Cardinal et al., 2008 and Carlson et al., 2011). This positive relationship suggests that this population would be ideal for selection of both high yield and high oil genotypes.

The negative relationship between seed oil and maturity has been observed by Bellaloui et al. (2009) and Carlson et al. (2011). Opposite from our results, Bellaloui et al. (2009) and Carlson et al. (2011) observed a positive relationship between protein concentration and maturity, but Recker et al. (2014) found a negative phenotypic correlation between protein concentration and maturity. Seed composition is dependent on the temperature during the seed filling period; Kumar et al. (2006) reported a significant positive correlation (0.438) between mean minimum day temperature and protein concentration. Different environments will have varying temperatures during seed development and this could explain the discrepancies among different studies. Our results indicate that early maturing lines in this population tend to have higher protein and oil concentration.

The results of this study could be helpful in the development of high protein cultivars without negatively affecting yield or oil concentrations. With small increases in protein the negative effect on oil can be minimized and vice versa. Yield of the Pro 5 lines is comparable to the yield checks without a significant reduction in oil content. Dissecting the quantitative traits using molecular markers to detect QTL that increase seed protein and oil content

without the negative affect on yield is one breeding approach to consider with these lines and will be discussed in the next chapter.

REFERENCES

American Soybean Association. 2013. SoyStats: A Reference Guide to Important Soybean Facts & Figures. http://www.soystats.com (accessed 21 Jan. 2014)

Bellaloui, N., J.R. Smith, J.D. Ray, A.M. Gillen. 2009. Effect of Maturity on Seed Composition in the Early Soybean Production System as Measured on Near-Isogenic Soybean Lines. Crop Sci. 49:608-620

Boerma, H.R., R.S. Hussey,D.V. Phillips, E.D.Wood,G.B. Rowan, and S.L. Finnerty. 1997. Registration of 'Benning' soybean. Crop Sci. 37:1982.

Brim, C.A. 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6:220

Brim, C.A. and J.W. Burton. 1979. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Sci. 19:494-498

Burton, J.W., C.A. Brim, and M.F. Young. 1987. Registration of 'Young' soybean. Crop Sci. 27:1093

Burton, J.W., T.E. Carter, Jr., and D.T. Bowman. 2005. Registration of NC-Roy soybean. Crop Science. 45:2654.

Cardinal, A.J., R.E. Dewey, and J.W. Burton. 2008. Estimating the individual effects of the reduced palmitic acid fapnc and fap1 alleles on agronomic traits in two soybean populations. Crop Sci. 48:633-639

Carlson, C.M., A.J. Cardinal, J.W. Burton, D.A. Dickey, J.B. Holland, J.M. Chaky. 2011. Genetic Control of Protein and Amino Acid Content in Soybean Determined in Two Genetically Connected Populations. M.S. Thesis. Available at: http://repository.lib.ncsu.edu/ir/handle/1840.16/7371; verified 18 December 2014.

Carter, T.E. Jr., P.E. Rzewnicki, J.W. Burton, M.R. Villagarcia, D.T. Bowman, E. Taliercio, and P. Kwanyuen. 2010. Registration of N6202 Soybean Germplasm with High Protein, Favorable Yield Potential, Large Seed, and Diverse Pedigree. Journal of Plant Registrations. 4:73-79

Eskandari, M., and E.R. Cober. 2013. Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. Theor Appl Genet. 126:1677-1687

Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. In Agric. And Home Economics Exp. Stn. Special Rep. no. 80. Iowa State Univ., Ames.

Holland, J.B. 2006. Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation with SAS Proc MIXED. Crop Sci. 46:642-654.

Hyten, D.L., V.R. Pantalone, C.E. Sams, A.M. Saxton , D. Landau-Ellis, T.R. Stefaniak, M.E. Schmidt. 2004. Seed quality QTL in a prominent soybean population. Theor Appl Genet. 109:552-561

Kim S.-D., E.-H. Hong, Y.-H. Kim, S.-H. Lee, Y.-K. Seong, K.-Y. Park, Y.-H. Lee, Y.-H. Hwang, E.-H. Park, H.-S. Kim, Y.-H. Ryu, R.-K. Park, and Y.-S. Kim. (1996) A new high protein and good seed quality soybean variety "Danbaegkong". RDA Journal of Agricultural Science- Upland and Industrial Crops 38:228-232.

Kumar, V., A. Rani, S. Solanki, and S.M. Hussain. 2006. Influence of growing environment on the biochemical composition and physical characteristics of soybean seed. J. Food. Comp. and Analysis. 19:188-195.

Littell, R. C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberber. 2006. SAS for Mixed Models. 2nd ed. SAS Institute, Inc., Cary, NC.

Pantalone, V.R., F.L. Allen, and D. Landau-Ellis. 2003. Registration of ‘5601T’ Soybean. Crop Sci.. 43:1123-1124

Recker, J.R., J.W Burton, A. Cardinal, L.M. Miranda. 2014. Genotypic and Phenotypic Correlations of Quantitative Traits in Two Long-Term, Randomly Mated Soybean Populations. Crop Sci. 54:939-943

SAS Institute Inc. 2012. SAS/STAT® 12.1 User’s Guide. Cary, NC: SAS Institute Inc..

Shipe, E.R., J.D. Mueller, S.A. Lewis, P.F. Williams, Jr., and J.P. Tompkins. 1997. Registration of ‘Dillon’ Soybean. Crop Sci. 37:1983

Steel R.G.D., J.H. Torrie, and D.A. Dickey.1997. Principles and Procedures of Statistics, 3d edn. McGraw-Hill, New York.

Wilcox J.R., R.M. Shibles. 2001. Interrelationships among seed quality attributes in soybean. Crop Sci. 41:11–14

Wilson, R.F. 2010. Outlook for soybeans and soybean products in 21st century markets. Lipid Technology. 22:199-202

Table 1.1 Early maturity group top 20 highest yielding lines for soybean [Glycine max (L.) Merr.] with values for protein concentration and oil concentration from population Pro 5 (G03-3385 x N06-10035)combined over locations, Clayton and Plymouth, NC 2013, compared to the highest yielding early maturity checks.

		Across 2013 locations		
Rank	**Line**	**Yield (kg ha^{-1})**	**Protein (g kg^{-1})**	**Oil (g kg^{-1})**
1	PRO5-133‡	3357.02	432.48	226.71
2	PRO5-158‡	3309.27	438.73	225.20
3	PRO5-15†	3223.55	488.06§	207.44
4	PRO5-102†	3140.13	481.15	199.48¶
5	PRO5-125‡	3134.57	445.51	222.82
6	PRO5-96	3122.38	460.05	218.33
7	PRO5-157	3108.14	432.30	218.92
8	PRO5-176‡	3104.37	438.46	220.58
9	PRO5-22	3104.36	466.69	211.65
10	PRO5-113†	3098.58	472.14	210.35
11	PRO5-145‡	3086.31	433.70	231.12§
12	PRO5-116‡	3070.81	430.29	222.02
13	PRO5-175‡	3034.33	451.68	222.33
14	PRO5-149‡	3017.50	449.42	228.64
15	PRO5-82†	3016.69	466.90	209.83
16	PRO5-4‡	2992.38	449.77	221.74
17	PRO5-88‡	2978.75	446.09	222.32
18	PRO5-134‡	2964.00	449.08	227.79
19	PRO5-98‡	2935.09	455.77	221.93
20	PRO5-2	2933.56¶	433.97	214.63
CK	N08-145	3570.73§	390.93¶	229.94
CK	5601T	3008.36	421.52	222.31

† Indicates the line is also in the top 20 for protein concentration
‡ Indicates the line is also in the top 20 for oil concentration
§ Indicates the highest value in each category
¶ Indicates the lowest value in each category
* significantly different from the check with the highest value for the trait at the 0.05 probability level (Dunnett's pairwise comparison)
** significantly different from the check with the highest value for the trait at the 0.01 probability level (Dunnett's pairwise comparison)
*** significantly different from the check with the highest value for the trait at the at the 0.001 probability level (Dunnett's pairwise comparison)

Table 1.2 Late maturity group top 20 highest yielding lines for soybean [Glycine max (L.) Merr.] with values for protein concentration and oil concentration from population Pro 5 (G03-3385 x N06-10035) grown at Clayton and Plymouth, NC 2013, and combined over locations compared to highest yielding late maturity checks.

	Clayton 2013				Plymouth 2013				Across 2013 locations			
Rank	**Line**	**Yield (kg ha^{-1})**	**Protein (g kg^{-1})**	**Oil (g kg^{-1})**	**Line**	**Yield (kg ha^{-1})**	**Protein (g kg^{-1})**	**Oil (g kg^{-1})**	**Line**	**Yield (kg ha^{-1})**	**Protein (g kg^{-1})**	**Oil (g kg^{-1})**
1	PRO5-178‡	4202.69§**	429.68	219.44	PRO5-28†‡	3186.72	467.93	222.05	PRO5-178‡	3407.2§	430.47	219.26
2	PRO5-76‡	3967.75	443.68	220.62	PRO5-99‡	3179.20	449.21	222.02	PRO5-76‡	3401.18	442.10	218.94
3	PRO5-159	3892.31	441.25	211.02	PRO5-11‡	3139.51	475.86	222.90	PRO5-99‡	3289.48	452.42	219.67
4	PRO5-156‡	3795.22	438.60	224.26	PRO5-45	3049.00	438.43	215.62	PRO5-159	3281.78	438.65	212.12
5	PRO5-147	3791.66	438.51	209.91	PRO5-27†	3020.20	481.71§	205.54	PRO5-160†	3222.52	473.99	206.50
6	PRO5-148	3781.78	468.05	211.28	PRO5-140†	3005.15	471.44	196.51	PRO5-147	3179.25	434.76	211.57
7	PRO5-84‡	3653.49	430.12	222.24	PRO5-41†	2990.13	471.94	207.13	PRO5-163†	3170.75	478.44	197.20
8	PRO5-126	3613.96	444.80	213.22	PRO5-163†	2987.63	480.74	193.05¶	PRO5-140†	3126.65	470.15	195.16¶
9	PRO5-160†	3610.81	473.92	205.35	PRO5-42‡	2871.26	435.38	221.85	PRO5-45	3118.07	434.25	217.34
10	PRO5-171	3530.83	448.67	210.03	PRO5-10†	2839.02	474.93	201.56	PRO5-80‡	3114.96	434.00	223.73
11	PRO5-165	3517.28	465.24	208.56	PRO5-76	2834.60	440.52	217.27	PRO5-77†	3112.74	469.81	208.39
12	PRO5-174	3497.17	447.86	213.32	PRO5-160†	2834.22	474.05	207.65	PRO5-151†	3103.21	480.83§	199.88
13	PRO5-80‡	3464.19	431.47	224.72§	PRO5-112	2831.52	441.71	218.83	PRO5-171	3096.49	446.59	209.56
14	PRO5-151†	3439.29	479.82§	201.60	PRO5-19	2824.80	439.82	217.36	PRO5-165	3089.43	464.04	209.05
15	PRO5-107‡	3430.44	444.55	219.54	PRO5-16	2823.24	460.93	210.21	PRO5-148†	3059.72	472.44	205.80
16	PRO5-154	3429.63	457.57	216.13	PRO5-90	2820.16	441.83	218.34	PRO5-117‡	3058.28	428.12	222.28§
17	PRO5-77†	3415.32	468.34	209.53	PRO5-77†	2810.16	471.29	207.25	PRO5-166	3043.22	429.44	210.26
18	PRO5-162†	3410.63	468.61	199.75¶	PRO5-35	2797.36	453.88	213.55	PRO5-126	3027.34	436.71	217.02
19	PRO5-117‡	3407.78	431.79	221.66	PRO5-34	2792.81	451.19	214.64	PRO5-41†	3011.04	473.25	206.66
20	PRO5-99‡	3399.76	455.63	217.31	PRO5-74	2785.73¶	440.13	209.90	PRO5-74	3008.71¶	438.65	210.24
CK	Dillon	3377.31	414.41¶	219.82	Dillon	3247.7§	398.83¶	224.12§	Dillon	3312.50	406.62¶	221.97
CK	Roy	3143.89¶	422.23	207.68	Roy	3219.01	408.50	216.19	Roy	3181.45	415.37	211.93

† Indicates the line is also in the top 20 for protein concentration
‡ Indicates the line is also in the top 20 for oil concentration
§ Indicates the highest value in each category
¶ Indicates the lowest value in each category
* significantly different from the check with the highest value for the trait at the 0.05 probability level (Dunnett's pairwise comparison)
** significantly different from the check with the highest value for the trait at the 0.01 probability level (Dunnett's pairwise comparison)
*** significantly different from the check with the highest value for the trait at the at the 0.001 probability level (Dunnett's pairwise comparison)

Table 1.3. Genotypic (above diagonal) and phenotypic (below diagonal) correlation coefficients, and their standard errors (in parentheses), between all pairs of traits (yield, protein, oil, and maturity date [R8]) for 132 soybean [Glycine max (L.) Merr.] lines from population Pro 5 grown at Clayton and Plymouth, NC, in 2013.

	Yield (kg ha^{-1})	**Protein (g kg^{-1})**	**Oil (g kg^{-1})**	**Maturity Date (R8)**
Yield (kg ha^{-1})		-0.3355†	0.4409†	-0.1691
		(0.1585)	(0.1534)	(0.1575)
Protein (g kg^{-1})	-0.0333		-0.7233†	-0.2156†
	(0.0412)		(0.0437)	(0.0864)
Oil (g kg^{-1})	0.1175†	-0.7123†		-0.2346†
	(0.0402)	(0.0370)		(0.0853)
Maturity Date (R8)	-0.0275	-0.1742†	-0.2208†	
	(0.0415)	(0.0760)	(0.0753)	

† Correlation coefficient significantly different from zero, when |correlation| > 1.96 * SE

CHAPTER 2: Genetic Mapping of Quantitative Trait Loci Associated with Seed Protein Content in a Soybean Population Derived from Germplasm Lines with Exotic Pedigree

ABSTRACT

Soybean [*Glycine max* (L.) Merr.] meal is the primary high-protein source in livestock feed, used predominately in poultry and swine production. Elevating the protein concentration of this expensive and important feed component, would benefit the livestock feed industry. The objectives of this study were to identify high protein QTL in a population derived from a cross between moderately high protein soybean genotypes G03-3385 and N06-10035, determine their relationship with oil QTL, and study the interaction of high protein QTL from different exotic sources. One hundred thirty two $F_{4:6}$ recombinant inbred lines were grown at two locations in 2013. The population was genotyped with 392 polymorphic markers and seed was analyzed by near-infrared reflectance spectroscopy to determine protein and oil concentration. Putative QTL were detected using multiple interval mapping (MIM). The across-locations analyses revealed 12 and 8 QTL significantly associated with protein and oil concentration, respectively. Although several of these were co-localized, eight of the protein QTL were independent of oil QTL and one had the same positive allele for both traits. The QTL and the epistatic interactions identified in this study explain 93.99% and 97.26% of the total phenotypic variation observed for oil and protein, respectively.

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] seed generally average 400 g kg^{-1} protein, although there are germplasm accessions with over 500 g kg^{-1} protein (Escalante and Wilcox, 1993; Panthee et al., 2005). Soybean meal, a by-product of oil extraction, provides a high-protein feed ingredient that is used predominantly in poultry and swine production. Soybean meal is an expensive component in livestock feed, and in the interest of both increasing livestock performance and profits, steps have been taken to elevate the feeding-value of soybean meal (Wilson, 2010).

Soybean seed protein is a quantitative trait having both large and small effects that can be influenced by the environment. Soybean seed protein content has been documented to have a significant negative correlation with both oil content and yield (Brim and Burton, 1979; Hyten et al., 2004). Increasing seed protein concentration while also increasing or maintaining seed oil concentration has been proven difficult because there is a negative correlation between the two traits. Recker et al. (2014) observed that pleiotropic effects rather than genetic linkage likely caused the correlation between protein and oil to remain negative in populations that were randomly mated for many generations and had linkage phase disequilibrium (LD) likely close to zero. Molecular markers can be used to identify quantitative trait loci (QTL) associated with soybean seed protein content which could be transferred into elite germplasm through the use of marker-assisted selection (MAS) programs. The effects of these individual loci on other agronomically important traits could be determined before they are transferred into elite germplasm. Currently there are more than

130 protein QTL reported on SoyBase (Fasoula et al., 2004; Hyten et al., 2004; Liang et al., 2010; Lu et al., 2012; Nichols et al., 2006; Palomeque et al., 2009; Pandurangan et al., 2012; Panthee et al., 2005; Soybase, 2014; Tajuddin et al., 2003). To detect seed composition QTL, some studies have utilized mapping populations with large parental differences. Others utilize mapping populations with plant introductions or exotic germplasm as one of the parents to detect new QTL not present in current elite lines (Hyten et al., 2004). An alternative approach to detect linkage between markers and QTL is selective genotyping of one or both phenotypic extremes of a population (Navabi et al., 2009). This method can also be used as a cost-effective way of validating QTL across populations.

In this study, two recombinant inbred line (RIL) populations derived from biparental crosses of moderately high protein and genetically unrelated germplasm lines, Pro 5 and Pro 16, were used to address the following objectives: (1) identify high protein QTL in a line derived from the high protein germplasm N6202 (Carter et al., 2010), (2) determine their relationship with oil QTL, and (3) to study the interaction of high protein QTL from different exotic sources.

MATERIALS AND METHODS

Plant Materials

Two populations (Pro 5 and Pro 16) were developed by biparental crosses performed in 2009 in Clayton, NC. F_1 plants were grown in Isabella, Puerto Rico in winter of 2009-2010. Plants

were advanced by single seed descent (Brim, 1966) to the F_4 generation. $F_{4:5}$ individual plants were harvested in Clayton, NC in the fall of 2011.

Pro 5 is a population of 132 lines derived from the cross G03-3385 x N06-10035. Pro 16 is a population of 133 lines derived from the cross G03-3621 x N06-10011. N06-10035 and N06-10011 are high protein selections from the cross 'Young' (Burton et al., 1987) x 'N6202' (Carter et al., 2010). G03-3385 and G03-3621 are high protein selections from the cross 'Benning' (Boerma et al., 1997) x Korean PI 'Danbaekkong' (PI 619083) (Kim, 1996).

Field Evaluation

In 2011, F_4 plants were grown in Clayton, NC and harvested individually. In 2012, a seed increase of $F_{4:5}$ lines were grown in un-replicated 3.7 m long one-row plots in Clayton, NC with the soil being Varina Loamy Sand. In 2012, flower color at the R2 reproductive stage, pubescence color, and maturity date at the R8 reproductive stage were recorded (Fehr and Caviness, 1977).

In 2013, 132 $F_{4:6}$ lines from Pro 5 were randomly divided into four sets for field evaluation. Within each set the lines were arranged in a randomized complete block design with three replications. The Pro 5 lines' parents G03-3385 and N06-10035, and 'N6202', a parent of N06-10035, were included as check lines. N08-145, 'NC-Roy', 'Dillon', and '5601T' were included as check lines for yield and maturity. NC-Roy is a maturity group VI cultivar (Burton et al., 2005), Dillon is a maturity group VI cultivar (Shipe et al., 1997), 5601T is a maturity group V cultivar (Pantalone et al., 2003). N08-145 is a high yielding MG

VI experimental line that was the highest yielding line within its MG in the NC Official Variety Testing in 2012 (http://www.ncovt.com/files/Soybeans2012_Table22.pdf). The tests were grown in 2013 in Plymouth, NC (planted 21 May) and Clayton, NC (planted 30 May). Soils were Portsmouth fine sandy loam and Norfolk loamy sand, respectively. Plots were three rows wide (0.97 m between rows) and 4.3 m long. Four hundred seed were planted in each plot. Plots were end-trimmed to 3.0 m and only the center row was harvested for yield and 100-seed weight determination. Flower color at the R2 reproductive stage, pubescence color, and maturity date at the R8 reproductive stage were recorded (Fehr and Caviness, 1977). Additionally, in 2013, as a validation population, the $F_{4:6}$ lines of Pro 16 were grown in un-replicated 3.7 m long one-row plots in Clayton, NC with flower color at the R2 reproductive stage, pubescence color, and maturity date at the R8 reproductive stage recorded (Fehr and Caviness, 1977). The soil was Norfolk loamy sand.

Seed Composition Analysis

In all years, protein and oil concentration analysis on a dry weight basis was analyzed by sending approximately 25 grams of seed samples from each plant/line to the USDA-ARS National Center for Agricultural Utilization Research in Peoria, IL. Analysis of the whole bean samples was completed by near infrared reflectance spectroscopy (NIR) in an Infratec 1241 Grain Analyzer (Foss North America, Inc., Eden Prairie, MN). The seed protein concentration from the 2011 and 2012 un-replicated seed increases were used to establish if there was a good correlation across years.

Molecular Marker Analysis

A single leaf was collected at random from ten plants in each $F_{4:5}$ line in 2012. A single punch was taken from each leaf using a cork borer and the punches for each plot were bulked for DNA extraction. DNA was extracted using a CTAB method (Stein et. al., 2001).

One hundred thirty two individuals of Pro 5 were genotyped using the 1536 SNPs (single nucleotide polymorphism) of the Illumina GoldenGate Universal Soy Linkage Panel (Hyten et al, 2010). Simple sequence repeat (SSR) markers were screened on the Pro 5 parents to identify polymorphic markers and to enhance the marker coverage of the 415 polymorphic SNPs from the Illumina GoldenGate Universal Soy Linkage Panel. Polymerase chain reactions (PCR) were carried out for SSR markers in a 384-well PTC-220 Peltier Thermal Cycler (MJ Research, Waltham, MA). PCR components were 1.2 µL sterile H_2O, 0.8 µL 10X PCR buffer, 1.6 µL 15 mM $MgCl_2$, 0.6 µL 3.12 mM dNTPs, 0.2 µL *taq* polymerase, 0.64 µL 1 µM M13-tailed forward primer, 0.48 µL 10 µM reverse primer, 0.48 µL 10 µM 6FAM-labeled M13, and 2 µL of 5-50 ng/µL DNA for a total reaction volume of 8 µL. Conditions for PCR were as follows: 2 minutes at 95ºC; 39 cycles of: 30 seconds at 92ºC, 30 seconds at 49ºC, and 30 seconds at 68ºC; ending with 5 minutes at 68ºC.

Selective Genotyping

In the validation population, Pro 16, a "trait based" selective genotyping analysis (Navabi et al., 2009) was used to confirm markers associated with protein concentration that were detected in the Pro 5 population and were also polymorphic in Pro 16. The analysis was

based on a normal approximation of a binomial distribution of allele frequencies, which was applied to the data obtained from bidirectional selective genotyping. Markers were confirmed to be associated with protein concentration if $|d_q| \geq Z_{(\alpha/2)}S_q$, where $|d_q|$ is the absolute value of the difference in allele frequencies between two tails, $Z_{(\alpha/2)}$ is the ordinate of the standard normal distribution such that the area under the curve from -∞ to $z_{(\alpha/2)}$ equals 1 - α/2, and S_q, the standard error of the difference between marker allele frequencies, was estimated as:

$$S_q = \sqrt{\frac{p_u q_u}{n_u} + \frac{p_1 q_1}{n_1}}$$

Where p_u and q_u are alternate allele frequencies for RILs selected in the upper tail, and p_1 and q_1 are alternant allele frequencies for RILs selected in the lower tail, and n_u and n_1 are the number of RILs in the upper and lower tails of the population, respectively (Eskandari et al., 2013b).

Primers were designed based on the BARC-061333-17169 SNP, BARC-044609-08737 SNP, BARC-057969-15031 SNP, BARC-044083-08609 SNP, and the BARC-029827-06444 SNP for use in the KBiosciences Competitive Allele Specific PCR (KASPar) SNP genotyping system (KBiosciences, Herts, UK).Three µl of DNA from each sample were transferred to a 384-well plate and air-dried overnight. Total volume per reaction was 4 ul, which consisted of 2 ul 2X KASP Reaction Mix, 0.11 ul 0.5X Assay Mix, 0.072 ul 50 mM $MgCl_2$, and 1.818 ul H2O. The 0.5X Assay Mix was prepared to a total volume of 20 ul which consisted of 1.2 uL 100 uM of each allele-specific primer, 3 ul 100 uM common reverse primer, and 14.6 uL 10 mM Tris pH 8.3. Thermocycling parameters were – 94°C

incubation for 15 minutes – 10 cycles of: 94°C for 20s, 65°C for 60s, -0.8°C/cycle – 30 cycles of: 94°C for 20s and 57°C for 60s. Endpoint fluorescence reading was performed using a Roche LightCycler® 480 (Penzberg, Germany). Allele calling was performed using Version 1.5 of the Roche LightCycler® 480 software. One SSR (Sat_345) was used for the analysis and PCR conditions and components were previously stated in the Molecular Marker Analysis section.

Statistical Analysis

Analysis of variance (ANOVA) for the protein and oil concentration was conducted using Proc MIXED (Littell et al., 2006) in SAS 9.4 (SAS Institute Inc, 2013). Locations, sets, and their interactions were considered fixed effects. The Pro 5 lines (nested within set), replicate, and the appropriate interactions with fixed effects were considered random effects. Similarly, a mixed model was used to analyze the field experiments data with each location considered separately. Set was considered a fixed effect. Line (nested within set), replicate, and the appropriate interactions with fixed effects were considered random. Best linear unbiased predictors (BLUPs) for protein and oil concentration were obtained for each line in the Pro 5 population in two analyses: across environments and separately within the two environments for the 2013 field season.

Heritability estimates of the $F_{4:}$ derived lines on a plot mean and entry mean basis, and their standard errors (Table 2.12), were obtained according to Holland et al. (2003).

Linkage and QTL Mapping

JoinMap 4 (van Ooijen, 2006) was used to identify markers showing segregation distortion in the Pro 5 marker data. Markers that displayed significant distortion were excluded from the dataset. Linkage mapping was conducted using JoinMap 4 with a minimum LOD score of 3.0 using the Kosambi (1944) mapping function. QTL analysis for protein concentration was conducted on the Pro 5 population with WinQTL cartographer v. 2.5 (Wang et al., 2012). The composite interval mapping (CIM) procedure was conducted with WinQTL Cartographer for protein concentration using BLUPs across environments and separately within the two environments. Firstly, a permutation test was conducted to obtain the $\alpha = 0.01$ genome-wide type I error rate using 1000 permutations to establish the appropriate LOD threshold for protein concentration in each environment and across environments (Churchhill and Doerge, 1994). After obtaining the appropriate LOD threshold, the "forward and backward regression" method was conducted with parameters: window size (5 cM) and walking speed (1 cM).

The CIM models for each trait and environment were used for the multiple interval mapping (MIM) procedure by scanning through the CIM results. The models were refined by using the "optimize positions", "search for new QTL" and "test existing QTL" options for both the main effects and epistatic interactions. The MIM models chosen were the minimum Bayesian Information Criterion (BIC) and "search for new QTL" was conducted until no new significant QTL were detected. The QTL effects and the corresponding proportion of

variation explained from each QTL was exported using the "summary" option. The same QTL mapping procedure was used for the oil concentration data.

RESULTS

Seed Composition Analysis

Across the 2013 environments the protein concentration in the Pro 5 population ranged from 424.6 to 491.0 g kg^{-1} and averaged 454.5 g kg^{-1}. G03-3385 averaged 443.7 g kg^{-1} across 2013 locations and N06-10035 averaged 452.5 g kg^{-1} across 2013 locations. The effect of location on total protein content was not significant in 2013. The Pro 5 and Pro 16 populations exhibited transgressive segregation for total protein content (Table 2.1 and Table 2.2). There was a high correlation noted for protein concentration among the 2013, 2012, and 2011 environments in the Pro 5 population (Table 2.3) and a moderate correlation for protein concentration between the 2013, 2012, and 2011 locations in the Pro 16 population (Table 2.4).

Linkage Mapping

A genetic linkage map (Figure 1) for the Pro 5 population was created using 392 polymorphic markers (14 SSRs and 378 SNPs) on 34 linkage groups with an approximate total length of 1752 cM. Broad coverage was provided by the markers with the exception of chromosome 2 with only 5 markers, chromosome 4 with 11 markers, chromosome 17 with 4 markers, and chromosome 19 with 8 markers. These gaps could not be narrowed with SSR markers because of lack of polymorphisms between the parents. The average number of

markers per chromosome was 11.53. The order of the markers and relative distance of the markers within the linkage groups did not demonstrate significant discrepancies from the consensus map (Hyten et al., 2010).

QTL Mapping of Protein Concentration

In the combined location analysis, there were a total of twelve QTL identified as associated with protein concentration, located on ten different chromosomes (Table 2.7). The individual phenotypic variances accounted for by each QTL ranged from 0.2% to 55.5%. Seven of the high protein QTL alleles were donated by N06-10035 and five by G03-3385. Four of these loci were also detected in both Clayton and Plymouth in the within location analyses (chromosomes 9, 11, 15, and 20). The QTL located on chromosome 9 and 15 had the positive allele donated by N06-10035 and the QTL located on chromosome 11 and 20 had the positive allele donated by G03-3385. From the remaining loci, one was common only to the Clayton location and five were common only to the Plymouth location.

Five additional QTL were detected only in the 2013 Clayton location on four different chromosomes (Table 2.5). Two of the high protein QTL alleles were contributed by N06-10035 and three of the high protein QTL alleles by G03-3385. The 2013 Plymouth environment had only two unique QTL associated with protein located on 2 different chromosomes (Table 2.6). Both of them had positive high protein alleles contributed by G03-3385. Significant epistatic interactions were detected in the across and within location analyses (Tables 2.5 – 2.7). The effect of these interactions ranged from -1.4 to 6.0.

QTL Mapping of Oil Concentration

Across environments there were a total of eight QTL identified as associated with oil concentration, located on six different chromosomes (Table 2.10). The individual phenotypic variances accounted for by each QTL ranged from 0.5% to 46.9%. Four of these were detected in both locations in the within location analyses and one was common only to the Plymouth. Four of the positive oil alleles were donated by G03-3385 and four from N06-10035. Five additional QTL were identified only in the 2013 Clayton location on five different chromosomes (Table 2.8). Two of the high oil QTL alleles were contributed by N06-10035 and the other three from G03-3385. In the Plymouth 2013 location six unique QTL associated with high oil were detected (Table 2.9). Four of the high oil alleles were contributed by N06-10035 and the other two from G03-3385. Significant epistatic interactions were detected in the across and within location analyses (Tables 2.8 – 2.10). The effect of these interactions ranged from 0.2 to 7.6.

Selective Genotyping

Six markers associated with significant protein QTL across 2013 locations in Pro 5 were also polymorphic between the parents of the Pro 16 population. Among these 6 markers, two were also confirmed as significantly associated with protein concentration in Pro 16 using bi-directional selective genotyping (Table 2.11).

Heritability

Heritability on an entry mean basis and on a plot mean basis, and their standard errors were calculated for protein concentration, and oil concentration (Table 2.12). Both seed protein and oil concentration had high heritability estimates.

DISCUSSION

The major high protein QTL on linkage group 20 that had the G03-3385 allele as a favorable allele was also detected in a previous study involving a Danbaekong derived high protein line (Warrington et al. 2011). This QTL had the largest effect in both studies. The QTL detected in this study with favorable alleles from G03-3385 and not in Warrington's et al. (2011) could be due to differences in the QTL analysis (CIM vs. MIM in this study), because of insufficient marker coverage in the Warrington research, or because some of the QTL in G03-3385 are from Benning and not Danbaekong.

In Clayton 2013, out of the 10 protein QTL, three are not co-localized with oil QTL. Among the QTLs that coincided, one had a positive allele for protein that also had a positive effect on oil. Three high protein QTL are in regions where no other protein QTL had been reported before on Soybase (2014) (www.soybase.org, verified 13 October 2014). In Plymouth 2013, out of the 10 protein QTL, eight are not co-localized with oil QTL. Among the QTL that coincided, none had positive alleles that also had a positive effect on oil. Three high protein QTL are in regions where no other QTL have been reported before on Soybase (2014) (www.soybase.org, verified 13 October 2014) using the GMComposite_2003 version

of the map. Across 2013 locations, out of the 12 protein QTL, eight are not co-localized with oil QTL. Among the QTL that coincided, none had positive alleles that also had a positive effect on oil. Four QTL, those located on chromosomes 3, 7, 14, and 15, are in regions where no other QTL have been reported before. On Soybase (2014) (www.soybase.org, verified 13 October 2014) there are a number of protein and protein-related QTL within 10cM upstream or downstream of those detected in this study: chromosome 1 (seed protein 3-5), chromosome 8 (seed protein 3-1, 21-1), chromosome 9 (seed protein 5-3, 35-4, 34-6), chromosome 11 (seed protein 34-7, 3-2, 16-1, 25-1, and 25-2), chromosome 12 (seed protein 34-8), and chromosome 20 (seed protein 26-4, 30-1, 1-4, 1-3, 11-1, 3-12, 10-1, 34-11, 31-1, 17-1, 15-1, 1-1, 1-2, 26-9 and cqSeed protein-003). The positive epistatic interactions of protein QTL with small effects observed in this study could contribute to the development of moderately high protein lines with adequate oil content. Despite their relative small effects, six of them were observed across environments and could be targeted for selection using molecular markers.

The protein QTL detected on chromosome 9, 11, 15, and 20 were detected in the Clayton, Plymouth, and across location analysis, which indicates they are stable across environments and are good targets for selection. The protein QTL with the largest effect was on chromosome 20 at BARC03815709983 and the location is consistent with the QTL found by Warrington et al. (2011). In general it was observed that the QTL with the largest effect for either trait also had a negative effect on the other; however, there were some exceptions: The positive allele for the protein QTL identified on chromosome 9 also had a significantly

favorable effect on total oil content in the Clayton analysis and a slightly favorable (but not significant effect on oil in Plymouth or across locations. The QTL detected on chromosome 11 did not coincide with oil QTL.

Out of the six QTL loci that were tested with selective genotyping in the validation population (Pro 16), only two met the threshold to be declared significant. These were not the QTL with the largest effect in Pro 5. Differences could also be due to the fact that the parental lines (N06-10035 and N06-10011) are sister lines but they are not identical and were selected only based on phenotypic data. Although they were both high protein sister lines, they might not carry the exact same high protein QTL.

The broad-sense heritability estimates calculated on a plot mean and entry mean basis were high for protein and oil concentrations and similar to those reported in previous studies (Hyten et al., 2004, Eskandari et al., 2013a, Eskandari et al., 2013b, and Recker et al., 2013). The heritability estimates for protein and oil concentration indicate that most of the phenotypic variation was genetic which suggests that genetic gain could be easily achieved through phenotypic selection of these traits.

In conclusion, the high protein QTL detected within location and across locations would be good targets for selection as they are stable across environments. Within those, the QTL detected on chromosome 9 was found to coincide with putative oil QTL and in the Clayton 2013 analysis it had a positive effect on both traits. This high protein QTL on chromosome 9, along with the others that were detected in the within location and across

location analyses and did not coincide with oil QTL, could be good targets for selection for high protein without negatively affecting oil.

REFERENCES

Boerma, H.R., R.S. Hussey,D.V. Phillips, E.D.Wood,G.B. Rowan, and S.L. Finnerty. 1997. Registration of 'Benning' soybean. Crop Sci. 37:1982.

Brim, C.A. 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6:220.

Brim, C.A. and J.W. Burton. 1979. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Sci. 19:494-498.

Burton, J.W., C.A. Brim, and M.F. Young. 1987. Registration of 'Young' soybean. Crop Sci. 27:1093.

Burton, J.W., T.E. Carter, Jr., and D.T. Bowman. 2005. Registration of NC-Roy soybean. Crop Science. 45:2654.

Carter, T.E. Jr., P.E. Rzewnicki, J.W. Burton, M.R. Villagarcia, D.T. Bowman, E. Taliercio, and P. Kwanyuen. 2010. Registration of N6202 Soybean Germplasm with High Protein, Favorable Yield Potential, Large Seed, and Diverse Pedigree. Journal of Plant Registrations. 4:73-79

Churchill, G.A. and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963-971.

Escalante, E.E. and J.R. Wilcox. 1993. Variation in seed protein among nodes of normal- and high-protein soybean genotypes. Crop Sci. 33:1164-1166.

Eskandari, M., and E.R. Cober. 2013a. Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. Theor Appl Genet. 126:1677-1687

Eskandari, M., E.R. Cober, and I. Rajcan. 2013b. Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents. Theor Appl Genet. 126:483-495.

Fasoula, Vasilia A., Donna K. Harris, H. Roger Boerma. 2004. Validation and Designation of Quantitative Trait Loci for Seed Protein, Seed Oil, and Seed Weight from Two Soybean Populations. Crop Sci. 44(4):1218-1225.

Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. In Agric. And Home Economics Exp. Stn. Special Rep. no. 80. Iowa State Univ., Ames.

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martinez. 2003. Estimating and interpreting heritability for plant breeding: An update. Plant Breed. Rev. 22:9-112.

Hyten, D.L., V.R. Pantalone, C.E. Sams, A.M. Saxton , D. Landau-Ellis, T.R. Stefaniak, M.E. Schmidt. 2004. Seed quality QTL in a prominent soybean population. Theor Appl Genet. 109:552-561

Hyten, D.L., I. Choi, Q. Song, J.E. Specht, T.E. Carter Jr., R.C. Shoemaker, E. Hwang, L.K. Matukumalli, P.B.Cregan. 2010 A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping. Crop Sci. 50:960-968

Kim S.-D., E.-H. Hong, Y.-H. Kim, S.-H. Lee, Y.-K. Seong, K.-Y. Park, Y.-H. Lee, Y.-H. Hwang, E.-H. Park, H.-S. Kim, Y.-H. Ryu, R.-K. Park, and Y.-S. Kim. (1996) A new high protein and good seed quality soybean variety "Danbaegkong". RDA Journal of Agricultural Science- Upland and Industrial Crops 38:228-232.

Kosambi D. (1944) The estimation of map distances from recombination values. Ann. Eugen 12:172-175.

Liang, H., Y. Yu, S. Wang, Y. Lian, T. Wang, Y. Wei, P. Gong, X. Liu, X. Fang, M. Zhang. 2010. QTL Mapping of Isoflavone, Oil and Protein Contents in Soybean (Glycine max L. Merr.). Ag. Sci. China. 9(8): 1108-1116.

Littell, R. C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberber. 2006. SAS for Mixed Models. 2nd ed. SAS Institute, Inc., Cary, NC.

Lu, W., Z. Wen, H. Li, D. Yuan, J. Li, H. Zhang, Z. Huang, S. Cui, W. Du. 2012. Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean. Theor. Appl. Gen. Online First.

Navabi A., D.E. Mather, J. Bernier, D.M. Spaner, G.N. Atlin. 2009. QTL detection with bidirectional and unidirectional selective genotyping: marker-based and trait-based analysis. Theor Appl Genet 118:347-358

Nichols, D.M., K.D. Glover, S.R. Carlson, J.E. Specht, B.W. Diers. 2006. Fine Mapping of a Seed Protein QTL on Soybean Linkage Group I and Its Correlated Effects on Agronomic Traits. Crop Sci. 46(2):834-839

Palomeque, L., L. Li-Jun, W. Li, B. Hedges, E. Cober, I. Rajcan. 2009. QTL in mega-environments: II. Agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted x high-yielding exotic soybean lines. Theor. Appl. Genet. 119(3):429-436

Pandurangan, S., A. Pajak, S. Molnar, E. Cober, S. Dhaubhadel, C. Hernandez-Sebastia, W. Kaiser, R. Nelson, S. Huber, F. Marsolais. 2012. Relationship between asparagine metabolism and protein concentration in soybean seed. J. Exp. Bot.

Pantalone, V.R., F.L. Allen, and D. Landau-Ellis. 2003. Registration of '5601T' Soybean. Crop Science. 43:1123-1124

Panthee, D.R., V.R. Pantalone, D.R. West, A.M. Saxton, and C.E. Sams. 2005. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Sci. 45:2015-2022.

Recker, J.R., J.W Burton, A. Cardinal, L.M. Miranda. 2013. Analysis of Quantitative Traits in Two Long-Term Randomly Mated Soybean Populations: I. Genetic Variances. Crop Sci. 53:1375-1383

Recker, J.R., J.W Burton, A. Cardinal, L.M. Miranda. 2014. Genotypic and Phenotypic Correlations of Quantitative Traits in Two Long-Term, Randomly Mated Soybean Populations. Crop Sci. 54:939-943

SAS Institute Inc. 2012. SAS/STAT® 12.1 User's Guide. Cary, NC: SAS Institute Inc.

Shipe, E.R., J.D. Mueller, S.A. Lewis, P.F. Williams, Jr., and J.P. Tompkins. 1997. Registration of 'Dillon' Soybean. Crop Science 37:1983

Soybase. 2014. Soybean Breeder's Toolbox Genetic Map Information. http://soybase.org/sbt/ (accessed 27 Jan. 2014)

Stein N., G. Herren, B. Keller. 2001. A new DNA extraction method for high-throughput marker analysis in a large genome species such as *Triticum aestivum*. Plant Breeding. 120:354-356

Tajuddin, T., S. Watanabe, N. Yamanaka, K. Harada. 2003. Analysis of Quantitative Trait Loci for Protein and Lipid Contents in Soybean Seeds Using Recombinant Inbred Lines. Breed. Sci. 53(2):133-140

van Ooijen, J.W. 2006. JoinMap 4: Software for the calculation of genetic linkage maps in experimental populations. Plant Research International B.V and Kyazma B.V., Wageningen, Netherlands.

Wang S., C. J. Basten, and Z.-B. Zeng. 2012. Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Warrington C.V., H.R. Boerma, E.C. Brummer, J.N. All, P.L. Raymer, D.A. Knauft. 2011. QTL Mapping And Optimum Resource Allocation For Enhancing Amino Acid Content In Soybean. University of Georgia.

Wilson, R.F. 2010. Outlook for soybeans and soybean products in 21st century markets. Lipid Technology. 22:199-202

Table 2.1. Protein concentration means for N06-10035, G03-3385, and for the 132 G03-3385 × N06-10035 RILs in all years and locations

Environment		Parents		F_4-derived RILs			
Location	**Year**	**N06-10035**	**G03-3385**	**Mean**	**Min**	**Max**	**Standard deviation**
		g kg^{-1}					
Clayton	2011	-	-	424.9	370.8	471.5	17.68
Clayton	2012	450.2	440.5	457.3	414.3	498.6	20.05
Clayton	2013	455.1	442.1	454.8	424.6	488.2	14.78
Plymouth	2013	449.8	445.3	453.8	423.9	495.8	16.40

Table 2.2. Protein concentrations for N06-10011, G03-3621, and for the 133 G03-3621 × N06-10011 RILs in all years

Environment		Parents		F_4-derived RILs			
Location	**Year**	**N06-10011**	**G03-3621**	**Mean**	**Min**	**Max**	**Standard deviation**
		g kg^{-1}					
Clayton	2011	-	-	412.6	376.3	447.7	14.08
Clayton	2012	446.3	444.5	443.6	414.0	470.9	12.21
Clayton	2013	445.6	433.6	438.1	397.1	467.9	14.28

Table 2.3. Protein concentration correlation coefficients for G03-3385 x N06-10035 RIL (Pro 5) population across all years

Year	2013	2012	2011
2013	1.00000	0.86276	0.78830
2012	0.86276	1.00000	0.70264
2011	0.78830	0.70264	1.00000

Table 2.4. Protein concentration correlation coefficients for G03-3621 x N06-10011 RIL (Pro 16) population across all years

Year	2013	2012	2011
2013	1.00000	0.52617	0.34380
2012	0.52617	1.00000	0.53582
2011	0.34380	0.53582	1.00000

Table 2.5. Estimates of increased protein QTL positions, effects and interactions identified by multiple interval mapping from Clayton 2013 location for the G03-3385 x N06-10035 RIL (Pro 5) population

QTL (pair)	Type	Chromosome	Marker	Position	LOD	Effect	Effect %
1	A	8.2	BARC01766503101	16.3	1.39	0.1824	2.0
2	A	8.1	BARC02761406619	38.6	1.91	-0.264	2.3
3	A	9	BARC04460908737	77.5	7.47	0.4659	0.9
5	A	11.1	BARC01461101591	22.2	7.15	-0.501	8.1
6	A	11.2	BARC05985116137	28	4.74	-0.3315	3.9
7	A	12	Sat_118	71.7	3.33	-0.2583	3.3
8	A	14.2	BARC01336500489	6	1.5	0.8087	3.0
9	A	15.1	BARC06127717149	12.4	3.67	0.3052	3.4
10	A	20.2	BARC03815709983	19.4	31.11	-1.2324	57.6
3 x6	AA				2.352	0.2507	1.2
5 x7	AA				2.428	-0.2175	0.4
2 x10	AA				1.994	0.2515	-1.4

Table 2.6. Estimates of increased protein QTL positions, effects and interactions identified by multiple interval mapping from Plymouth 2013 location for the G03-3385 x N06-10035 RIL (Pro 5) population

QTL (pair)	Type	Chromosome	Marker	Position	LOD	Effect	Effect %
1	A	1	BARC05513113049	49.912	16.76	0.5478	8.7
2	A	3	BARC01017900543	46.642	9.86	0.3447	1.3
3	A	6.1	BARC06325918282	64.5121	12.52	-0.5198	7.0
4	A	7.1	BARC04166708063	21.516	3.19	-0.1471	1.6
5	A	8.3	BARC05488712192	18.653	17.51	-0.5626	9.8
6	A	9	BARC04460908737	81.4060	12.16	0.3299	1.7
7	A	11.1	BARC01461101591	39.149	9.35	-0.3342	3.3
8	A	14.2	BARC02855305949	6.8	13.94	0.4128	0.8
9	A	15.1	BARC05796915031	13.12	12.14	0.2922	3.7
10	A	20.2	BARC03815709983	20.387	39.36	-1.1343	48.0
3 x 5	AA				3.69	0.1282	-0.4
2 x 7	AA				3.079	0.1663	0.6
3 x 7	AA				4.115	0.0864	0.2
4 x 7	AA				8.861	-0.3349	2.5
6 x 7	AA				5.217	-0.1472	-0.6
2 x 9	AA				12.295	-0.391	6.0
8 x 9	AA				5.417	-0.1742	0.9

Table 2.7. Estimates of increased protein QTL positions, effects and interactions identified by multiple interval mapping from across locations (Clayton 2013 and Plymouth 2013) location for the G03-3385 x N06-10035 RIL (Pro 5) population

QTL (pair)	Type	Chromosome	Marker	Position	LOD	Effect	Effect %
1	A	1	BARC05513113049	50.0	13.53	0.8635	13.6
2	A	3	BARC02146504122	39.6	6.63	0.2915	-1.2
3	A	7.2	BARC00732000155	50.8	6.22	0.3217	1.1
4	A	8.3	BARC05488712192	18.5999	2.14	-0.1089	1.6
5	A	9	BARC04460908737	78.4	13.76	0.4896	3.1
6	A	11.1	BARC01461101591	22.2	10.46	-0.3403	4.1
7	A	11.2	Satt197	1.0	6.62	-0.3921	3.1
8	A	12	BARC05590713843	75.2	7.20	-0.2774	3.0
9	A	14.2	BARC02855305949	6.9	8.48	0.2380	0.2
10	A	15.1	BARC05796915031	13.2	13.91	0.5090	5.9
11	A	20.2	BARC03815709983	19.4	36.87	-1.2095	55.5
12	A	20.1	BARC02930106148	40.1	7.09	0.3080	4.4
1 x 5	AA				1.892	-0.0901	0.2
3 x 5	AA				5.17	-0.1893	2.1
6 x 8	AA				3.145	-0.1286	0.0
3 x 10	AA				2.164	0.1115	-0.4
7 x 11	AA				4.494	0.1546	0.9
11 x 12	AA				5.03	-0.1734	1.1

Table 2.8. Estimates of increased oil QTL positions, effects and interactions identified by multiple interval mapping from Clayton 2013 location for the G03-3385 x N06-10035 RIL (Pro 5) population

QTL (pair)	Type	Chromosome	Marker	Position	LOD	Effect	Effect %
1	A	8.1	BARC02761406619	40.6	7.24	0.2253	7.4
2	A	8.1	BARC06366318423	78.9	-0.15	-0.276	1.2
3	A	9	BARC05553313402	70	4.73	0.1622	4.8
4	A	11.2	Satt197	9	16.8	0.3989	23.9
5	A	12	BARC04920910821	97.4	4.21	-0.475	14
6	A	14.2	BARC02202504259	28	2.37	-0.0878	-1.1
7	A	15.1	BARC05796915031	14.1	5.76	-0.1504	2.1
8	A	16	BARC02815905778	41.3	2.02	-0.0702	1.8
9	A	20.2	BARC03815709983	18.5	28.42	0.5112	41.3
2 x 8	AA				1.337	0.0573	0.6

Table 2.9. Estimates of increased oil QTL positions, effects and interactions identified by multiple interval mapping from Plymouth 2013 location for the G03-3385 x N06-10035 RIL (Pro 5) population

QTL (pair)	Type	Chromosome	Marker	Position	LOD	Effect	Effect %
1	A	1	BARC03888307384	1	4.13	0.1889	2.7
2	A	1	BARC03097306982	30.2	4.15	-0.2186	4.1
3	A	3	BARC05712914594	15.6	4.71	-0.1834	0.4
4	A	8.1	BARC05985316139	9.8	5.41	0.2419	5.9
5	A	8.1	BARC06366318423	78.9	3.73	-0.1664	0.4
6	A	11.2	Satt197	11	11.16	0.3365	13.2
7	A	11.2	BARC03537907181	91.2	4.22	0.1932	2.2
8	A	13.2	BARC04164708054	45.4	5.57	-0.2358	8
9	A	15.1	BARC05796915031	13.2	5.39	-0.231	3.5
10	A	16	BARC01602702038	0.1	2.63	-0.1507	2.7
12	A	20.2	BARC03815709983	19.4	2.14	0.2864	20.6
4 x 6	AA				1.549	0.1201	0.6
3 x 9	AA				2.556	-0.1551	2.1

Table 2.10. Estimates of increased oil QTL positions, effects and interactions identified by multiple interval mapping from across locations (Clayton 2013 and Plymouth 2013) location for the G03-3385 x N06-10035 RIL (Pro 5) population

QTL (pair)	Type	Chromosome	Marker	Position	LOD	Effect	Effect %
1	A	8.1	BARC05985316139	10.8	5.59	0.1451	3.8
2	A	8.1	BARC06366318423	78.9	5.8	-0.139	0.5
3	A	11.2	Satt197	10	18.81	0.4008	19.7
4	A	13.1	Sat_039	5	6.68	0.1502	2.8
5	A	13.2	BARC06386318477	27.8	4.97	-0.145	4.7
6	A	14.2	BARC02202504259	28	11.2	-0.2605	-2.4
7	A	15.1	BARC02822105799	18.1	8.44	-0.4139	4.8
8	A	20.2	BARC03815709983	18.4	32.54	0.5983	46.9
1 x 4	AA				3.283	0.1014	0.2
4 x 6	AA				4.31	0.1269	2.4
5 x 6	AA				7.946	0.2023	7.6

Table 2.11. Putative QTL associated with soybean seed protein concentration confirmed using a 'trait-based' bidirectional selective genotyping analysis (Navabi et al, 2009) of Pro 16.

Locus	Chr	Allele frequency				D_q^c	S_q^d	P value
		High protein		Low protein				
		G_a	N_b	G_a	N_b			
Sat_345	1	0.571429	0.428571	0.305556	0.694444	0.265873	0.11354	≤0.05
BARC-044609-08737	9	0.324324	0.675676	0.636364	0.363636	0.312039	0.113732	≤0.05

[a] The frequency of the allele shared by G03-3621

[b] The frequency of the allele shared by N06-10011

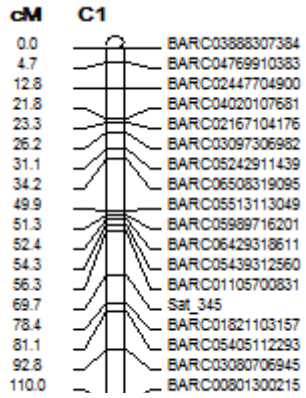
[c] The absolute value of the difference in marker allele frequencies

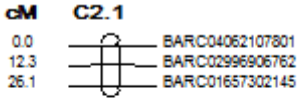
[d] The standard error of the difference between marker allele frequencies

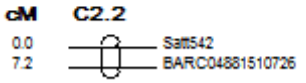
Table 2.12. Broad-sense heritability (H^2) estimates for both entry mean and plot basis with their standard errors for yield, seed protein concentration, and seed oil concentration for G03-3385 x N06-10035 RIL (Pro 5) population across locations (Clayton, NC 2013 and Plymouth, NC 2013)

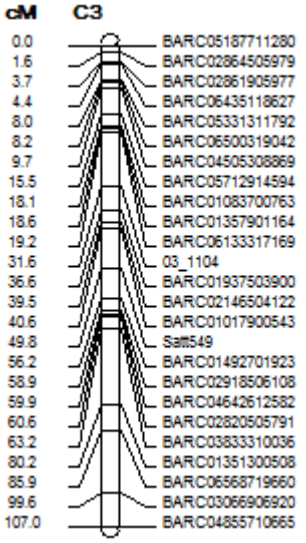
Trait	H^2 (SE) entry	H^2 (SE) plot
Protein	0.939 (0.01)	0.798 (0.02)
Oil	0.951 (0.009)	0.822 (0.02)

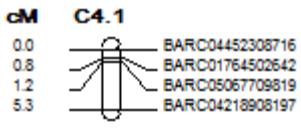
Figure 1. Linkage Map of the G03-3385 x N06-10035 RIL population (Pro 5)

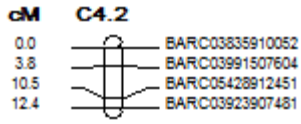

cM C1
0.0 BARC03888307384
4.7 BARC04769910383
12.8 BARC02447704900
21.8 BARC04020107681
23.3 BARC02167104176
26.2 BARC03097306982
31.1 BARC05242911439
34.2 BARC06508319095
49.9 BARC05513113049
51.3 BARC05989716201
52.4 BARC06429318611
54.3 BARC05439312560
56.3 BARC01105700831
69.7 Sat_345
78.4 BARC01821103157
81.1 BARC05405112293
92.8 BARC03080706945
110.0 BARC00801300215

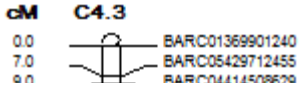

cM C2.1
0.0 BARC04062107801
12.3 BARC02996906762
26.1 BARC01657302145

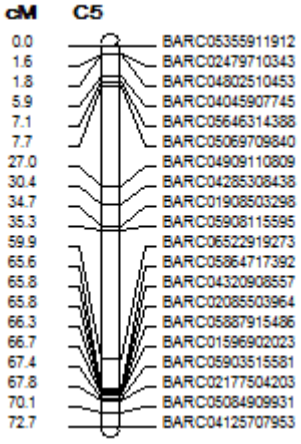

cM C2.2
0.0 Satt542
7.2 BARC04881510726

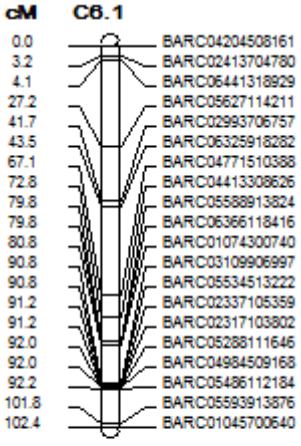

cM C3
0.0 BARC05187711280
1.6 BARC02864505979
3.7 BARC02861905977
4.4 BARC06435118627
8.0 BARC05331311792
8.2 BARC06500319042
9.7 BARC04505308869
15.5 BARC05712914594
18.1 BARC01083700763
18.6 BARC01357901164
19.2 BARC06133317169
31.6 03_1104
36.6 BARC01937503900
39.5 BARC02146504122
40.6 BARC01017900543
49.8 Satt549
56.2 BARC01492701923
58.9 BARC02918506108
59.9 BARC04642612582
60.6 BARC02820505791
63.2 BARC03833310036
80.2 BARC01351300508
85.9 BARC06568719660
99.6 BARC03066906920
107.0 BARC04855710665

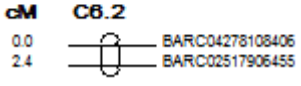

cM C4.1
0.0 BARC04452308716
0.8 BARC01764502642
1.2 BARC05067709819
5.3 BARC04218908197

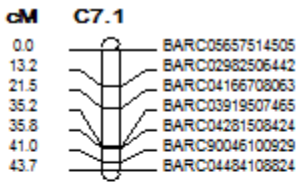

cM C4.2
0.0 BARC03835910052
3.8 BARC03991507604
10.5 BARC05428912451
12.4 BARC03923907481

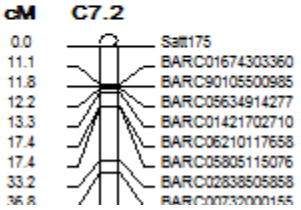

cM C4.3
0.0 BARC01369901240
7.0 BARC05429712455

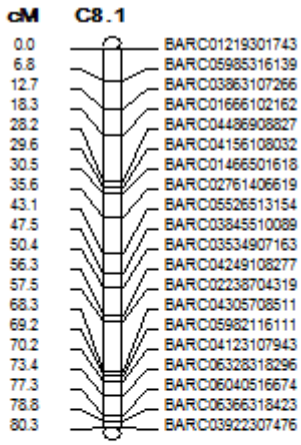

cM C5
0.0 BARC05355911912
1.6 BARC02479710343
1.8 BARC04802510453
5.9 BARC04045907745
7.1 BARC05646314388
7.7 BARC05069709840
27.0 BARC04909110809
30.4 BARC04285308438
34.7 BARC01908503298
35.3 BARC05908115595
59.9 BARC06522919273
65.6 BARC05864717392
65.8 BARC04320908557
65.8 BARC02085503964
66.3 BARC05887915486
66.7 BARC01596902023
67.4 BARC05903515581
67.8 BARC02177504203
70.1 BARC05084909931
72.7 BARC04125707953

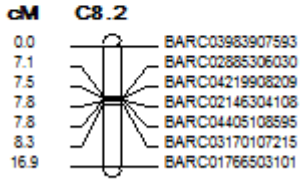

cM C6.1
0.0 BARC04204508161
3.2 BARC02413704780
4.1 BARC06441318929
27.2 BARC05627114211
41.7 BARC02993706757
43.5 BARC06325918282
67.1 BARC04771510388
72.8 BARC04413308626
79.8 BARC05588913824
79.8 BARC06366118416
80.8 BARC01074300740
90.8 BARC03109906997
90.8 BARC05534513222
91.2 BARC02337105359
91.2 BARC02317103802
92.0 BARC05288111646
92.0 BARC04984509168
92.2 BARC05486112184
101.8 BARC05593913876
102.4 BARC01045700640

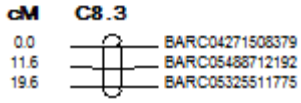

cM C6.2
0.0 BARC04278108406
2.4 BARC02517906455

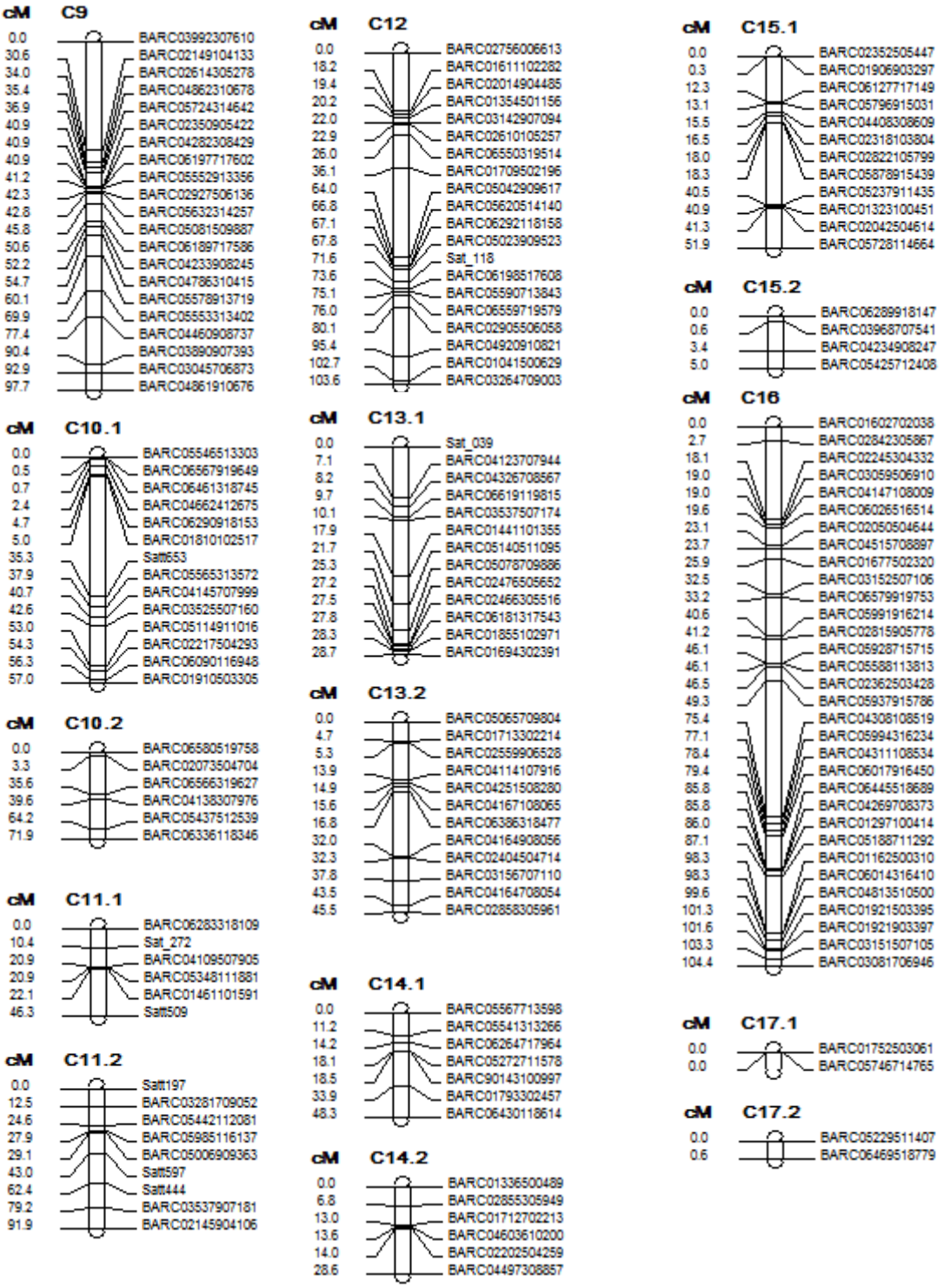

cM C7.1
0.0 BARC05657514505
13.2 BARC02982506442
21.5 BARC04166708063
35.2 BARC03919507465
35.8 BARC04281508424
41.0 BARC90046100929
43.7 BARC04484108824

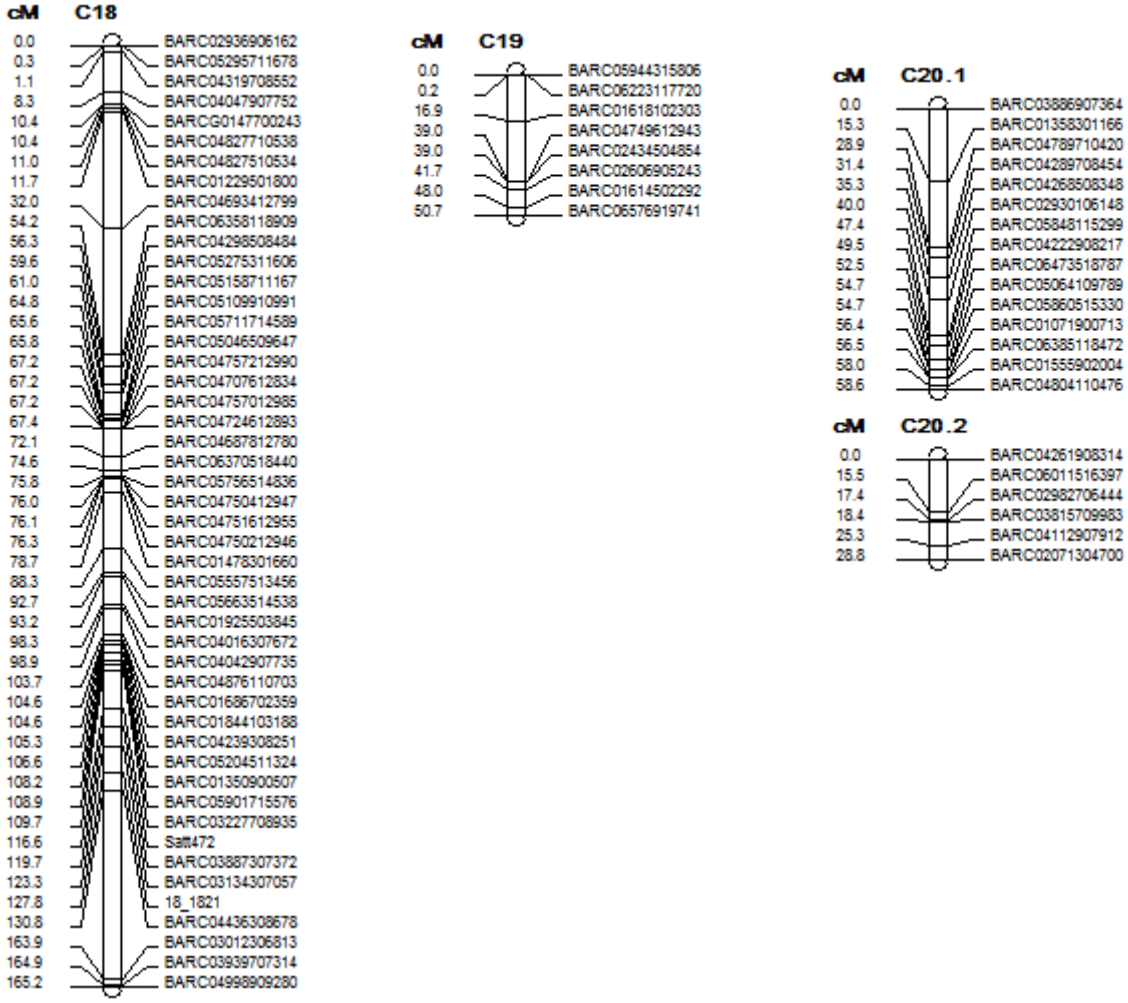

cM C7.2
0.0 Satt175
11.1 BARC01674303360
11.8 BARC90105500985
12.2 BARC05634914277
13.3 BARC01421702710
17.4 BARC06210117658
17.4 BARC05805115076
33.2 BARC02838505858


cM C8.1
0.0 BARC01219301743
6.8 BARC05985316139
12.7 BARC03863107266
18.3 BARC01666102162
28.2 BARC04486908827
29.6 BARC04156108032
30.5 BARC01466501618
35.6 BARC02761406619
43.1 BARC05526513154
47.5 BARC03845510089
50.4 BARC03534907163
56.3 BARC04249108277
57.5 BARC02238704319
68.3 BARC04305708511
69.2 BARC05982116111
70.2 BARC04123107943
73.4 BARC06328318296
77.3 BARC06040516674
78.8 BARC06366318423
80.3 BARC03922307476


cM C8.2
0.0 BARC03983907593
7.1 BARC02885306030
7.5 BARC04219908209
7.8 BARC02146304108
7.8 BARC04405108595
8.3 BARC03170107215
16.9 BARC01766503101


cM C8.3
0.0 BARC04271508379
11.6 BARC05488712192
19.6 BARC05325511775


cM C9
0.0 BARC03992307610
30.6 BARC02149104133
34.0 BARC02614305278
35.4 BARC04862310678
36.9 BARC05724314642
40.9 BARC02350905422
40.9 BARC04282308429
40.9 BARC06197717602
41.2 BARC05552913356
42.3 BARC02927506136
42.8 BARC05632314257
45.8 BARC05081509887
50.6 BARC06189717586
52.2 BARC04233908245
54.7 BARC04786310415
60.1 BARC05578913719
69.9 BARC05553313402
77.4 BARC04460908737
90.4 BARC03890907393
92.9 BARC03045706873
97.7 BARC04861910676
cM C10.1
0.0 BARC05546513303
0.5 BARC06567919649
0.7 BARC06461318745
2.4 BARC04662412675
4.7 BARC06290918153
5.0 BARC01810102517
35.3 Satt553
37.9 BARC05565313572
40.7 BARC04145707999
42.6 BARC03525507160
53.0 BARC05114911016
54.3 BARC02217504293
56.3 BARC06090116948
57.0 BARC01910503305
cM C10.2
0.0 BARC06580519758
3.3 BARC02073504704
35.6 BARC06566319627
39.6 BARC04138307976
64.2 BARC05437512539
71.9 BARC06336118346
cM C11.1
0.0 BARC06283318109
10.4 Sat_272
20.9 BARC04109507905
20.9 BARC05348111881
22.1 BARC01461101591
46.3 Satt509
cM C11.2
0.0 Satt197
12.5 BARC03281709052
24.6 BARC05442112081
27.9 BARC05985116137
29.1 BARC05006909363
43.0 Satt597
62.4 Satt444
79.2 BARC03537907181
91.9 BARC02145904106
cM C12
0.0 BARC02756006613
18.2 BARC01611102282
19.4 BARC02014904485
20.2 BARC01354501156
22.0 BARC03142907094
22.9 BARC02610105257
26.0 BARC06550319514
36.1 BARC01709502196
64.0 BARC05042909617
66.8 BARC05620514140
67.1 BARC06292118158
67.8 BARC05023909523
71.6 Sat_118
73.6 BARC06198517608
75.1 BARC05590713843
76.0 BARC06559719579
80.1 BARC02905506058
95.4 BARC04920910821
102.7 BARC01041500629
103.6 BARC03264709003
cM C13.1
0.0 Sat_039
7.1 BARC04123707944
8.2 BARC04326708567
9.7 BARC06619119815
10.1 BARC03537507174
17.9 BARC01441101355
21.7 BARC05140511095
25.3 BARC05078709886
27.2 BARC02476505652
27.5 BARC02466305516
27.8 BARC06181317543
28.3 BARC01855102971
28.7 BARC01694302391
cM C13.2
0.0 BARC05065709804
4.7 BARC01713302214
5.3 BARC02559906528
13.9 BARC04114107916
14.9 BARC04251508280
15.6 BARC04167108065
16.8 BARC06386318477
32.0 BARC04164908056
32.3 BARC02404504714
37.8 BARC03156707110
43.5 BARC04164708054
45.5 BARC02858305961
cM C14.1
0.0 BARC05567713598
11.2 BARC05541313266
14.2 BARC06264717964
18.1 BARC05272711578
18.5 BARC90143100997
33.9 BARC01793302457
48.3 BARC06430118614
cM C14.2
0.0 BARC01336500489
6.8 BARC02855305949
13.0 BARC01712702213
13.6 BARC04603610200
14.0 BARC02202504259
28.6 BARC04497308857
cM C15.1
0.0 BARC02352505447
0.3 BARC01906903297
12.3 BARC06127717149
13.1 BARC05796915031
15.5 BARC04408308609
16.5 BARC02318103804
18.0 BARC02822105799
18.3 BARC05878915439
40.5 BARC05237911435
40.9 BARC01323100451
41.3 BARC02042504614
51.9 BARC05728114664
cM C15.2
0.0 BARC06289918147
0.6 BARC03968707541
3.4 BARC04234908247
5.0 BARC05425712408
cM C16
0.0 BARC01602702038
2.7 BARC02842305867
18.1 BARC02245304332
19.0 BARC03059506910
19.0 BARC04147108009
19.6 BARC06026516514
23.1 BARC02050504644
23.7 BARC04515708897
25.9 BARC01677502320
32.5 BARC03152507106
33.2 BARC06579919753
40.6 BARC05991916214
41.2 BARC02815905778
46.1 BARC05928715715
46.1 BARC05588113813
46.5 BARC02362503428
49.3 BARC05937915786
75.4 BARC04308108519
77.1 BARC05994316234
78.4 BARC04311108534
79.4 BARC06017916450
85.8 BARC06445518689
85.8 BARC04269708373
86.0 BARC01297100414
87.1 BARC05188711292
98.3 BARC01162500310
98.3 BARC06014316410
99.6 BARC04813510500
101.3 BARC01921503395
101.6 BARC01921903397
103.3 BARC03151507105
104.4 BARC03081706946
cM C17.1
0.0 BARC01752503061
0.0 BARC05746714765
cM C17.2
0.0 BARC05229511407
0.6 BARC06469518779


cM C18
0.0 BARC02936906162
0.3 BARC05295711678
1.1 BARC04319708552
8.3 BARC04047907752
10.4 BARCG0147700243
10.4 BARC04827710538
11.0 BARC04827510534
11.7 BARC01229501800
32.0 BARC04693412799
54.2 BARC06358118909
56.3 BARC04298508484
59.6 BARC05275311606
61.0 BARC05158711167
64.8 BARC05109910991
65.6 BARC05711714589
65.8 BARC05046509647
67.2 BARC04757212990
67.2 BARC04707612834
67.2 BARC04757012985
67.4 BARC04724612893
72.1 BARC04687812780
74.6 BARC06370518440
75.8 BARC05756514836
76.0 BARC04750412947
76.1 BARC04751612955
76.3 BARC04750212946
78.7 BARC01478301660
88.3 BARC05557513456
92.7 BARC05663514538
93.2 BARC01925503845
98.3 BARC04016307672
98.9 BARC04042907735
103.7 BARC04876110703
104.6 BARC01686702359
104.6 BARC01844103188
105.3 BARC04239308251
106.6 BARC05204511324
108.2 BARC01350900507
108.9 BARC05901715576
109.7 BARC03227708935
116.6 Satt472
119.7 BARC03887307372
123.3 BARC03134307057
127.8 18_1821
130.8 BARC04436308678
163.9 BARC03012306813
164.9 BARC03939707314
165.2 BARC04998909280
cM C19
0.0 BARC05944315806
0.2 BARC06223117720
16.9 BARC01618102303
39.0 BARC04749612943
39.0 BARC02434504854
41.7 BARC02606905243
48.0 BARC01614502292
50.7 BARC06576919741
cM C20.1
0.0 BARC03886907364
15.3 BARC01358301166
28.9 BARC04789710420
31.4 BARC04289708454
35.3 BARC04268508348
40.0 BARC02930106148
47.4 BARC05848115299
49.5 BARC04222908217
52.5 BARC06473518787
54.7 BARC05064109789
54.7 BARC05860515330
56.4 BARC01071900713
56.5 BARC06385118472
58.0 BARC01555902004
58.6 BARC04804110476
cM C20.2
0.0 BARC04261908314
15.5 BARC06011516397
17.4 BARC02982706444
18.4 BARC03815709983
25.3 BARC04112907912
28.8 BARC02071304700