5-2022

Breeding Soybean [Glycine max (L) Merr.] Under Reduced Irrigation

Francia Seconde Ravelombola
University of Arkansas, Fayetteville

Breeding Soybean [*Glycine max* (L) Merr.] Under Reduced Irrigation

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Crop, Soil, and Environmental Sciences

by

Francia Seconde Ravelombola
University of Antananarivo, Madagascar
Master of Science in Agronomy, 2016
Szent Istvan University, Hungary
Master of Horticulture, 2018

May 2022
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Ehsan Shakiba, Ph.D.
Dissertation Director

_____
Leandro A. Mozzoni, Ph.D
Committee Member

_____
Larry C. Purcell, Ph.D.
Committee Member

_____
Fernando Grignola, Ph.D.
Committee Member

_____
Trenton L. Roberts, Ph.D.
Committee Member

_____
Christopher G. Henry, Ph.D.
Committee Member

Abstract

Soybean [*Glycine max* (L) Merr.], a legume species native to East Asia in the *Fabaceae* family, ranks among the most important food crops in the world. It is widely grown and known for its high protein and oil concentration. Soybean is valuable because its seeds have multiple applications in food, feed, pharmaceutical, and industrial enterprises. Even though seed yield is the most important trait, breeders have recently given a significant attention to quality traits, such as high protein or modified oil concentration. Soybean seed protein inheritance has been extensively studied; however, genetics of high-protein 'BARC-7' soybean are still unknown.

On the other hand, soybean production in the United States Mid-South relies heavily on irrigation with 85% of soybean surfaces are supplemented with water in Arkansas. The most common irrigation practice in the U.S. Mid-South is furrow irrigation. Furrow irrigation could add an extra dimension of variation because of water gradients on the front and back of field, and potentially unequal flows between rows. However, the National Centers for Environmental Information reported a water shortage in eastern Arkansas, causing a reduction in water levels at irrigation reservoirs and generating concern on water availability for crop irrigation during reproductive stages. A reduction in groundwater availability could result in farmers having to skip or delay irrigation at a certain reproductive stage. Reduction in irrigation due to water quantity restrictions will significantly affect soybean yield, making variety selection increasingly important. Also, exploring molecular approaches to increase yield genetic gain has been one of the main challenges for soybean breeders and geneticists.

Therefore, the objectives of this study were 1) to map of high-protein 'BARC-7' gene using F_2-derived lines 2) to assess if irrigation onsets at different reproductive stages affect wilting, seed yield and key agronomic traits on determinate maturity group 5 (MG 5) soybean 3)

to conduct a nested association mapping (NAM) for wilting, maturity, and seed yield and to identify superior individuals in seed yield using genomic approach under different irrigation onsets 4) to evaluate the spatial variability of furrow-irrigated soybean for seed yield, wilting, and maturity under four different irrigation onsets.

Results suggested that QTL for protein and oil inherited from 'BARC-7' were identified on chromosomes 6, 13, and 20. The known major QTL on chromosome 20 was not detected. Results also indicated significant differences in wilting and yield but no significant differences in maturity, protein, oil content, and 100-seed weight across different irrigation onsets. Results revealed that a total of 4, 39, and 7 SNPs were found to be significantly associated with canopy wilting, maturity, and seed yield, respectively, using the combined data under different irrigation onsets obtained over four environments (location-year combination). Overall genomic selection accuracy was moderate ranging from 0.39 to 0.44, and genomic selection was efficient to select superior soybean lines under reduced irrigation. The spatial models displayed better data fitting (lower AIC and/or BIC) than the block model in each different irrigation level across different environments and traits. Indeed, genotype ranking for seed yield was different between the block model and the best spatial model, suggesting that spatial adjustment may be necessary for soybean breeding operations under furrow irrigation. Further validation in a breeding yield trial demonstrated similar results of the effectiveness in terms of AIC and/or BIC of the spatial model compared to the block model for soybean seed yield. The results from this study could contribute to proceed to a fine-mapping to the regions associated with high protein and oil in 'BARC-7' genetic background, to better understand mild drought on populations in order to define the breeding objectives and subsequent deployment of soybean lines under limited irrigation, and to suggest a spatial adjustment for soybean breeding operations under furrow irrigation.

Acknowledgments

I am extremely grateful to Dr. Leandro Mozzoni and Dr. Ehsan Shakiba who are one of the best advisors I ever met, for accepting me to work under their guidance and directions. Without their full and endless support, I would never reach this step. I would like also to thank my committee members, Dr. Larry Purcell, Dr. Trent Roberts, Dr. Fernando Grignola, and Dr. Christopher Henry for their suggestions, and academic and research support. I would like also to thank Dr. Song in the USDA for helping us with the genotyping part of the project. Many thanks to all the graduate students: Maria de Oliveria, Derrick Harrison, and Josh Winter; all undergraduate students: John Grey, Julia Kanwisher, Visiting Scholar and Scientists: Vitor Bortolotti, Rodrigo Botelho, Diego Belapart, Jefferson Prass Pimentel and Elaine Batista; all the soybean Breeding Programs composed by Ali Ablao, Daniel Rogers, Dr. Liliana Florez, Dr. Andrea Acuna , Dr. Marcos da Silva, Dr. Wu; and all the technicians and staffs in different research stations: Jonathan McCoy, Jody Hedge, Scott Hayes and Matthew Ryan Young.

I would like also to thank my dad, my mom, my brother, and my husband for their full support, assistance, and unconditional love.

Dedication

This dissertation is dedicated to:

My dad: Second Modeste Velombola

My mom: Francine Razanamala

My brother: Waltram Second Ravelombola

My husband: Toky Ramananjatovo

Preface

This dissertation includes six chapters. The first chapter (Chapter I) is related to the introduction and literature review to introduce different topics. The second chapter (Chapter II) is discussing on mapping a high protein gene “BARC7” gene inheritance with is a unrelated data to the chapter III, IV and V. Chapter III evaluates the impact of delaying irrigation on wilting, seed yield, and other agronomic traits of determinate MG5 soybeans. Chapter IV is presenting a nested association mapping for wilting, maturity, seed yield and seed yield genomic selection under reduced irrigation in two RILs soybean populations. Then, Chapter V is assessing the spatial models for seed yield, wilting, and maturity in furrow-irrigated soybean plots. Finally, Chapter VI is giving an overall conclusion and some recommendations.

Table of Contents

CHAPTER I: Introduction and Literature Review 1

Production and Importance of Soybean 2

Soybean Seed Composition 2

Soybean Production in Arkansas 3

Water management in soybean production 5

Drought effects on vegetative stages 6

Drought effects on yield and yield components 6

Drought effects on protein and oil content 7

Mechanisms of drought tolerance 8

Physiological traits associated with drought trait tolerance 9

Water use efficiency 9

Transpiration 10

Canopy coverage 10

Canopy temperature 10

Canopy wilting 11

Biochemical mechanisms of drought tolerance 11

Enzyme activity 11

Abscisic acid accumulation 12

Traditional and Molecular Breeding of Soybean 12

Quantitative Trait Loci analysis 13

Genome wide association study 13

Genomic selection 14

Breeding for modified protein content 16

Breeding drought-tolerant and slow wilting lines 16

References 18

CHAPTER II: Genetics of seed protein and oil inherited from "BARC-7" soybean in two F2-derived mapping populations 34

ABSTRACT 35

INTRODUCTION 36

MATERIALS AND METHODS 37

Plant materials and phenotyping 37

Genotyping and QTL mapping 39

RESULTS 40

Phenotypic Variation for Seed Protein and Oil Content 40

Trait and Environment Correlation, and Heritability 41

QTL Analysis of Seed Protein 42

QTL analysis of Seed Oil 43

DISCUSSION 44

CONCLUSIONS 48

REFERENCES 49

CHAPTER III: Impact of Delaying Irrigation on Wilting, Seed Yield, and Other Agronomic Traits of Determinate MG5 Soybean 67

ABSTRACT 68

INTRODUCTION 69

MATERIALS AND METHODS 71

Plant Materials and Experimental Design 71

Weather Conditions and Soil Sampling 73

Soybean Phenotyping 76

Data Analysis 76

RESULTS 77

Weather Conditions and Soil Sampling ..... **Error! Bookmark not defined.**

Canopy Wilting 80

Maturity 81

Seed yield 81

Seed protein and oil content, and 100-seed weight 81

DISCUSSION 82

CONCLUSIONS 87

REFERENCES 88

CHAPTER IV: Nested association mapping for wilting, maturity, seed yield and seed yield genomic selection under reduced irrigation in two RILs soybean populations 120

ABSTRACT 121

INTRODUCTION 122

MATERIALS AND METHODS 125

Plant materials and experimental design 125

Phenotyping 126

Genotyping and quality control 126

Best linear unbiased prediction (BLUP) 127

Nested association Mapping and candidate gene discovery 127

Candidate gene(s) discovery 128

Genomic estimated breeding values (GEBVS) 128

Lines advancement and selection based on Yield GEBVs 129

RESULTS 130

Nested association mapping (NAM) 130

Candidate genes 130

Genomic Selection 131

Lines advancement and selection 132

DISCUSSION 133

CONCLUSIONS 135

REFERENCES 136

CHAPTER V: Spatial models for seed yield, wilting, and maturity in furrow-irrigated soybean plots 145

ABSTRACT 146

INTRODUCTION 147

MATERIALS AND METHODS 149

Plant materials and Experimental design 149

Phenotyping 150

Data Analysis 150

Validation of spatial adjustment in breeding trial datasets 152

RESULTS 153

Comparison of the different models for wilting, maturity, and yield 153

Validation of spatial adjustment in breeding trial datasets 156

DISCUSSION 157

CONCLUSION 160

REFERENCES 161

CHAPTER VI: Overall Conclusions 199

List of Tables

CHAPTER I

Table 1.1. Soybean water use (ET) by growth stage (ASGROW, 2016)..... 27

Table 1.2. Water requirements for soybean during late reproductive stages through maturity (Table modified from Yonts, C.D. et al., 2008. Predicting the last irrigation of the season. NebGuide G1871. University of Nebraska-Lincoln Extension.)..... 27

Table 1.3. Allowable Deficits-Soybean as reported by Henry et al. (2014) 27

Table 1.4. Atmometer Setting as reported by Henry et al. (2014). 28

Table 1.5. Major seed protein QTL identified using bi-parental mapping population 29

Table 1.6. Major seed protein QTL identified using diverse germplasm (GWAS)..... 31

Table 1.7. Reported QTL related to soybean drought tolerance in bi-parental populations. 31

CHAPTER II

Table 2.1. Descriptive statistics of seed protein content (% dry matter) in 250 F_2-derived breeding lines each of two different breeding populations when evaluated in four locations and assessed with near-infrared spectroscopy. 53

Table 2.2. Descriptive statistics of seed oil content (% dry matter) in 250 F2-derived breeding lines each of two different breeding populations when evaluated in four locations and assessed with near-infrared spectroscopy. 54

Table 2.3. QTL analysis of seed protein content for two breeding populations each consisting of 250 F2-derived breeding lines. Analysis conducted using composite interval mapping of phenotypic data consisting of least-square means from 4-environment 1-rep trials. Confidence intervals in centimorgan (cM) within each population. Previously-reported QTL in the reported region are presented, as available on Soybase.org as of October 10, 2021. 55

Table 2.4. QTL analysis of seed oil content for two breeding populations each consisting of 250 F2-derived breeding lines. Analysis conducted using composite interval mapping of phenotypic data consisting of least-square means from 4-environment 1-rep trials. Confidence intervals in centimorgan (cM) within each population. Previously-reported QTL in the reported region are presented, as available on Soybase.org as of October 10, 2021. 56

Table S 2.1. ANOVA table for seed protein content for 250 F2-derived breeding lines originated from the cross of UA 5115C/UA 5814HP when grown on non-replicated RCBD planted in four environments. 60

Table S 2.2. ANOVA table for seed protein content for 250 F2-derived breeding lines originated from the cross of R13-532/UA 5814HP when grown on non-replicated RCBD planted in four environments. 60

Table S 2.3. QTL analysis for seed protein content in 250 F2-derived progenies of UA 5115C/UA 5814HP 61

Table S 2.4. QTL analysis for seed protein content in 250 F2-derived progenies R13-532/UA 5814HP. 62

Table S 2.5. QTL analysis for oil protein content in 250 F2-derived progenies of UA 5115C/UA 5814HP. 63

Table S 2.6. QTL analysis for oil protein content in 250 F2-derived progenies of R13-532/UA 5814HP 64

CHAPTER III

Table 3.1. Soybean seed protein, oil, and 100-seed weight of each wilting class under different onset irrigations (R1, R2, R3, and R4) evaluated in four environments (location-year combination 19ROH, 19STU, 20ROH, and 20STU) 95

Table S 3.1. Average minimum (Min_T), maximum (Max_T), average (Avg_T) temperature in °C monthly from planting to harvest evaluated in four environments (location-year combination 19ROH, 19STU, 20ROH, and 20STU) 102

Table S 3.2. Soil type in each environment based on the percentage of sand, silt and clay 102

Table S 3.3. Characteristic of elements of soils from soil samplings for pre and post samplings for different irrigation onsets (R1, R2, R3, and R4) in four environments (19ROH, 19STU, 20ROH, and 20STU), paired t- test, mean difference, t ratio and p-value 103

CHAPTER IV

Table 4.1. Significant SNPs associated with canopy wilting, maturity, seed yield with their respective LOD (-log10(p_value)) value, annotated gene found within a 40-kb genomic region flanking the significant SNP, and functional annotation corresponding to the candidate gene. 140

Table 4.2. Yield of different selected soybean lines based on GEBVs under irrigated and non-irrigated conditions in kg/ha grown in on location with three replication as a randomized complete block design during the summer of 2021. LSD least Significant Difference at level 0.05 143

CHAPTER V

Table 5.1. Comparison of three different models (block model, exponential model, and gaussian model) in four environments (19ROH, 19STU, 20ROH, and 20STU) in terms of AIC (Akaike's Information Criteria), BIC (Bayesian Information Criteria), and significant of genotype effects, irrigation effects, and the interaction of the irrigation and genotype effect on soybean wilting. 165

Table 5.2. Comparison of three different models (block model, exponential model, and gaussian model) in four environments (19ROH, 19STU, 20ROH, and 20STU) in terms of AIC (Akaike's Information Criteria), BIC (Bayesian Information Criteria), and significant of genotype effects, irrigation effects, and the interaction of the irrigation and genotype effect on soybean maturity. 166

Table 5.3. Comparison of three different models (block model, exponential model, and gaussian model) in four different environments (19ROH, 19STU, 20ROH, and 20STU) in terms of AIC (Akaike's Information Criteria), BIC (Bayesian Information Criteria), and significant of genotype effects, irrigation effects and the interaction of the irrigation and genotype effect on soybean seed yield..... 167

Table 5.4. Comparison of three different models (block model, exponential model, and gaussian model) for the 2019 and 2020 Final Yield Trials (19AF4E and 20AF4E) grown in four environments in terms of AIC (Akaike's Information Criteria), BIC (Bayesian Information Criteria), and significant of genotype effects on soybean seed yield. 168

Table 5.5. Heritability for seed yield for block model and exponential model for each environment for the 2019 Arkansas Final yield trials maturity Group IV Early (19AF4E) and 2020 Arkansas Final yield trials maturity group IV Early (20AF4E) 168

Table S 5.1. Yield BLUPs for each irrigation level and ranking of soybean genotype for the block model and the exponential model grown at four Arkansas environments. 177

Table S 5.2. Soybean Yield BLUP for the 2019 Final Yield Trials Maturity group IV (19 AF4E) grown in four Arkansas environment under a randomized-complete block design (RCBD) with two replications and ranking of the genotypes after analysis using an exponential model and RCBD models. 197

Table S 5.3. Soybean Yield BLUP for the 2020 Final Yield Trials Maturity group IV (20AF4E) grown in two or four Arkansas environments as a randomized-complete block design (RCBD)

with two replications, and ranking of the genotypes after analysis using an exponential model and a block model. 198

List of Figures

CHAPTER I

Figure 1.1. Soybean production in the United States (USA) over 40 years (1978-2018) by USDA-NASS 32

Figure 1.2. Amino acid composition of soybean protein. Red essential amino acid, black nonessential (Asif & Acharya, 2013)..... 32

Figure 1.3. Deviation from average precipitation May to July 2018. Source NOAA, 2018. 33

Figure 1.4. Trend in soybean yield for 40 years (1977-2017) in Arkansas by USDA-NASS..... 33

Figure 2.1. Pearson correlation analysis of oil content (%) and protein content (%) for 250 F2-derived lines derived from UA 5115C/UA 5814HP (a) and R13-532/UA 5814HP (b) each. The light blue color circle represents the 95% confidence interval for the correlation; the thinner the circle, the higher the correlation. 57

CHAPTER II

Figure 2.2. Pearson correlation analysis of seed protein among 250 F2-derived lines from UA 5115C/UA 5814HP (a) and R13-532/UA 5814HP (b). The shaded area represents the 95% confidence interval for the correlation. 58

Figure 2.3. Pearson correlation analysis of seed oil among 250 F2-derived lines from UA 5115C/UA 5814HP (a) and R13-532/UA 5814HP (b) each. The shaded area represents the 95% confidence interval for the correlation. 59

CHAPTER III

Figure 3.1: Daily distribution of rainfall and irrigation (mm) during the two growing seasons (2019 and 2020) in Rohwer (ROH). DOY: day of the year. The ▽ symbol indicates the time of canopy wilting rating. 97

Figure 3.2: Daily distribution of rainfall and irrigation (mm) during the two growing seasons (2019 and 2020) in Stuttgart (STU). DOY: day of the year. The ▽ symbol indicates the time of canopy wilting rating. 98

Figure 3.3. Canopy witling of each wilting class under different onset irrigations (R1, R2, R3, and R4). Same letters are not significantly different (Tukey's; $p < 0.05$). Whiskers denote standard error of the mean. 99

Figure 3.4. Maturity of each wilting class under different onset irrigations (R1, R2, R3, and R4). Same letters are not significantly different (Tukey's; $p < 0.05$). Whiskers denote standard error of the mean. 100

Figure 3.5. Soybean seed yield of each wilting class under different onset irrigations (R1, R2, R3, and R4). Same letters are not significantly different (Tukey's; $p < 0.05$). Whiskers denote standard error of the mean. 101

CHAPTER IV

Figure 4.1. Spearman's correlation among the GEBVs in each irrigation treatment 144

Figure 4.2. Principal component analysis among the GEBVs for R1, R2, R3, R4 stages..... 144

CHAPTER V

Figure 5.1. Aerial image of the layout of the experimental design in the 20STU environment. Each strip represents a different irrigation level (R1, R2, R3, R4 growth stages). Each irrigation level is separated with four rows of borders. 169

Figure 5.2. QQ plots for block model, exponential model, and gaussian model in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean wilting..... 170

Figure 5.3. Student residuals for block model, exponential model, and gaussian model in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean wilting..... 171

Figure 5.4. QQ plots for block model, exponential model, and gaussian model in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean maturity..... 172

Figure 5.5. Student residuals for linear mixed model with blocking structure (a), exponential model (b), and gaussian model (c) in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean maturity. 173

Figure 5.6. QQ plots linear mixed model with blocking structure (a), exponential model (b), and gaussian model (c) in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean seed yield. 174

Figure 5.7. Student residuals linear mixed model with blocking structure (a), exponential model (b), and gaussian model (c) in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean seed yield. 175

Figure 5.8. Kendall correlation for BLUP yield in each irrigation level at R1, R2, R3, and R4 growth stages between the block model and the exponential model, and p-value of the significance of the Kendall correlation 176

Figure 5.9: Exemplary of a spatial trend in a range/column manner for unadjusted (raw) seed yield and seed yield of three models including: block, exponential, and gaussian models for the irrigation R1 in 20STU environment 177

List of Publications

CHAPTER II: Ravelombola, F., Chen, P., Vuong, T., Nguyen, H., Mian, R., Acuña, A., ... & Mozzoni, L. 2022. Genetics of seed protein and oil inherited from "BARC-7" soybean in two F2-derived mapping populations. *Journal of Crop Improvement*, 1-17.

CHAPTER III: Ravelombola, F.; Acuña, A.; Florez-Palacios, L.; Wu, C.; Harrison, D.; deOliveira, M.; Winter, J.; DaSilva, M.; Roberts, T.; Henry, C.; Grignola, F.; Shakiba, E.; Mozzoni, L. 2022. Impact of Delaying Irrigation on Wilting, Seed Yield, and Other Agronomic Traits of Determinate MG5 Soybean.*Agronomy*,*12*, 1115. https://doi.org/10.3390/agronomy12051115

CHAPTER V: Ravelombola, F., Acuña, A., Florez-Palacios, L., Wu, C., Harrison, D., de Oliveira, M., Winter, J., Da Silva, M.P., and Mozzoni, L. Spatial models for seed yield, wilting, and maturity in furrow-irrigated soybean plots. *Journal of Crop Improvement.* (Accepted)

CHAPTER I

Introduction and Literature Review

Production and importance of soybean

The major oilseed crops in the United States (U.S.) are soybean [*Glycine max* (L) Merr.], cottonseed, sunflower seed, canola, rapeseed, and peanut, with soybean accounting for 90% of the total production (https://www.fas.usda.gov/). Second to *Zea mays* (corn), soybean are the most cultivated crop in the U.S. (https://www.nass.usda.gov/), and the U.S. is the largest soybean producer in the world, followed by Brazil and Argentina (https://www.nass.usda.gov/). Global soybean production has risen dramatically over the past 40 years in the U.S. **(Figure 1.1).** Key export markets for soybeans include Europe, Japan, Mexico, Taiwan, China, and the Republic of Korea. Soybean is one of the most economically important crops in Arkansas, providing a substantial source of revenue to growers, as Arkansas ranks among the top ten producers in the U.S. (https://www.nass.usda.gov/). In fact, a total of 4.84 million tons of soybean are produced (NASS USDA, 2021) by more than 6,800 Arkansans farmers in 1.4 Mha making it the largest crop by acreage in the state (AFBF, 2021). The farmers' production accounted for $747,098,000 cash receipts in 2001, which represented 13.4% of the total cash receipts from marketing of all commodities in Arkansas and 37% of the total cash receipts for crops (Coats & Ashlock, 2001),. The demand of soybean has dramatically increased in the past couple of decades (Wilson, 2016). Indeed, it is a primary source protein and oil. Soybean is called the "green gold" and the "miracle crop" (Liu, 1997) because of its high value in active compounds.

Soybean seed composition

Dry Soybeans seed typically contain 38% protein and 18% oil by weight **(Figure 1.2)**. The remainder consists of 30% carbohydrates, 9% of water, and ash **(Figure 1.2)**. The majority of carbohydrates within soybean seeds is either sucrose (41%-68%), stachyose (12-35%) or raffinose (5-16%) (Verma & Shoemaker, 1996). The average fatty acid composition of

commercial soybean oil is about 10% palmitic acid (16:0), 4% stearic acid (18:0), 22% oleic acid (18:1), 54% linoleic acid (18:2), and 10% linolenic acid (18:3). Globulins and albumins are the two major components of seed storage protein **(Figure 1.2)**. Soybean seed proteins contain 70% of globulin that is predominately composed of glycinin and β-conglycinin (Warrington et al., 2015). β-conglycinin is important for soybean improvement, for it is an essential amino acid that monogastric animals cannot synthesize and must be incorporated into their diet (Ma et al., 2016; Panthee et al., 2005).

Protein and oil have a negative correlation with one another (Diers et al., 1992; Sebolt et al., 2000; Nichols et al., 2006; Pathan et al., 2013), where increasing the seed content by weight of one will cause a linear decrease of the other. Genome-wide association studies (GWAS) (Hwang et al., 2014; Vaughn et al., 2014; Bandillo et al., 2015) and Quantitatitve trait loci (QTL) analysis (Nichols et al., 2006; Pathan et al., 2013) have shown QTL, haplotype or genomic loci in chromosome 20, 15, and 5 associated to both oil and protein, indicating a negative pleiotropic effect. In addition to pleiotropic effects of protein on oil, variation in seed protein concentration significantly affects seed size, crop growth, and development (Poeta et al., 2016). High-protein genotypes showed lower leaf area and harvest index when compared with high-yielding genotypes (Poeta et al., 2016). A high-protein large seed was associated with more assimilate availability per seed during seed filling, while high-protein small seed showed higher leaf area at the beginning of seed fill, more canopy biomass production, and low levels of assimilate per seed (Poeta et al., 2016).

Soybean production in Arkansas

Soybean is grown in 45 of the 75 counties in Arkansas, which are concentrated in the eastern half of the state comprising the Mississippi delta region. It is also grown in counties that

lie in the Arkansas River valley and in the southwestern corner of the state (Spradley, 2005). It is successfully grown in a wide range of soil types in Arkansas, including sandy loam, silt loam, and heavier textured, predominantly clay soils. Approximately 85% of Arkansas soybean are irrigated (AFBF, 2021). Commercial production of soybean in Arkansas uses maturity groups (MG) 3 through 5.

A soybean crop will produce approximately 0.13 t/ha for every 25.4 mm of water used through the season (Riley, 2014). Soybean yields ranged from 2.69 to 3.36 t/ha require 508 to 635 mm of available soil moisture during the growing season. A total of 254 to 381 mm of irrigation water is required to assure a standard soybean yield (Riley, 2014); however, the irrigation needed will vary on the soil moisture and rainfall. There are several methods of soybean irrigation used in Arkansas, including furrow irrigation, flood irrigation, and pivot systems. Furrow irrigation is a practice in which soybean are planted on raised beds, allowing a furrow between each row. Water is then directed down the furrows when irrigation is needed. This is usually done using poly pipe (Tacker & Vories, 1998), which is rolled out across the top of the soybean field and connected to an irrigation riser (Enciso & Peries, 2005). Another irrigation method is flood irrigation, in which soybean are planted on flat land and simply runs poly pipe along the top of the field and allows water to flow across the field (Tacker & Vories, 1998). The problem with this method is getting the water to spread evenly over the field. Difficulties with flood irrigation led to the development of the border irrigation method in which a large volume of water over a flat surface in a short period is flushed. A fourth irrigation method is pivot sprinkler irrigation in which sprinklers mounted on rollers are moved across a field to irrigate. The rule-of-thumb is that irrigation for soybean is necessary whenever the available soil water falls to 50% (Tacker & Vories, 1998).

The timing of irrigation is referred to as irrigation scheduling. Correct timing is critical to maximizing yield. Too often growers irrigate by the appearance of the crop. Visual stress, especially during bloom and pod set, results in yield loss. Also, once irrigation is started, the time required to finish a field will result in part of the crop suffering even greater stress. If the soil moisture can be determined, then irrigation timing decisions can be improved. Determining the soil moisture by visual observation or by kicking the soil surface is difficult and can be misleading (Tacker & Vories, 1998). One available method to apply irrigation in a timely manner is the use of atmometer (Nygren & Ingram, 2016). An atmometer or evaporimeter or Evapotranspiration Gauge is a scientific instrument used for measuring the rate of water evaporation from a wet surface to the atmosphere (Nygren & Ingram, 2016). The setting of the atmometer is based on the type of the soil, the type of irrigation, and the growth stages of the soybeans **(Table 1.2 & Table 1.3)**.

Water management in soybean production

According to a study reported by the Natural Resources Defense Council (www.nrdc.org), Arkansas is at risk of water shortage by 2050. Both the 1990 and 2014 Arkansas Water Plans identified major gaps in groundwater availability when forecasting 40-year irrigation needs for eastern Arkansas' agriculture (Arkansas Natural Resources Commission [ANRC], 2014). These shortages, as much as 80% of demand, could result in future groundwater restrictions for agricultural use. The need for increased water withdrawal in the State, which refers to the total volume or removed from a water source such as lake or river and returned to the source to be used again, could result in significant costs for the agricultural community (Winthrop Rockefeller Foundation, 2008). Groundwater levels dropped from 1980 to 2005 (Winthrop Rockefeller Foundation, 2008). The drought of 1980–1981 raised concerns about

water shortage in Arkansas, including the impact on crop irrigation (Looney, 1984). By 1981, groundwater levels in the agriculture dependent Delta counties had dropped from 6 to 9 m below the surface to over 12 to 15 m deep (AFBF, 1981). Depletion of the Alluvial Aquifer in the Delta had been occurring, and documented, since the 1920s (AFBF, 1981), but during the early 1980s, some farmers' wells were drying up because the groundwater level had been lowered below the depth to which many wells had been drilled (AFBF, 1981). Depletion and drop in groundwater levels were observed during the summer of 2018, as a large portion of the Arkansas Mississippi delta experienced as low as 25 to 50% of normal precipitation for the interval May to July **(Figure 1.3)**.

Drought effects on vegetative stages

Irrigation is generally not required during germination, unless moisture conditions at planting were inadequate. Soybean uses little amounts of water during the seedling stage, and the demand increases while during vegetative growths **(Table 1.1 & Table 1.2)**. However, too much water in the early season can prolong the vegetative growth stage, which can result in delays in flowering, increased plant height, and lodging. Limiting water supply in the early season might encourage plants to develop stronger, healthier root systems that grow deeper. Farmers rely on soil moisture and natural precipitation as much as possible during the early growth stages (McWilliams et al., 1999).

Drought effects on yield and yield components

Unpredictable and inadequate rainfall and temperatures reduce soybean yield up to 40% (Specht et al., 1999), and 40-year average yields of non-irrigated compared to irrigated AR soybean crop show consistent underperformance of non-irrigated soybean (USDA-NASS) **(Figure 1.4)**.Also, soybean performance tests include several varieties that consistently produce

in the 3.36 t/ha range when irrigated, but the same varieties average 0.67 to 1.34 t/ha less without irrigation.

Water requirements for soybean range from 381 to over 635 mm depending on planting date, maturity group, location, and weather (Frederick et al., 2001). Irrigation is required during flowering on soils with an insufficient water holding capacity, such as sandy soils or when conditions are exceptionally dry. Soybean are most sensitive to water stress during the mid- to late-reproductive stages: pod development (R3 to R4) and seed fill (R5 to R6) (Fehr et al., 1971; Korte et al., 1983; Specht et al., 1999). Water stress during pod development and early seed fill can have the greatest impact on yield, and result in a reduced number of seeds per pod and smaller seed size (Desclaux et al., 2000) **(Table 1.1 &Table 1.2)**. Discontinuing irrigation before physiological maturity can considerably reduce yield if the soil water content is not adequate (Desclaux et al., 2000).

Drought effects on protein and oil concentration

Drought might affect protein synthesis as well as the protein structure (Frota & Tucker, 1978; Termaat & Munns, 1986). The incorporation of amino acids into the leaf proteins might be inhibited because of the removal of the hydration shell of the protein that is dissociated. Dornbos & Mullen, (1992) investigated soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. Chemical composition of the seed was altered and reduction in yield, viability and vigor were found during seed filling under drought. Protein content was increased by 4.4% while oil content decreased by 2.9% under drought stress (Dornbos & Mullen, 1992). The increase of protein content is positively and linearly correlated to the stress level, whereas oil content is negatively and linearly correlated to the stress degree (Dornbos & Mullen, 1992). An unchanged composition of fatty acid of the oil was denoted;

however, the higher the air temperature, the lower the proportion of the polyunsaturated components (Dornbos & Mullen, 1992).

Mechanisms of drought tolerance

To cope with drought, soybean plants utilize different regulation and metabolic pathways. There are three different mechanisms to handle drought: drought escape, drought avoidance, and drought tolerance (Carrow, 1996; Levitt, 1980), making drought a very complex trait to discern.

Drought escape refers to a soybean completing its life cycle before the beginning of the drought, by shortening the life cycle to match water supply. One example of drought escape is the early soybean planting system used in the southern part of the U.S. Those short season soybeans are planted in March or early April in zones where later maturing cultivars have traditionally been grown. These early maturing cultivars begin blooming in late April to early May; starting setting seed in late May to early June and reach full seed setting by mid-July to early August. In the southern part of the U.S., the rainfall is usually plentiful from April to early July allowing the soybean crop to reach the critical reproductive stage with ample water prior to July and August where conditions often favor drought stress (Heatherly & Elmore, 2004). The mechanism involves physiological and developmental characteristics that usually result in earlier than anticipated seed production through shortening of the life cycle (Carrow, 1996).

Drought avoidance is characterized by mechanisms that maintain high water potentials in plant tissues under mild or moderate water deficit conditions. It can be accomplished by adopting different strategies, such as increasing rooting depth, promoting an efficient root system, reducing stomatal conductance, leaf rolling or folding, reducing the evaporation surface,

increasing wax accumulation on the leaf surface, and enhancing water storage abilities in specific organs (Carrow, 1996; Fang & Xiong, 2015; Ludlow & Muchow, 1990; O'toole % Bland, 1987).

Drought tolerance allows plants to maintain turgor and cell volume at low leaf water potential to continue metabolic activity longer under water stress through osmotic adjustment, antioxidant capacity, and cell membrane stability. Species capable of osmotic adjustment can maintain turgor at lower leaf water potentials that in non-adjusted species. Thus, osmotic adjustment clearly promotes drought tolerance, because such plants show improved growth when supplied with limited water than do non-adjusted species. Turgor maintenance under drought may sustain cell expansion and stomatal conductance. Soybean is not very efficient in maintaining turgor pressure (Hanson and Hitz, 1982). Soybean lack significant amounts of osmotica (K^+, NO_3^-, Na^+, Cl^-); the reason why their adaption to drought stress is poor (Nuccio et al., 1999).

Physiological traits associated with drought trait tolerance

Tuberosa (2012) established some physiological traits to screen for drought tolerance in soybeans such as water use efficiency (WUE), transpiration, canopy coverage (CC), canopy temperature (CT) and canopy witling (CW).

Water use efficiency

Water use efficiency is defined as the total dry matter produced (amount of carbon dioxide (CO_2) used fixed via photosynthesis) by plants per unit of water used (total quantity of water applied) (Taiz & Zeiger, 2006). It is the reciprocal of the transpiration ratio. It refers to how much dry matter can the plant take up per volume of water used.

$$WUE = \frac{D}{W}$$

where D is the mass of dry matter, W is the mass of water used (de Wit, 1958).

Transpiration

Transpiration represents the water flow through the plant and released into the atmosphere, while evaporation characterize the water directly released from the soil surface to the atmosphere. The transpiration coefficient is defined as the amount of water (weight or volume) required to produce a weight unit of plant dry matter (Mengel, 2001).

Canopy coverage

Canopy coverage is the proportion of the soil covered by the leaves. A closed canopy could reduce soil evaporation by improving WUE. In fact, a canopy coverage establishment may be useful to minimize the water loss as it may store water content in the soil for later developmental stages (Purcell & Specht, 2004; Rebetzke et al., 2007; Slafer et al., 2005). Greater canopy coverage also improves solar radiation interception, and it is positively associated with crop growth and yield (Edwards & Purcell, 2005; Liebisch et al., 2015).

Canopy temperature

Canopy temperature (CT) as measured by thermal imaging is the difference in temperature between the canopy surface and the surrounding air. In the field, genotypes with a higher canopy temperature use more of the available water in the soil to avoid excessive dehydration (Tuberosa, 2012). The surface temperature of the canopy is related to the amount of transpiration resulting in evaporative cooling. CT could be measured remotely using an infrared thermometer, and has advantages compared to other methods for stress detection, such as stomatal structure conductance and water potential, because it integrates a larger area of plant measurement, is non-destructive, does not interfere with stomata, and is faster and not laborious.

Canopy wilting

Wilting mechanisms appear to be related to soil moisture conservation even before drought stress becomes severe (Fletcher et al., 2007; King et al., 2009). Canopy wilting is the first visible symptom of soil water deficit in soybean. Soybean genotypes differ in the time of onset and the severity of canopy wilting in response to drought (Sloane et al., 1990). A second possible mechanism for slow wilting is soil water conservation. When soil water is plentiful, some slower-wilting genotypes can relatively maintain lower transpiration rates compared to conventional cultivars, and thus do not deplete the soil-moisture reservoir as rapidly. As the drought builds, enough soil moisture is available for slow-wilting genotypes to prolong transpiration and leaf turgor for several days compared to fast-wilting genotypes. One promising soybean trait for improving drought tolerance is delayed-canopy wilting (King et al., 2009). Field observations of delayed wilting in soybean were first noted in the early 1980s in a program that screened several hundred soybean plant introductions for agronomic drought tolerance in North Carolina (Sloane et al., 1990). Under drought conditions in the field, a few soybean genotypes wilted slowly compared with most other genotypes of similar maturity (King et al., 2009).

Biochemical mechanisms of drought tolerance

Enzyme activity

Drought stress might also influence enzymatic reactions. Enzyme activities associated with photosynthesis and respiration as well as ATPase activity are increased by drought stress in order to maintain increased energy demand (Lawlor, 2002).

Abscisic acid accumulation

The relationship between drought stress and phytohormones are complex. Under that drought stress, there is a rapid accumulation of abscisic acid (ABA). Accumulation of ABA induces stomatal closure and thus inhibits transpiration pathway (Herppich & Peckmann, 1997). Nevertheless, moderate drought stress appears to have a little effect on stomatal closure (Hsiao, 1973; Niinemets et al., 1999). For instance, soybeans showed no reduction in gaseous exchange, indicative of substantial closure of stomata, until the water potential of the leaves and fallen to as low as -1.0MPa, called threshold value. This indicate that due to turgor loss at relatively low water potentials gaseous exchanged is inhibited and thus the diffusion of CO_2 from the atmosphere through the stomata into the leaf tissue (Kramer & Boyer, 1995).

Traditional and molecular breeding of soybean

The center of origin of soybean is in China, and it has spread out to more than 50 countries in the last five centuries (Wilcox, 2004). Soybean belongs to the *Fabaceae* family, commonly called legumes, and is a cross-compatible plant. Soybean is a paleo-polyploid with diploid behavior with 20 pairs of chromosomes (2n=40) (Kim et al., 2010). There are approximately 3,500 accessions of perennial *Glycine* species in germplasm collections throughout the world (Verma & Shoemaker, 1996).

Genetic gain in soybean has been accomplished through conventional breeding methods and molecular-based strategies. Each cycle of genetic improvement begins with the breeder making choices as to the parents to be used to create segregating populations. Those populations are then advanced toward homozygosity with or without selection to produce relatively-homozygous lines that are the subject to selection for key traits, including yield, maturity, and disease tolerance. A given cycle ends when its best lines are released as improved pure-line

cultivars. Since a cycle is initiated each year, genetic improvement is a continuous process (Boerma & Specht, 2004). Selection of parents is extremely important to create segregating population that can be either elite cultivar or exotic material.

Quantitative trait loci analysis

The application of QTL analysis allows the identification of chromosomal regions conditioning the phenotypic variation in quantitative traits, such as protein, drought tolerance or resistance, and identifies the desirable alleles at these QTL for use in marker-assisted selection (MAS). With the massive advancement in genomics knowledge about the physiological and functional aspects of traits and metabolic pathways controlling trait expression, the candidate gene approach has become a powerful technique to associate traits to functional genes. This approach can increase the precision of the genetic mapping and increase the accuracy of detecting QTL related to the trait of interest (Zhu & Zhao, 2007).

Genome wide association study

Association mapping or Genome wide association study (GWAS) analyses the link of a molecular marker with a phenotypic trait of interest in unrelated individuals of a population, rather than using a mapping population of known pedigree. GWAS does not require any crossing and is suitable for fine scale mapping with a greater possibility for recombination to take place than traditional pedigree studies (Nordborg & Tavaré, 2002). Linkage analysis gives a more precise location of the QTL that controls the trait of interest (Glazier et al., 2002; Gupta et al., 2005).

Genomic selection

Genomic selection is a form of marker-assisted selection in which genetic markers consider the whole genome to predict phenotype based on the genotype (Goddard & Hayes, 2007). It can enhance genetic gain by speeding up the breeding cycles, as it has the potential for making selection of individuals before phenotyping. Instead of utilizing limited number of significant linked molecular markers in conventional marker-assisted selection, genomic selection estimates genome-wide molecular marker effects simultaneously, generates the genomic estimated breeding values (GEBVs) for lines, and selects the superior lines based on their GEBVs (Bernardo & Yu, 2007; Piyasatian et al., 2007). Genomic selection utilizes a training population, with individuals that are genotyped and phenotyped, and a testing population of individuals that are genotyped but not phenotyped. Genome-wide markers are random effects and all marker effects on the phenotype are estimated simultaneously in a single model. Prediction models attempt to capture the total additive genetic variance to estimate breeding value of individuals based on sum of all marker effects (Goddard & Hayes, 2007).

Genomic selection has been extensively studied in animal breeding (Hayes et al., 2009; Legarra et al., 2008; Tribout et al., 2012), and is becoming a powerful tool in plant breeding (Heffner et al., 2009; Jannink et al., 2010). Genomic prediction has been carried out in maize (Huang et al., 2016), soybean (Xavier et al., 2016; Zhang et al., 2016), rice (Onogi et al., 2016), canola (Jan et al., 2016), and wheat (Battenfield et al., 2016). Indeed, genomic selection has been frequently and widely applied in wheat breeding. The prediction accuracy of genomic selection in wheat breeding has been investigated using cross-validation methodology across multiple environments (Dawson et al., 2013) and multiple breeding cycles (Michel et al., 2016). Genomic selection has been assessed for quantitative traits (Poland et al., 2012), quality traits (Heffner et

al., 2011), and disease resistance traits (Rutkoski et al., 2012; Rutkoski et al., 2011; Rutkoski et al., 2014) in wheat breeding. In contrast, soybean breeding programs have rarely addressed the application of genomic selection. Genomic selection has been evaluated in soybean breeding for agronomic traits such as soybean cyst nematode (SCN) resistance (Bao et al., 2014), seed weight (Zhang et al., 2016), and yield components (Xavier et al., 2016). It is reported that genomic selection was more accurate than conventional MAS against SCN (Bao et al., 2014).

The size of the training population, the choice of prediction model, and the marker density influence the accuracy of genomic selection. Zhong et al. (2009) reported that the accuracy of the genomic prediction increased when the training population size increased. However, the gain on genomic prediction accuracies was more relevant at small training population size and tends to reach a plateau as training population size increased (Isidro et al., 2015). No minimum size of the training of population was reported. Wong & Bernardo (2008) demonstrated on the studies on the selection of oil palm that a training population size of 30, 50, and 70 has been successful. Genomic selection models employed in soybean breeding included ridge regression best linear unbiased prediction (rrBLUP) (Bernardo & Yu, 2007), ridge regression best linear unbiased prediction with major genes fitted as fixed effects (Bao et al., 2014), BayesA, BayesB, BayesC, Bayesian LASSO regression (BLR) (De Los Campos et al., 2009), reproducing kernel hilbert space , standard genomic best linear unbiased predictor (GBLUP) (Gao et al., 2012) which only includes additive effects, extended version of GBLUP which includes both additive effects and additive-by-additive effects (Cockerham, 1954; Xu et al., 2014), Bayesian Cp (BCP) (Habier et al., 2011), support vector machine (SVM) (Long et al., 2011), and random forest (RF) (González-Recio & Forni, 2011). In general, additive linear

models such as RR and RRF models outperformed sophisticated models such as Bayesian and machine learning in prediction accuracy (Bao et al., 2014).

Breeding for modified protein concentration

Major QTLs for protein and oil were consistently mapped in chromosome 5, 6, 9, 12, 14, 15, 17, 18, 20 and confirmed using bi-parental mapping population: derived lines, back cross, and RILs **(Table 1.5)**. Additionally, using a GWAS mapping approach, QTLs associated with seed protein and oil content were identified on chromosomes, 3, 5, 6, 8, 11, 12, 13, 15, 17, 20 **(Table 1.6)**. However, due to the lack of large effects, the negative relationship with oil and yield, and the inconsistency across environments, very few protein QTL were further used or incorporated in breeding programs (Wang et al., 2015).

Breeding drought-tolerant and slow wilting lines

Although Soybase, a USDA Soybean Genetics Database, reports a large number of QTL for agronomic, physiological, seed composition traits, biotic and abiotic factors in soybean (Grant et al., 2010), only a handful of QTLs have been reported for drought in bi-parental populations **(Table 1.7)**. In most cases, small population sizes have been used for QTL detection. Conducting GWAS of canopy wilting in diverse soybeans (373 maturity group 4 genotypes), Kaler et al. (2017) reported 61 environment-specific significant SNP-canopy wilting associations, and 21 SNPs that associated with canopy wilting in more than one environment. There were 34 significant SNPs associated with canopy wilting when averaged across environment, and the significant SNPs identified were located within a gene or very close to genes that had a reported biological connection to transpiration or water transport (Kaler et al., 2017).

Breeding lines and plant introductions (PIs) have been developed that wilt more slowly than other existing varieties. Two PIs, PI416937 and PI471938, are slow wilting and present drought tolerant traits. These lines were among the best drought tolerant sources and have been used in most of the breeding programs for drought tolerance. PI416937 was the first slow wilting line identified and is the most studied among drought tolerant lines in the U.S. to identify mechanisms for tolerance to drought (Pathan et al., 2013). Several possible explanations demonstrated why this line wilts more slowly under drought than other soybean lines. PI416937 had more highly-branched roots in the upper soil profile than drought sensitive lines (Busscher et al., 2000; Hudak & Patterson, 1996), and possesses the genetic capability to continue root growth on compacted soils. The drought tolerant lines PI416937 and PI471938 have been utilized to develop slow wilting varieties that perform well relative to other varieties regardless of water regime. N98-9683, a North Carolina State University soybean, G00-3209, a University of Georgia soybean line (Paris & Shelton, 2001), and R10-2436 and R11-2933, University of Arkansas soybean lines, are examples of slow wilting soybean lines (Ross, 2016).

Breeding drought-tolerant and high-yielding soybeans is a solution to cope with production under non-irrigated conditions. However, there is very little knowledge of the differences in response of cultivars with slow wilting versus those that do not possess the trait when grown under irrigation schedules that skip certain reproductive stages because of water scarcity. Therefore, it is critical to know the crop responses and appropriate breeding methodologies for selecting materials adapted to these situations.

References

Arkansas Farm Bureau Federation (AFBF). (1981). Water: Its uses and the implications for Arkansas agriculture. Retrieved January 29, 2019, from http://www.wrfoundation.org/media/1387/waterissues_full_resources.pdf

Arkansas Farm Bureau Federation (AFBF). (2021). Soybeans. Retrieved January 29, 2021, from https://www.arfb.com/pages/arkansas-agriculture/commodity-corner/soybean/

Arkansas Natural Resources Commission [ANRC],. AS(2014). Water Management. Retrieved January 29, 2019, from https://www.anrc.arkansas.gov/divisions/water-resources-management/arkansas-water-plan/

Ashlock, L., Mayhew, W., Windham, T., Keisling, T., Klerk, R., Beaty, D., & Lorenz, G. (2000). Production Systems and Economics. Retrieved January 29, 2019, from https://www.uaex.edu/publications/pdf/mp197/chapter16.pdf

Asif, M., & Acharya, M. (2013). Phytochemicals and nutritional health benefits of soy plant. *International Journal of Nutrition, Pharmacology, Neurological Diseases*, *3*(1), 64. https://doi.org/10.4103/2231-0738.106998

Bandillo, N., Jarquin, D., Song, Q., Nelson, R., Cregan, P., Specht, J., & Lorenz, A. (2015). A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *The Plant Genome, 8*(3).

Bao, Y., Vuong, T., Meinhardt, C., Tiffin, P., Denny, R., Chen, S., … Young, N. D. (2014). Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. *The Plant Genome, 7*(3).

Battenfield, S. D., Guzmán, C., Gaynor, R. C., Singh, R. P., Peña, R. J., Dreisigacker, S., … Poland, J. A. (2016). Genomic selection for processing and end-use quality traits in the CIMMYT spring bread wheat breeding program. *The Plant Genome*, *9*(2).

Bernardo, R., & Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Science*, *47*(3), 1082–1090.

Boerma, H. R., & Specht, J. E. (2004). *Soybeans: improvement, production and uses.* American Society of Agronomy.

Brummer, E. C., Graef, G. L., Orf, J., & Wilcox, J. R. (1997). Mapping QTL for Seed Protein and Oil Content in Eight Soybean Populations. *Crop Science*, *37*(2), 370. https://doi.org/10.2135/cropsci1997.0011183X003700020011x

Busscher, W. J., Frederick, J. R., & Bauer, P. J. (2000). Timing effects of deep tillage on penetration resistance and wheat and soybean yield. *Soil Science Society of America Journal*, *64*(3), 999–1003.

Carrow, R. N. (1996). Drought avoidance characteristics of diverse tall fescue cultivars. *Crop Science*, *36*(2), 371–377.

Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., … Specht, J. E. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Science*, *43*(3), 1053–1067.

Coats, R., & Ashlock, L. (2001). The Arkansas soybean industry, p. 1-6. In Arkansas Soybean Handbook. Publication MP 197. Cooperative Extension Service, University of Arkansas. Retrieved January 29, 2019, from https://www.uaex.edu/publications/pdf/mp197/chapter1.pdf

Cockerham, C. C. (1954). An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics*, *39*(6), 859.

Dawson, J. C., Endelman, J. B., Heslot, N., Crossa, J., Poland, J., Dreisigacker, S., … Jannink, J.-L. (2013). The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crops Research*, *154*, 12–22.

De Los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., … Cotes, J. M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigrees. *Genetics*.

de Wit, C. T. (1958). *Transpiration and crop yields*. Unknown Publisher.

Desclaux, D., Huynh, T.-T., & Roumet, P. (2000). Identification of soybean plant characteristics that indicate the timing of drought stress. *Crop Science*, *40*(3), 716–722.

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*(5), 608–612.

Dornbos Jr, D. L., & Mullen, R. E. (1992). Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. *Journal of the American Oil Chemists' Society*, *69*(3), 228–231.

Du, W., Wang, M., Fu, S., & Yu, D. (2009). Mapping QTLs for seed yield and drought susceptibility index in soybean (Glycine max L.) across different environments. *Journal of Genetics and Genomics = Yi Chuan Xue Bao*, *36*(12), 721.

Edwards, J. T., & Purcell, L. C. (2005). Soybean yield and biomass responses to increasing plant population among diverse maturity groups. *Crop Science*, *45*(5), 1770–1777.

Enciso, J., & Peries, X. (2005). Using Flexible Pipe (poly-pipe) Surface Irrigation. Retrieved January 29, 2019, from http://cotton.tamu.edu/Irrigation/FlexiblePipeFactSheet.pdf

Fang, Y., & Xiong, L. (2015). General mechanisms of drought response and their application in drought resistance improvement in plants. *Cellular and Molecular Life Sciences*, *72*(4), 673–689.

Fasoula, V. A., Harris, D. K., & Boerma, H. R. (2004). Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. *Crop Science*, *44*(4), 1218–1225.

Fehr, W. R., Caviness, C. E., Burmood, D. T., & Pennington, J. S. (1971). Stage of development descriptions for soybeans, Glycine Max (L.) Merrill 1. *Crop Science*, *11*(6), 929–931.

Flanders, A. (2014). Crop Enterprise Budgets for Arkansas Field Crops Planted in 2014. University of Arkansas Division of Agriculture, Research and Extension. Retrieved January 29, 2019, from https://www.uaex.edu/farm-ranch/economics-marketing/farm-planning/budgets/crop-budgets.aspx

Fletcher, A. L., Sinclair, T. R., & Allen Jr, L. H. (2007). Transpiration responses to vapor pressure deficit in well watered 'slow-wilting'and commercial soybean. *Environmental and Experimental Botany*, *61*(2), 145–151.

Frederick, J. R., Camp, C. R., & Bauer, P. J. (2001). Drought-stress effects on branch and mainstem seed yield and yield components of determinate soybean. *Crop Science*, *41*(3), 759–763.

Frota, J. N. E., & Tucker, T. C. (1978). Absorption Rates of Ammonium and Nitrate by Red Kidney Beans under Salt and Water Stress 1. *Soil Science Society of America Journal*, *42*(5), 753–756.

Gao, H., Christensen, O. F., Madsen, P., Nielsen, U. S., Zhang, Y., Lund, M. S., & Su, G. (2012). Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genetics Selection Evolution*, *44*(1), 8.

Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science*, *298*(5602), 2345–2349.

Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, *124*(6), 323–330.

González-Recio, O., & Forni, S. (2011). Genome-wide prediction of discrete traits using Bayesian regressions and machine learning. *Genetics Selection Evolution*, *43*(1), 7.

Gupta, P. K., Rustgi, S., & Kulwal, P. L. (2005). Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Molecular Biology*, *57*(4), 461–485.

Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*, *12*(1), 186.

Hanson, A. D., & Hitz, W. D. (1982). Metabolic responses of mesophytes to plant water deficits. *Annual Review of Plant Physiology*, *33*(1), 163–203.

Hart, T. (2019). Soybean Breeding and Genetics Program - The Miracle Bean. Retrieved January 29, 2019, from https://www.themiraclebean.com/soybean-breeding-and-genetics-program/

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, *92*(2), 433–443.

Heatherly, L. G., & Elmore, R. W. (2004). Managing inputs for peak production. *Soybeans: Improvement, Production, and Uses. Madison (USA): Agronomy Monograph*, *16*, 451–536.

Heffner, E. L., Jannink, J.-L., Iwata, H., Souza, E., & Sorrells, M. E. (2011). Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. *Crop Science*, *51*(6), 2597–2606. https://doi.org/10.2135/cropsci2011.05.0253

Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Science*, *49*(1), 1–12.

Henry, C. G., Espinoza, L., Ismanov, M., Francis, P., & Ross, J. (2014). Scheduling Irrigation using an Atmometer (ET Gauge) for Arkansas Soybeans. Retrieved January 28, 2019, from https://www.uaex.edu/environment-nature/water/docs/irrig-ET-Gage-Sheet-Soybean.pdf

Herppich, W. B., & Peckmann, K. (1997). Responses of gas exchange, photosynthesis, nocturnal acid accumulation and water relations of Aptenia cordifolia to short-term drought and rewatering. *Journal of Plant Physiology*, *150*(4), 467–474.

Hsiao, T. C. (1973). Plant responses to water stress. *Annual Review of Plant Physiology*, *24*(1), 519–570.

Huang, M., Cabrera, A., Hoffstetter, A., Griffey, C., Van Sanford, D., Costa, J., … Sneller, C. (2016). Genomic selection for wheat traits and trait stability. *Theoretical and Applied Genetics*, *129*(9), 1697–1710.

Hudak, C. M., & Patterson, R. P. (1996). Root distribution and soil moisture depletion pattern of a drought-resistant soybean plant introduction. *Agronomy Journal*, *88*(3), 478–485.

Hwang, E.-Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., & Cregan, P. B. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics*, *15*(1), 1.

Isidro, J., Jannink, J.-L., Akdemir, D., Poland, J., Heslot, N., & Sorrells, M. E. (2015). Training set optimization under population structure in genomic selection. *Theoretical and Applied Genetics*, *128*(1), 145–158.

Jan, H. U., Abbadi, A., Lücke, S., Nichols, R. A., & Snowdon, R. J. (2016). Genomic prediction of testcross performance in canola (Brassica napus). *PLoS One*, *11*(1), e0147769.

Jannink, J.-L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics*, *9*(2), 166–177.

Juvik, G. A., Bernard, R. L., & Kauffman, H. E. (1985). Directory of germplasm collections: 1. II. Food legumes (soyabean).

Kaler, A. (2017). Genomic and Physiological Approaches to Improve Drought Tolerance in Soybean.

Kim, M. Y., Lee, S., Van, K., Kim, T.-H., Jeong, S.-C., Choi, I.-Y., … Ma, J. (2010). Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. *Proceedings of the National Academy of Sciences*, *107*(51), 22032–22037.

King, C. A., Purcell, L. C., & Brye, K. R. (2009). Differential wilting among soybean genotypes in response to water deficit. *Crop Science*, *49*(1), 290–298.

Korte, L. L., Williams, J. H., Specht, J. E., & Sorensen, R. C. (1983). Irrigation of Soybean Genotypes During Reproductive Ontogeny. I. Agronomic Responses 1. *Crop Science*, *23*(3), 521–527.

Kramer, P. J., & Boyer, J. S. (1995). *Water relations of plants and soils*. Academic press.

Lawlor, D. W. (2002). Limitation to Photosynthesis in Water-stressed Leaves: Stomata vs. Metabolism and the Role of ATP. *Annals of Botany*, *89*(7), 871–885.

Lee, S. H., Bailey, M. A., Mian, M. A. R., & Carter, T. E. (1996). RFLP loci associated with soybean seed protein and oil content across populations and locations. *THEORETICAL AND APPLIED GENETICS*, *93*(5–6), 649–657.

Legarra, A., Robert-Granié, C., Manfredi, E., & Elsen, J.-M. (2008). Performance of genomic selection in mice. *Genetics*.

Levitt, J. (1980). *Responses of Plants to Environmental Stress, Volume 1: Chilling, Freezing, and High Temperature Stresses.* Academic Press.

Liebisch, F., Kirchgessner, N., Schneider, D., Walter, A., & Hund, A. (2015). Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. *Plant Methods*, *11*(1), 9.

Liu, K. S. (1997). Soybeans: Chemistry, Technology, and Utilization; Aspen Publication. *Inc.: New York*.

Long, N., Gianola, D., Rosa, G. J., & Weigel, K. A. (2011). Application of support vector regression to genome-assisted prediction of quantitative traits. *Theoretical and Applied Genetics*, *123*(7), 1065.

Looney, J. W. (1984). Modification of Arkansas water law: Issues and alternatives. *Ark. L. Rev.*, *38*, 221.

Lu, W., Wen, Z., Li, H., & Yuan, D. (2013). Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *126*(2), 425.

Ludlow, M. M., & Muchow, R. C. (1990). A critical evaluation of traits for improving crop yields in water-limited environments. In *Advances in agronomy* (Vol. 43, pp. 107–153). Elsevier.

Ma, Y., Kan, G., Zhang, X., Wang, Y., Zhang, W., Du, H., & Yu, D. (2016). Quantitative trait Loci (QTL) mapping for glycinin and β-conglycinin contents in soybean (Glycine max L. Merr.). *Journal of Agricultural and Food Chemistry*, *64*(17), 3473–3483.

Martin, D. L., Dorn, T. W., Melvin, S. R., Corr, A. J., & Kranz, W. L. (2011). Evaluating energy use for pumping irrigation water. In *Proceedings of the 23rd Annual Central Plains Irrigation Conference* (pp. 22–23).

McWilliams, D. A., Berglund, D. R., & Endres, G. J. (1999). Soybean growth and management quick guide.

Mengel, K. (2001). Principles of plant nutrition (Fifth ed.). Dordrecht;Boston;: Kluwer Academic Publishers.

Mian, M. A. R., Ashley, D. A., & Boerma, H. R. (1998). An Additional QTL for Water Use Efficiency in Soybean. *Crop Science*, *38*(2), 390. https://doi.org/10.2135/cropsci1998.0011183X003800020020x

Mian, M. A. R., Bailey, M. A., Ashley, D. A., & Wells, R. (1996). Molecular Markers Associated with Water Use Efficiency and Leaf Ash in Soybean. *Crop Science*, *36*(5), 1252. https://doi.org/10.2135/cropsci1996.0011183X003600050030x

Michel, S., Ametz, C., Gungor, H., Epure, D., Grausgruber, H., Löschenberger, F., & Buerstmayr, H. (2016). Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theoretical and Applied Genetics*, *129*(6), 1179–1189.

Nichols, D. M., Glover, K. D., Carlson, S. R., Specht, J. E., & Diers, B. W. (2006). Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. *Crop Science*, *46*(2), 834–839.

Niinemets, Ü., Bilger, W., Kull, O., & Tenhunen, J. D. (1999). Responses of foliar photosynthetic electron transport, pigment stoichiometry, and stomatal conductance to interacting environmental factors in a mixed species forest canopy. *Tree Physiology*, *19*(13), 839–852.

Nordborg, M., & Tavaré, S. (2002). Linkage disequilibrium: what history has to tell us. *TRENDS in Genetics*, *18*(2), 83–90.

Nuccio, M. L., Rhodes, D., McNeil, S. D., & Hanson, A. D. (1999). Metabolic engineering of plants for osmotic stress resistance. *Current Opinion in Plant Biology*, *2*(2), 128–134.

Nygren, A., & Ingram, T. (2016). Atmometer (ETgage) Installation Tips -. Retrieved January 29, 2019, from https://cropwatch.unl.edu/2016/atmometer-etgage-installation-tips

Onogi, A., Watanabe, M., Mochizuki, T., Hayashi, T., Nakagawa, H., Hasegawa, T., & Iwata, H. (2016). Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. *Theoretical and Applied Genetics*, *129*(4), 805–817.

Orf, J. H., Diers, B. W., & Boerma, H. R. (2004). Genetic improvement: conventional and molecular-based strategies. *Soybeans: Improvement, Production, and Uses*, (soybeansimprove), 417–450.

O'toole, J. C., & Bland, W. L. (1987). Genotypic variation in crop plant root systems. In *Advances in Agronomy* (Vol. 41, pp. 91–145). Elsevier.

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*(5), 2015–2022.

Paris, R. L., & Shelton, G. W. (2001). The uniform soybean tests, southern states. *USDA-ARS, Stoneville, MS*.

Pathan, S. M., Vuong, T., Clark, K., Lee, J.-D., Shannon, J. G., Roberts, C. A., … Hyten, D. L. (2013). Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. *Crop Science*, *53*(3), 765–774.

Phansak, P., Soonsuwon, W., Hyten, D. L., & Song, Q. (2016). Multi-Population Selective Genotyping to Identify Soybean [*Glycine max* (L.) Merr.] Seed Protein and Oil QTLs. *G3: Genes, Genomes, Genetics*, *6*(6), 1635–1648. https://doi.org/10.1534/g3.116.027656

Piyasatian, N., Fernando, R. L., & Dekkers, J. C. M. (2007). Genomic selection for marker-assisted improvement in line crosses. *Theoretical and Applied Genetics*, *115*(5), 665–674.

Poeta, F., Borrás, L., & Rotundo, J. L. (2016). Variation in seed protein concentration and seed size affects soybean crop growth and development. *Crop Science*, *56*(6), 3196–3208.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., … Sorrells, M. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome*, *5*(3), 103–113.

Purcell, L. C., & Specht, J. E. (2004). Physiological traits for ameliorating drought stress. *Soybeans: Improvement, Production, and Uses*, (soybeansimprove), 569–620.

Rebetzke, G. J., Richards, R. A., Fettell, N. A., Long, M., Condon, A. G., Forrester, R. I., & Botwright, T. L. (2007). Genotypic increases in coleoptile length improves stand establishment, vigour and grain yield of deep-sown wheat. *Field Crops Research*, *100*(1), 10–23.

Riley, T. L. (2014). Arkansas water use reporting. Retrieved January 29, 2019, from https://arwaterplan.arkansas.gov/issues%20&%20recommendations/march%202014%20Irrigation%20pres%20%20for%20SWP%20mtg%20.pdf

Ross, J. (2016). Arkansas soybean research studies, 2014. Retrieved February 3, 2019, from http://cdm16039.contentdm.oclc.org/cdm/ref/collection/p266101coll7/id/31967

Rutkoski, J., Benson, J., Jia, Y., Brown-Guedira, G., Jannink, J.-L., & Sorrells, M. (2012). Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *The Plant Genome*, *5*(2), 51–61.

Rutkoski, J. E., Heffner, E. L., & Sorrells, M. E. (2011). Genomic selection for durable stem rust resistance in wheat. *Euphytica*, *179*(1), 161–173.

Rutkoski, J. E., Poland, J. A., Singh, R. P., Huerta-Espino, J., Bhavani, S., Barbier, H., … Sorrells, M. E. (2014). Genomic selection for quantitative adult plant stem rust resistance in wheat. *The Plant Genome*, *7*(3).

Sebolt, A. M., Shoemaker, R. C., & Diers, B. W. (2000). Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Science*, *40*(5), 1438–1444.

Slafer, G. A., Araus, J. L., ROYO, C., & DEL MORAL, L. F. G. (2005). Promising eco-physiological traits for genetic improvement of cereal yields in Mediterranean environments. *Annals of Applied Biology*, *146*(1), 61–70.

Sloane, R. J., Patterson, R. P., & Carter, T. E. (1990). Field drought tolerance of a soybean plant introduction. *Crop Science*, *30*(1), 118–123.

Sonah, H., O'Donoughue, L., Cober, E., & Rajcan, I. (2015). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnology Journal*, *13*(2), 211–221. https://doi.org/10.1111/pbi.12249

Specht, J. E., Chase, K., Macrander, M., & Graef, G. L. (2001). Soybean Response to Water. *Crop Science*, *41*(2), 493. https://doi.org/10.2135/cropsci2001.412493x

Specht, J. E., Hume, D. J., & Kumudini, S. V. (1999). Soybean yield potential—a genetic and physiological perspective. *Crop Science*, *39*(6), 1560–1570.

Spradley, P. (2005). Crop Profile for Soybeans in Arkansas - (IPM) Database. Retrieved January 29, 2019, from https://ipmdata.ipmcenters.org/documents/cropprofiles/ARsoybean.pdf

Stearns, F. W. (2010). One hundred years of pleiotropy: a retrospective. *Genetics*, *186*(3), 767–773.

Tacker, P., & Vories, E. (1998). Irrigation. Arkansas Soybean Handbook, University of Arkansas. Cooperative Extension Service. Retrieved from https://www.uaex.edu/publications/pdf/mp197/chapter8.pdf

Taiz, L., & Zeiger, E. (2006). Water and plant cells. *Plant Physiology, 4th Edn. Sinauer Associates Inc*, *672*.

Termaat, A., & Munns, R. (1986). Use of concentrated macronutrient solutions to separate osmotic from NaCl-specific effects on plant growth.[Symposium paper]. *Australian Journal of Plant Physiology (Australia)*.

Tribout, T., Larzul, C., & Phocas, F. (2012). Efficiency of genomic selection in a purebred pig male line. *Journal of Animal Science*, *90*(12), 4164–4176.

Tuberosa, R. (2012). Phenotyping for drought tolerance of crops in the genomics era. *Frontiers in Physiology*, *3*, 347.

Valliyodan, B., Dan Qiu, Patil, G., Zeng, P., Huang, J., Dai, L., … Nguyen, H. T. (2016). Landscape of genomic diversity and trait discovery in soybean. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep23598

Vaughn, J. N., Nelson, R. L., Song, Q., Cregan, P. B., & Li, Z. (2014). The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. *G3: Genes, Genomes, Genetics*, g3–114.

Verma, D. P. S., & Shoemaker, R. C. (1996). *Soybean: genetics, molecular biology and biotechnology*. Cab International.

Wang, J., Chen, P., Wang, D., & Shannon, G. (2015). Identification and mapping of stable QTL for protein content in soybean seeds. *Molecular Breeding*, *35*(3). https://doi.org/10.1007/s11032-015-0285-6

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 X Danbaekkong soybean population. *Theoretical and Applied Genetics*, *128*(5), 839–850.

Wilcox, J. R. (2004). World distribution and trade of soybean. *Soybeans: Improvement, Production, and Uses*, (soybeansimprove), 1–14.

Wilson, J. R. (2016). Soybean Consumption on the Rise Worldwide. Retrieved January 29, 2019, from https://worldagnetwork.com/soybean-consumption-on-the-rise-worldwide/

Winthrop Rockefeller Foundation. (2008). Water issues in arkansas,A companion report to water issues in Arkansas – an unfinished story -summary report. Retrieved January 29, 2019, from http://www.wrfoundation.org/media/1387/waterissues_full_resources.pdf

Wong, C. K., & Bernardo, R. (2008). Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theoretical and Applied Genetics*, *116*(6), 815–824.

Xavier, A., Muir, W. M., & Rainey, K. M. (2016). Assessing predictive properties of genome-wide selection in soybeans. *G3: Genes, Genomes, Genetics*, g3–116.

Xu, S., Zhu, D., & Zhang, Q. (2014). Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proceedings of the National Academy of Sciences*, *111*(34), 12456–12461.

Zhang, J., Song, Q., Cregan, P. B., & Jiang, G.-L. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (Glycine max). *Theoretical and Applied Genetics*, *129*(1), 117–130.

Zhang, Y., Li, W., Lin, Y., & Zhang, L. (2018). Construction of a high-density genetic map and mapping of QTLs for soybean (Glycine max) agronomic and seed quality traits by specific length amplified fragment sequencing. *BMC Genomics*, *19*. https://doi.org/10.1186/s12864-018-5035-9

Zhong, S., Dekkers, J. C., Fernando, R. L., & Jannink, J.-L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics*.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., … Tian, Z. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, *33*(4), 408–414. https://doi.org/10.1038/nbt.3096

Zhu, M., & Zhao, S. (2007). Candidate gene identification approach: progress and challenges. *International Journal of Biological Sciences*, *3*(7), 420.

Table 1.1. Soybean water use (ET) by growth stage (Reference)

Crop Development	Water Use (mm/day)
Germination and seedling	1.27-2.54
Rapid vegetative growth	2.54-5.08
Flowering to pod fill (full canopy)	5.08-7.62
Maturity to harvest	1.27-5.08

Table 1.2. Water requirements for soybean during late reproductive stages through maturity (Table modified from Yonts, C.D. et al., 2008. Predicting the last irrigation of the season. NebGuide G1871. University of Nebraska-Lincoln Extension.)

Growth stage	Approximate days to maturity	Water use to maturity (mm)
End of pod elongation	37	228.6
Beginning seed enlargement	29	165.1
End of seed enlargement	18	88.9
Leaves begin to yellow	10	48.26
Maturity	0	0.0

Table 1.3. Allowable Deficits-Soybean as reported by Henry et al. (2014)

Predominant Soil	Flood, Furrow, Border (mm)	Sprinkler/Center Pivot (mm)
Clay	50.8	38.1
Silt loam w/pan†	44.45	31.75
Silt loam wo/pan ‡	63.5	50.8
Sandy loam	57.15	44.45
Sandy	50.8	38.1

† Wo/pan – without pan, without shallow restrictive layer

‡ W/pan – with pan, shallow restrictive layer at 25.4 cm or less below soil surface

Table 1.4. Atmometer Setting as reported by Henry et al. (2014). Yellow Soil moisture deficit

Stage of Growth	**31.75†**	**38.1**	**44.45**	**50.8**	**57.15**	**63.5**
V1 1st Node	137.16‡	165.1	193.04	220.98	248.92	276.86
V2 2nd Node	78.74	96.52	111.76	127	142.24	160.02
V3 3rd Node	53.34	63.5	73.66	83.82	96.52	106.68
V4	48.26	58.42	68.58	78.74	88.9	96.52
V5	43.18	50.8	58.42	68.58	76.2	83.82
V6	40.64	48.26	55.88	63.5	71.12	78.74
R1 Begin Bloom	38.1	45.72	53.34	60.96	66.04	73.66
R2 Full Bloom	35.56	43.18	48.26	55.88	63.5	71.12
FULL CANOPY	**33.02** ‡‡	**40.64**	**45.72**	**53.34**	**60.96**	**66.04**
R3 Begin Pod	33.02	40.64	45.72	53.34	60.96	66.04
R4 Full Pod	33.02	40.64	45.72	53.34	60.96	66.04
R5 Begin Seed	33.02	40.64	45.72	53.34	60.96	66.04
R6 Full Seed	33.02	40.64	45.72	53.34	60.96	66.04

† Value (in mm) of allowable deficits based on soil type and irrigation soil type (Table 3)

‡ Set Atmometer to this value based on soil type and irrigation system

‡‡ Set Atmometer to this value if canopy closes before growth stage is reached

Table 1.5. Major seed protein QTL identified using bi-parental mapping population

Population	Parent	Markers	QTL analysis	Number of QTL	Major QTL Chromosomes	G*E effect	References
60 F2:3 breeding lines	A81356022 X PI 468916	243 RFLP	Single-factor ANOVA	3	20, 15, 18	No	Diers et al. (1992)
Backcross populations BC4 and BC5 lines	A81356022 X PI 468916	SSR and AFLP	Single-factor ANOVA	3	20	No	Nichols et al. (2006)
60 to 100 F2-derived breeding lines	M82806 X HHP	21 to 85 RFLP	Single-factor ANOVA	9	20, 15, 18	No	Brummer et al. (1997)
120 F4-derived lines and 111 F2-derived lines	Young X PI 416937 and PI 97100 x 'Coker 237	RFLP	single-factor ANOVA and interval mapping	4	15	No	Lee et al. (1996)
BC3 F4:6 lines	Parker X PI 468916	SSR and RFLP	single-factor ANOVA	2	20	Yes	Sebolt et al. (2000)
75 F5-derived RILS	A3733 X PI 437088A	329 RPADS and 103 SSR	Linear regression, simple interval mapping, markers, composite interval mapping	3	20	Yes (oil)	Chung et al. (2003)
180 F2:4 lines	PI 97100 X Coker 237 and Young X PI416937	SSR and RAFLP	Single factor ANOVA	4	20, 15	No	Fasoula et al. (2004)
131 F6-derived lines RILs	Essex X Williams	SSR	Composite interval mapping	4	6	Yes	Hyten et al. (2004)

Table 1.5 (Cont.)

Population	Parent	Markers	QTL analysis	Number of QTL	Major QTL Chromosomes	G*E effect	References
101 F6–derived RILs	N87-984-16 X TN93-99	SSR	Single factor ANOVA, composite interval mapping	3	18	No	Panthee et al. (2005)
212 F2:9 RILs	ZDD09454 X Yudou12	SSR	Composite interval mapping	11	20, 18	Yes	Lu et al. (2013)
216 and 156 RILs	Magellan × PI 438489B and Magellan × PI 567516C	SSR and SNP	Interval mapping	4	15, 5, 6	Yes	Pathan et al. (2013)
242 and 214 RILs	R05-1415 × R05-638 and V97-1346 × R05-4256	SSR and/or SNP	Composite interval mapping	4	14, 20	No	Wang et al. (2015)
140 F5 -derived RILs	Benning X Danbaekkong	SSR and SNP	Composite interval mapping	4	14, 17, 15, 20	No	Warrington et al. (2015)
48 F2:3 lines	Multiple biparental populations	SNP	Interval mapping	35	20, 15, 10	No	Phansak et al. (2016)
148 F8:11 RILs	Huapidou X Qihuang 26	SLAF	Inclusive composite Interval Mapping and Composite interval mapping	35	5	No	Zhang et al. (2018)

Table 1.6. Major seed protein QTL identified using diverse germplasm (GWAS)

Population size	Markers	No. Loci	Chromosomes	References
298	SoySNP50K	13	20	Hwang et al. (2014)
3,000	SoySNP50K	20	20	Vaughn et al. (2014)
139	GBS-47K	8	20, 5, 8	Sonah et al. (2015)
302	WGRS	20	13, 03, 17, 12, 11, 15	Zhou et al. (2015)
>12,000	SoySNP50K	20	20, 15, 6	Bandillo et al. (2015)
106	WGRS	21	20	Valliyodan et al. (2016)

Table 1.7. Reported QTL related to soybean drought tolerance in bi-parental populations.

Population type	Pedigree	Traits	Markers	Major QTL observed	QTL analysis	References
120 F4	Yong X PI416937	WUE	RFLP	5	ANOVA	Mian et al. (1996)
116 F2	S-100 X Tokyo	WUE	RFLP	2	Single-factor ANOVA, interval mapping	Mian et al. (1998)
236 RILs	Minsoy X Noir 1	Yield	-	1	Composite interval mapping	Specht et al. (2001)
184 F2:7:11 lines	Kefeng1 X Nannong1138-2	Yield and drought soybean index	SSR	19	Composite interval mapping	Du et al. (2009)

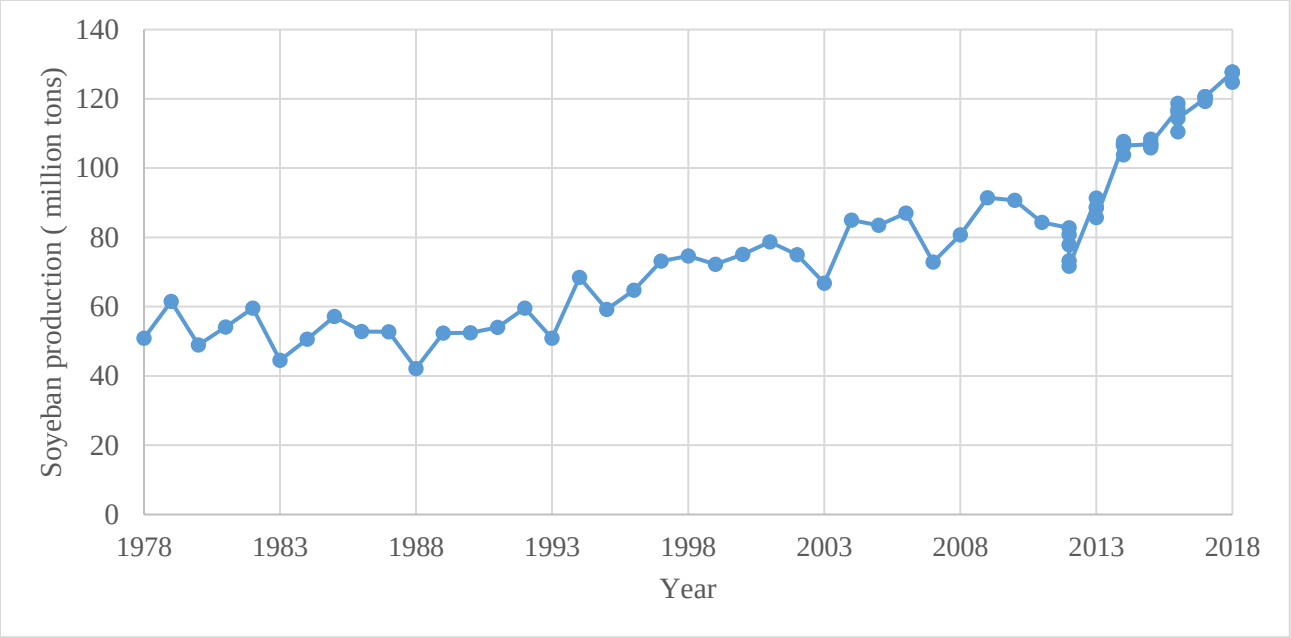

Figure 1.1. Soybean production in the United States (USA) over 40 years (1978-2018) by USDA-NASS

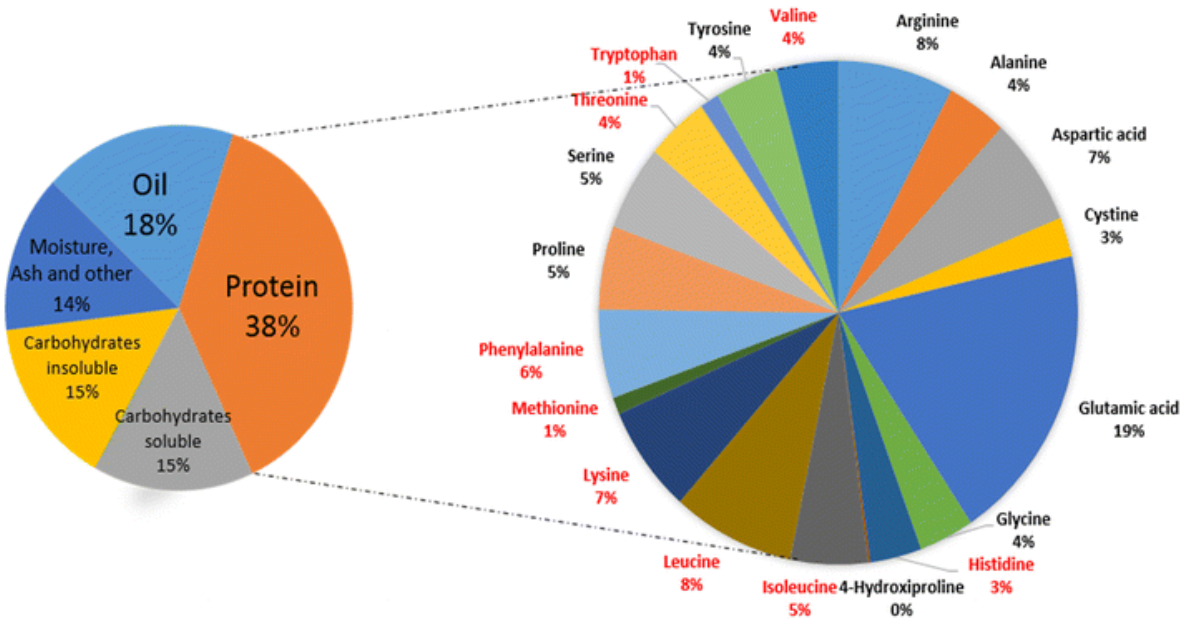

Figure 1.2. Amino acid composition of soybean protein. Red essential amino acid, black nonessential (Asif & Acharya, 2013)

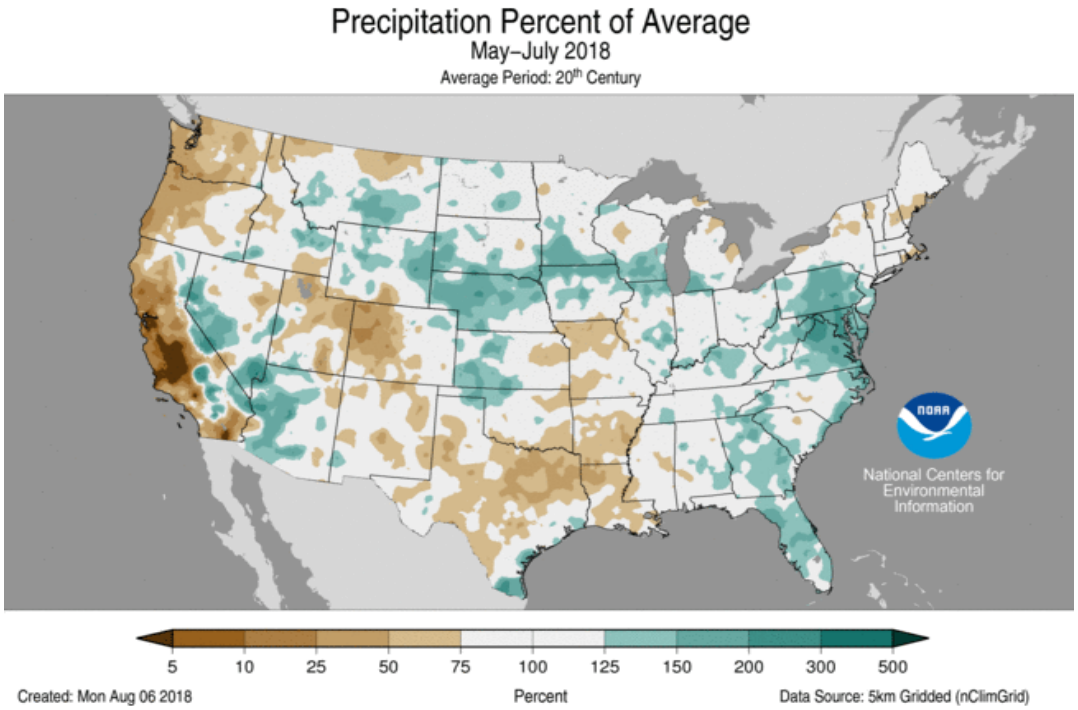

Figure 1.3. Deviation from average precipitation May to July 2018. Source NOAA, 2018.

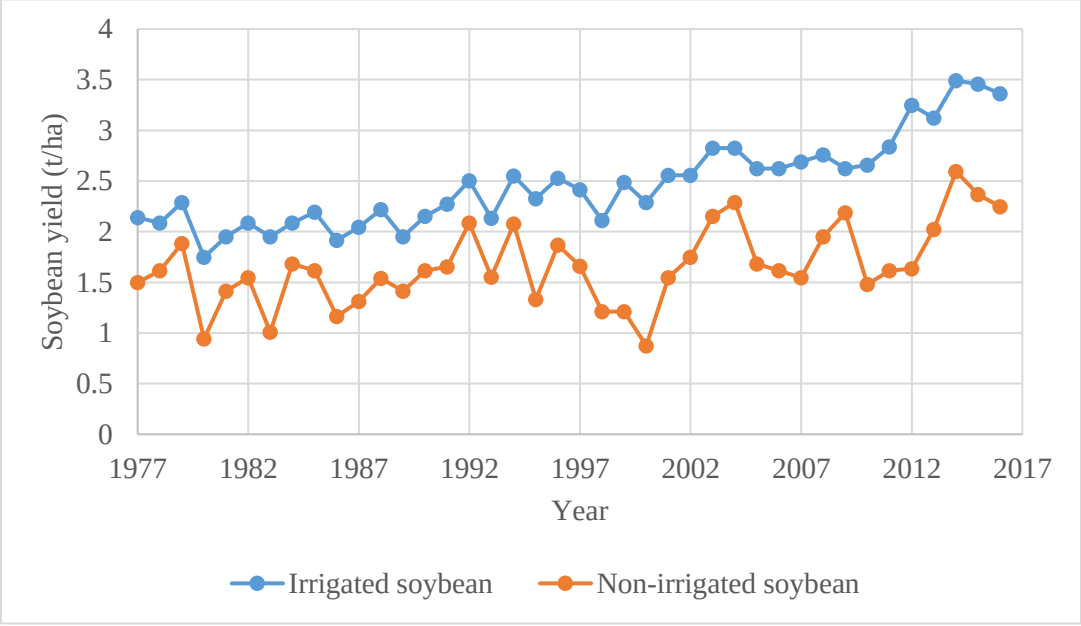

Figure 1.4. *Trend in soybean yield for 40 years (1977-2017) in Arkansas by USDA-NASS*

CHAPTER II

Genetics of seed protein and oil inherited from "BARC-7" soybean in two F2-derived mapping populations

Ravelombola, F.[1], Chen, P.[2], Vuong, T.[3], Nguyen, H.[3], Mian, R.[4], Acuña, A.[1], Florez-Palacios, L.[1], Wu, C.[1], Harrison, D.[1], deOliveira, M.[1], Winter, J.[1], Ford, M.[1], DaSilva, M.P.[1], and Mozzoni, L.[1]*

[1]Department of Crop, Soil, and Environmental Sciences, University of Arkansas, Fayetteville, AR 72701.

[2]Fisher Delta Research Center, University of Missouri, Portageville, MO 63873.

[3]Division of Plant Sciences, University of Missouri, Columbia, MO 65201.

[4]USDA – ARS, Soybean and Nitrogen Fixation Research, Raleigh, NC 27606.

*Corresponding author (lmozzon@uark.edu)

Ravelombola, F., Chen, P., Vuong, T., Nguyen, H., Mian, R., Acuña, A., ... & Mozzoni, L. (2022). Genetics of seed protein and oil inherited from "BARC-7" soybean in two F2-derived mapping populations. *Journal of Crop Improvement*, 1-17.

ABSTRACT

Soybean [*Glycine max* (L.) Merr.] seed protein inheritance has been extensively studied; however, genetics of high-protein 'BARC-7' soybean are still unknown. In this study, we used 250 F_2-derived lines from each of two soybean populations for quantitative trait loci (QTL) mapping. UA 5814HP, with high protein content, tracing to BARC-7 as maternal grandfather, was a common parent. Field experiments were conducted using a randomized complete block design with one replication across four environments. Seed protein and oil were quantified using near-infrared (NIR) instrument. Genetic linkage maps were constructed using the Infinium Soy6KSNP Beadchips. QTL analysis was performed using a composite interval mapping method. QTL for protein and oil were identified on chromosomes 6, 13, and 20. The known major QTL on chromosome 20 was not detected; but a novel QTL further downstream on chromosome 20 (only detected in population two) had high-protein alleles inherited from BARC-7-derived parent. Fine mapping efforts are currently ongoing for confirmation of these results.

INTRODUCTION

One of the biggest challenges for plant breeders is feeding an increasing world population, forecasted to reach ~9 billion by 2050. Soybean [*Glycine max* (L) Merr.], a legume species native to East Asia in the *Fabaceae* family, ranks among the most important food crops in the world. It is widely grown and known for its high protein and oil contents (Wilson, 2004). Soybean is valuable because its seeds have multiple applications in food, feed, pharmaceutical, and industrial enterprises (Birt et al., 2004; García et al., 1997; Lusas, 2004). Even though seed yield is the most important trait, breeders have recently given a significant attention to quality traits, such as high protein or modified oil content (Lee et al., 2019; Carneiro et al., 2020; Singh et al., 2020). Soybean seeds contain approximately 40% protein and 20% oil (Clemente & Cahoon, 2009); however, there is a negative correlation between yield and protein content in soybean (Novikova, 2018; Lee et al., 2019; Sobko et al., 2020; Finoto et al., 2021), and between seed protein and oil content (Lee et al., 2019). Broad-sense heritability of protein and oil content in soybean is relatively high, ranging between 0.57 to 0.97(Chung et al., 2003; Panthee et al., 2005; Jain, 2018; Tian, 2020; Jiang, 2020; Zhang, 2021; Arnold, 2021).

As many as 248 QTL associated with seed protein and 320 with seed oil have been reported in Soybase (Grant et al., 2010). Some QTL for protein and oil content were detected at the same position, suggesting either closely linked QTL or QTL with pleiotropic effect on both traits. Nonetheless, according to the rules established by the Soybean Genetics Committee (error rate lower than 0.01 and confirmation study showing alleles at the same locus are segregating in all the test populations), only two of those marker associations are accepted (Fasoula et al., 2004; Nichols et al., 2006). Two protein QTL, located on chromosomes (Chr.) 15 and 20, were commonly identified in several studies. Of these, QTL on Chr. 20 was considered a major QTL

with the highest proportion of phenotypic variance explained. The region of the interval on Chr. 15 is between 10 to 30 centimorgan (cM), whereas that on Chr. 20 is between 20 to 40 cM (Grant et al., 2010). Warrington et al. (2015) used a recombinant inbred line population derived from the cross Benning/Danbaekkong, and mapped a major protein QTL on Chr. 20 carrying the Danbaekkong allele that explained 55% of phenotypic variation in protein content.

Leffel (1992) released a series of high protein soybean germplasm lines, including BARC-6 (Reg.no.GP-127, PI 555396), BARC-7 (Reg.no. GP-128, PI 555397), BARC-8 (Reg. no. GP-129, PI 555398) and BARC-9 (Reg. GP-130, PI 555399). Among these, BARC-7 was derived from the cross of CX797-21/D80-6931. D80-6931 is a high protein maturity group (MG) VI BC3 line, in which PI 86490 was the high protein donor parent and 'Tracy' was the recurrent parent. BARC-7 is a MG IV germplasm line with purple flowers, tan and brown pods, and determinate stem growth habit. The mean seed protein of BARC-7 is 491 g kg^{-1} (Leefel, 1992). BARC-7 was a parent used in the breeding program at the University of Arkansas System Division of Agriculture, and had progeny with high seed protein levels, including 'UA 5814HP' (Chen et al., 2017) and 'R11-7999' (Florez-Palacios et al., 2020). However, the genetic architecture of protein and oil content in many BARC-7-derived soybean elite lines is unknown. Therefore, the goal of this study was to perform QTL mapping for seed protein and oil content in two breeding populations that traced high seed protein to BARC-7 soybean germplasm line.

MATERIALS AND METHODS

Plant materials and phenotyping

Four initial crosses (UA 5615C/UA 5814HP, UA 5115C/UA 5814HP, R13-359/UA 5814HP, R13-532/UA 5814HP) between high-yielding lines and the high-protein cultivar UA 5814HP were made at the Milo J. Shult Agricultural Research and Extension Center in

Fayetteville, AR, in 2017. 'UA 5814HP' (Chen et al., 2017) is the progeny of 'R95-1705' (high protein) and 'S00-9980-22' (regular protein content); 'R95-1705', in turn, is the progeny of 'Hutcheson' (regular protein) and 'BARC-7' (high protein) (Leffel, 1992). On the other hand, 'UA 5615C', 'UA 5515C' (Florez-Palacios et al., 2019), R13-359, and R13-532 are commodity MG5 soybean varieties and lines with regular seed protein levels. A total of 13, 19, 20, and 26 F_1 seeds of the four populations, respectively, were sent to Costa Rica during the winter of 2018, and bulk harvested by population. Approximately 800 seeds of F_2 generation for each population were then planted in 8 rows of 4.6 m length during the summer of 2018 in Fayetteville, AR. All parental lines were screened for marker polymorphisms using the Infinium Soy6KSNP Beadchips (Song et al., 2020) (data not shown). Based on parental polymorphism and agronomic field adaptation, two populations, UA 5115C/UA 5814HP (Pop1) and R13-532/UA 5814HP (Pop2), were selected for QTL analysis. A total of 250 F_2 plants for each selected population were randomly selected and individually harvested for generation advancement. A sample of 50 to 100 seeds per $F_{2:3}$ line was sent to a winter nursery (Costa Rica) in 2019 to advance the population and increase the number of seeds via bulk harvesting. Two-hundred fifty F_2-derived lines from each population were planted in four environments (location-year combination) for phenotyping, using a randomized complete block design (RCBD) with one replication. Environments included Upala, Costa Rica (inceptisol soil order -series unknown) in 2018 (18CR); Portageville, MO (Tiptonville silt loam soil (19MO)), and Rohwer, AR (Sharkey and Desha silt loam soils (19RO)) in 2019; and Fayetteville, AR (Captina silt loam soil in 2020 (20FA)). Plots single rows in 19CRwere 0.76 mapart with 4.6 m long with a 1.5 m alley as single rows in 19CR. For all the environments plot were two row plots and were 0.96 m, 0.81m and 0.91m apart with 4.6 m long with 1.5 m alley in 19RO, 20FA, 19MO respectively. Entries within

each population were randomly divided into four experiments, and each experiment had three checks (P53A67X, AG55X7, and AG56X8). A sub sample of 50-seed from each line was used for protein and oil estimation via Near-Infrared Spectroscopy using a DA 7250 NIR analyzer (Perten, Sweden).

Genotyping and QTL mapping

DNA was extracted from fresh young leaves using the hexadecymethylammonium bromide (CTAB) protocol (Doyle, 1990). Genotyping was done using the Infinium Soy6KSNP Beadchips (Song et al., 2020) in the Soybean Improvement Laboratory USDA-ARS, Beltsville.

Data were analyzed using analysis of variance (ANOVA) and treating environments as replications. Genotypes were treated as a fixed effect in JMP 16.0. The statistical model for the analysis was:

$$y_{ij} = \mu + g_i + b_j + \varepsilon_{ij}$$

where y_{ij} is the mean response (protein content or oil content) associated with the *i*th genotype in the *j*th environment, μ is the overall mean of protein or for oil content, g_i is the genotype effect (fixed effect), b_j is the environmental effect (fixed effect), and $\varepsilon_{ij}$is the experimental error associated with the *ij*th observation. Pearson correlation between protein and oil content across environments was computed using JMP 16.0. Broad-sense heritability was estimated using the following equation:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{l}}$$

where h^2 the is broad-sense heritability, σ_g^2 is the variance of the genotype, and σ_e^2 is the variance of error, and l is the number of environments.

Linkage genetic maps were constructed using JoinMap v.4.1 (Kyazma B.V., Wageningen, Netherlands). The segregation distortion for single nucleotide polymorphisms (SNPs) was analyzed through a Chi-square test. A total of 1,423 and 1,115 polymorphic markers for Pop1 and Pop2, respectively were used to construct the genetic map. The genetic distance was estimated using the Kosambi mapping function to address the inference. Based on the recombination frequencies, 24 linkages were created for Pop1 and Pop2, representing the 20 haploid chromosomes in the soybean genome.

QTL analysis was conducted through WinQTL cartograph v2.5 (Wang et al., 2012). Composite interval mapping (CIM) was the statistical model for the QTL search and to estimate the magnitude of their effects and their phenotypic variances. Cofactors were added on a backward regression analysis to increase the likelihood of finding a QTL. Genomic regions with a LOD (log-likelihood) >3 were considered significant QTL (Brody, 2019).

RESULTS

Phenotypic variation for seed protein and oil content

The phenotypic variation for seed protein content for each population across the four environments (18CR, 19RO, 19MO, and 20FA) showed a bell-shaped distribution, typical of quantitative traits **(Supplementary Figure S 2.1)**. In 18CR, seed protein content (on a % dry matter basis) for Pop1 and Pop2 ranged from 38.12 to 46.57 and 38.96 to 47.56, with a mean of 42.16±1.63% and 43.20 ± 1.47%, respectively. Similarly, for 19MO, seed protein showed a range (mean ± standard deviation) of 40.8% to 49.16% (44.8±1.54%) and 39.93% to 49.29%

(45.36±1.56%) for Pop1 and Pop2, respectively. In 19RO, range, mean and standard deviation for Pop1 and Pop2 were 42.5%-49.42% (45.35±1.23 %) and 41.06%-49.78% (45.83±1.35%), respectively. The observed range, mean and standard deviation were 38.43%-45.22% (41.98±1.09%) and 38.06%-45.39% (41.69±1.31%) for Pop1 and Pop2, respectively for 20FA **(Table 2.1)**. Analysis of variance revealed a high significant effect of environment on protein content in both populations **(Supplementary Tables S 2.1 & S 2.2).**

Oil content also showed a bell-shaped distribution **(Supplementary Figure S 2.2).** The mean oil content was 20.47±0.72%, 21.29±0.85%, 21.38±0.85% and 23.27±0.95% for Pop1 in 20FA, 19MO, 19RO, and 18CR, respectively. For Pop2, the oil content was 22.12±0.97% in 18CR, 20.28±0.91% in 19MO, 20.90±0.93% in 19RO, and 20.10±0.74% in 20FA. Similar to the protein content, environments showed highly significant effects on oil content **(Table 2.2**).

Trait and environment correlation, and heritability

Pearson correlation analysis showed a highly significant negative correlation ($p < 0.0001$) between seed oil and protein content varying from r= -0.545 to -0.775 for Pop 1 and r= -0.587 to -0.655 for Pop2 **(Figure 2.1)**. Additionally, correlation of protein levels among environments was found to be significant ($p<0.05$) and moderately positive, ranging from r = 0.28 to 0.47 and r = 0.27 to 0.47 for Pop1 and Pop2, respectively **(Figure 2.2)**. Similarly, correlation of oil levels among environments was positive, ranging from 0.23 to 0.43 for Pop1 and from 0.30 to 0.54 for Pop2 **(Figure 2.3)**.

The broad-sense heritability for protein and oil content (h^2) was 91.24% and 90.33% for Pop1, and 93.44% and 93.91% for Pop2, respectively. These results indicate that soybean seed

protein and oil contents were highly heritable and mainly influenced by genetic factors under our experimental conditions **(Supplementary Table S 2.1 & S 2.2).**

QTL analysis of seed protein

Analysis of QTL associated with protein content in individual environments for Pop1 showed eight QTL detected on six chromosomes (Gm03, Gm06, Gm13, Gm15, Gm16, and Gm18). Of these eight QTL, one was identified in 19RO, one in 19MO, two in 20FA, and four in 18CR **(Supplementary Table S 2.3)**. These QTL had absolute additive effects that ranged from 0.15 to 0.95, and explained 4% to 15% of the phenotypic variation. Of the QTL, a region on Chr. Gm03 covering a confidence interval of 28-49cM was observed both in 18CR and 20FA environments, albeit the actual SNP closest to peak was different. In addition, two nearby regions in Chr. Gm13 were associated with protein content in 19RO and 18CR **(Supplementary Table S 2.3)**. All other QTL were not consistent across environments. In an across-environment analysis, three QTL were identified on Chr. Gm06, Gm13, and Gm18, with an absolute additive effect of 0.42, 0.36, and 0.31, and explaining 12%, 9%, and 7% of phenotypic variation, respectively **(Table 2.3)**. The negative value of additive effects -0.36 (Chr.13 for Pop1) and -0.24, -0.34, -0.45, -0.11, -0.50 (Chr. Gm04, Gm05, Gm13, Gm16, and Gm20, correspondingly) for Pop 2 described in Table 3 indicated that favorable alleles for increasing protein were from UA 5814HP, except for the QTL on chromosomes Gm06 and Gm18.

In Pop2, for protein content, individual-environment results showed 14 QTL located on 10 chromosomes. Two QTL were found in the environments 18CR and 20FA, whereas four QTL were identified in the environment 19RO and six in 19MO **(Supplementary Table S 2.4)**. Absolute additive effects ranged from 0.03 to 0.81, with 2% to 13% of phenotypic variation explained. Similar to Pop1, the negative value of the additive effects was related to the favorable

allele for increasing protein coming from male parent UA 5814HP on chromosomes Gm04, Gm05, Gm6, Gm17, and Gm20. The QTL analysis across environment for Pop2 showed six QTL across six chromosomes (Gm04, Gm05, Gm06, Gm13, Gm16, and Gm20), with absolute additive effect values ranging from 0.11 to 0.51, and explaining 1% to 18% of the phenotypic variation **(Table 2.3)**.

Comparing results across populations, we observed that QTL Gm06_46078974_G_A on chromosome Gm06 was consistently identified in both Pop1 and Pop2 in the across-environment analysis, with mean effects of 0.42 and 0.51 for Pop1 and Pop2, respectively **(Table 2.3)**. However, the higher protein allele was not inherited from BARC-7. Additionally, a QTL on chromosome Gm13 was identified in both populations and traced to BARC-7, with 0.36 and 0.45 absolute allelic effects; however, the location in the linkage map was not the same for the Pop1 and Pop2 resulting in two different QTL **(Table 2.3)**.

QTL analysis of seed oil

A total of nine QTL for seed oil content were mapped in Pop1 (UA 5115C/UA 5814HP) in single-environment analyses **(Supplementary Table S 2.5)**. Of these, three QTL were mapped in 18CR, and two QTL each in 19MO, 19RO, and 20FA. These QTL were identified on chromosome Gm06 (3 QTL), Gm08 (1), Gm10 (1), Gm13 (2), Gm18 (1), and Gm20 (1). QTL on Gm06, Gm08, Gm10, Gm18 showed negative additive effects (-0.38, -0.31, -0.22, -0.02) and contributed an average of phenotypic variation of 8%, 4%, 10%, and 1%, respectively, indicating the negative effect on oil content from the UA 5814HP. The QTL on Gm06 (2 QTL), Gm13 (2), and Gm20 (1) had a positive additive effect, ranging from 0.10 to 0.64, explaining phenotypic variation from 5% to 33%, which indicated that they traced to the normal protein parent. An across-environment analysis revealed three QTL associated with oil content, found on

chromosomes Gm06, Gm13, and Gm15, with an additive effect of -0.22, +0.42, and +0.34, which explained an average of 9%, 15% and 23% of phenotypic variation (Table 4).

In Pop2 (R13-532/UA 5814HP), in single-environment analyses, a total of 18 QTL on 10 chromosomes were associated with seed oil content. Seven QTL were observed in the environment 18CR, 4 each in 19MO and 19RO, and 3 in 20FA. These QTL were found on Gm01 (1 QTL), Gm04 (3), Gm06 (3), Gm08 (2), Gm11 (1), Gm13 (3), Gm14 (1) Gm17 (1), Gm18 (1), and Gm20 (2); with a range between 3% and 22% of the phenotypic variation explained, and absolute additive effects varied from 0.07 to 0.52 (Supplementary Table S6). There were 10 QTL with negative additive effects, indicating that the alleles were from the parent UA 5814HP. In the across-environments analysis, we observed four QTL, mapped on Chr. Gm04, Gm06, Gm13, and Gm20, with LOD > 3.0 **(Table 2.4)**. Similar to the protein results, the QTL on Chr. Gm06 and Gm13 were detected for both populations and detected in most environments. We also observed in Table 4 that the additive effect for oil for Gm06 was negative (-0.22 for Pop1 and -0.29 for Pop2), and Gm13 was positive (0.42 for Pop1 and 0.28), which was in the contrast with the additive effect for protein on the same chromosome, as the high correlation coefficient between protein and oil contents.

DISCUSSION

The present study investigated the genetic control of the high protein content inherited from BARC-7 source, through QTL mapping in two F_2-derived populations from a cross between UA 5115C/UA 5814HP and R13-532/UA 5814HP, which were evaluated in four different environments. Both populations exhibited a typical normal distribution, with protein and oil ranges within expected values based on parents. UA 5814HP is a line with high protein, averaging 45.5% protein, and moderate oil content (20.5%) on a dry weight basis (Chen et al.,

2017); and although neither UA 5814HP nor the other parents were evaluated in the trials, the progeny showed a wide range for protein and oil values, as expected from transgressive segregation.

The average protein content of both populations was low in environments 18CR and 20FA compared to 19MO and 19RO. This is likely due to the higher temperature in 18CR and 20FA than in 19RO and 19MO during the growing season (data not shown). Conversely, higher temperature tends to increase oil content (Mourtzinis et al., 2017). That is the case in our current study, where environments 19CR and 20 FA had a high oil content compared to 19RO and 19MO. Similar observations on the effect of temperature on protein and oil were also reported in previous studies (Dornbos & Mullen, 1992; Piper & Boote, 1999; Specht et al., 2001; Mourtzinis, 2017; Novikova, 2018; Mertz-Henning et al., 2018; Lee et al., 2019, Sobko et al., 2020). Results also showed a negative correlation between protein and oil content. Highly negative phenotypic correlations between protein and oil are well documented in the literature (Cober & Voldeng, 2000; Assefa, 2018; Mertz-Henning et al., 2018; Novikova, 2018; Lee, 2019; Kambhampati, 2020; Yao, 2020; Li, 2021). This indicated that increasing seed protein concentration using phenotypic selection may occur at the expense of oil concentration and vice versa (Chung et al., 2003). In addition, we found a significant correlation between environments for protein and oil content. The correlation indicated that the environment is a factor that affects protein and oil; therefore, it is crucial to evaluate such traits across different environments in a mapping study. However, our study also showed high heritability for protein and oil content, which indicated that the traits were under a high level of genetic control. The high heritability of protein and oil has also been previously reported (Jain, 2018; Tian, 2020; Jiang, 2020; Zhang, 2021; Arnold, 2021).

Previous studies reported many QTL for protein content identified on Chr. Gm06, Gm15, Gm18, and Gm20 (Diers et al., 1992; Brummer et al., 1997; Warrington et al., 2015). In our study, we identified eight QTL associated with seed protein on Chr. Gm04, Gm05, Gm06, Gm13, Gm16, Gm18, and Gm20. The proportion of the phenotypic variance explained by a given QTL (R^2 value) is a parameter in deciding whether marker-assisted selection can be more efficient than conventional phenotypic selection alone (Bernardo, 2001; Bernardo & Charcosset, 2006). The QTL with large effect ($R^2 \geq 10\%$) were present on Chr. Gm06, Gm13 and Gm20. It is important to note that the major QTL identified in this study on Chr. Gm06 from Pop1, and on Chr. Gm13 from Pop1 and Pop2 have been reported in many previous studies **(Table 2.3)**. It is crucial to highlight that the QTL on Chr. Gm06 from Pop2 has been reported as "cqSeedProtein-012" (Pathan et al., 2013). Indeed, the "cq" designation in SoyBase, indicates a "confirmed QTL." There are only 16 QTL for protein listed as cq in SoyBase and the remaining QTL have not been confirmed to date (Grant et al., 2010). The major QTL on Chr. Gm20 (118.7-166.6 cM) from Pop2, and Chr. Gm06 (137-166.7 cM) from Pop1 have not been reported yet. Those could be potential novel QTL. In fact, the known QTL on Chr. Gm20 (20-40 cM) has been reported in multiple studies (Diers et al., 1992; Chung et al., 2003; Nichols et al., 2006; Bandillo et al., 2015; Warrington et al., 2015), but was not found to be significant in our study. Previous studies revealed that high protein alleles at that locus have historically been associated with decreases in both seed yield and oil content. Some sources of high protein alleles include PI 437088A, PI 407780A, PI 468916, and Danbaekkong (PI 619083) (Chung et al., 2003; Warrington et al., 2015; Kim et al., 2016; Diers et al., 1992; Nichols et al., 2006). The results of our study suggest that BARC-7 may carry alleles different from Danbaekkong; this could be useful for breeders to diversify sources of higher protein.

Apart from the major QTL, some QTL had relatively small effects and were environment-specific. The inconsistency of the protein QTL could be explained by either genotype specificity or sensitivity to environmental conditions (Patil et al., 2017). This could also be elucidated by the fact that the protein is a complex, quantitative, heritable trait controlled by multiple genes and affected by environmental conditions and each of them might express differently under given environments (Akond, 2014; Li et al., 2019). These inconsistent QTL across environments might bring the challenges for breeders to select by using a few markers in a breeding program.

A total of seven QTL associated with the oil content in this study were mainly detected on chromosomes Gm04, Gm06, Gm13, Gm15, and Gm20 for both populations across all environments. Of these, the QTL on Chr. Gm13 (175-181 cM) for Pop1 and on Chr. Gm04 for Pop2 (77.4-87.4 cM) are novel QTL. Akin to the protein, the inconsistency of the oil QTL can be explained by the specificity of the genotype or by its sensitivity to environmental conditions (Patil et al., 2017). The oil QTL that has the largest effect is the one on Chr. Gm15 for Pop1 (R^2=23%). Mao et al. (2013) reported the same region on Chr. Gm15 (SeedOil 43-15) in Soybase.

In our study, not all QTL associated with oil and protein content co-located in the same exact regions, in agreement with different studies of diverse genetic backgrounds (Feng, 2009; Pathan et al., 2013; Mao et al., 2013; Rossi et al., 2013). However, the oil QTL detected on Chr. Gm20 (R^2=22%) and on Chr. Gm13 (R^2=11%) did overlap with protein QTL in Pop2. Moreover, the sign of the additive effect is also flipped for the oil content as compared to those of the protein QTL. This emphasizes how protein and oil are correlated, as has also been reported in many previous studies (Cober & Voldeng, 2000; Assefa, 2018; Mertz-Henning et al., 2018;

Novikova, 2018; Lee, 2019; Kambhampati, 2020; Sobko et al., 2020Yao, 2020; Finoto et al., 2021; Li, 2021).

CONCLUSIONS

A preliminary mapping using 250 F_2-derived lines each from two populations, showed a QTL on Chr. Gm13, explaining approximately 10% of variation for seed protein content, and one QTL further downstream in Chr. 20 (only detected on population two), explaining 18% of protein variation. An ongoing fine-mapping using an advanced inbred line mapping approach will help confirm and fine-map the regions associated with high protein and oil in BARC-7 genetic background.

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*, *2014*.

Arnold, B., Menke, E., Mian, M. A., Song, Q., Buckley, B., & Li, Z. (2021). Mining QTL for elevated protein and other major seed composition traits from diverse soybean germplasm. *Molecular Breeding*, *41*(8), 1-18.

Assefa, Y., Bajjalieh, N., Archontoulis, S., Casteel, S., Davidson, D., Kovács, P., ... & Ciampitti, I. A. (2018). Spatial characterization of soybean yield and quality (amino acids, oil, and protein) for United States. *Scientific reports*, *8*(1), 1-11.

Carneiro, R. C., Duncan, S. E., O'Keefe, S. F., Yin, Y., Neill, C. L., & Zhang, B. (2020). Sensory and consumer studies in plant breeding: a guidance for edamame development in the US. *Frontiers in Sustainable Food Systems*, *4*, 124.

Bandillo, N., Jarquin, D., Song, Q., Nelson, R. L., Cregan, P., Specht, J., & Lorenz, A. (2015). A population structure and genome-wide association analysis on the USDA soybean germplasm collection.

Bernardo, R. (2001). What if we knew all the genes for a quantitative trait in hybrid crops?. *Crop Science*, *41*(1), 1-4.

Bernardo, R., & Charcosset, A. (2006). Usefulness of gene information in marker-assisted recurrent selection: A simulation appraisal. *Crop Science*, *46*(2), 614-621.

Birt, D. F., Hendrich, S., Alekel, D. L., & Anthony, M. (2004). Soybean and the prevention of chronic human disease. *Soybeans: Improvement, production, and uses*, *16*, 1047-1117.

Brody, L. 2019. LOD Score. *in* Talking Glossary of Genetic Terms. National Human Genome Research Institute. Available online at https://www.genome.gov/genetics-glossary/LOD-Score. Accessed 10/14/2021.

Brummer, E. C., Graef, G. L., Orf, J., Wilcox, J. R., & Shoemaker, R. C. (1997). Mapping QTL for seed protein and oil content in eight soybean populations. *Crop Science*, *37*(2), 370-378.

Chen, P., Florez-Palacios, L., Orazaly, M., Manjarrez-Sandoval, P., Wu, C., Rupe, J. C., ... & Robbins, R. T. (2017). Registration of 'UA 5814HP'Soybean with High Yield and High Seed-Protein Content. *Journal of Plant Registrations*, *11*(2), 116-120.

Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., ... & Specht, J. E. (2003). The seed protein, oil, and yield QTL on soybean linkage group I.

Clemente, T. E., & Cahoon, E. B. (2009). Soybean oil: genetic approaches for modification of functionality and total content. *Plant physiology*, *151*(3), 1030-1040.

Cober, E. R., & D Voldeng, H. (2000). Developing high-protein, high-yield soybean populations and lines. *Crop Science*, *40*(1), 39-42.

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*(5), 608-612.

Dornbos, D. L., & Mullen, R. E. (1992). Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. *Journal of the American Oil Chemists Society*, *69*(3), 228-231.

Fasoula, V. A., Harris, D. K., & Boerma, H. R. (2004). Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. *Crop Science*, *44*(4), 1218–1225.

Feng, J. P., Shi, Q. H., & Wang, X. F. (2009). Effects of exogenous silicon on photosynthetic capacity and antioxidant enzyme activities in chloroplast of cucumber seedlings under excess manganese. *Agricultural Sciences in China*, *8*(1), 40-50.

Finoto, E. L., Soares, M. B. B., Albuquerque, J. D. A. A. D., Monteiro Neto, J. L. L., Maia, S. D. S., Dona, S., ... & Abanto-Rodriguez, C. (2021). Selection of soybean genotypes for yield, size, and oil and protein contents. *Australian Journal of Crop Science*, *15*(1), 48-50.

Florez-Palacios, L., Orazaly, M., Manjarrez-Sandoval, P., Wu, C., Mozzoni, L., Dombek, D. and Chen, P., 2019. Registration of 'UA 5115C'Soybean Cultivar. Journal of Plant Registrations, 13(1), 38-41.

Florez-Palacios, L., Mozzoni, L., Orazaly, M., Manjarrez-Sandoval, P., Wu, C., Dombek, D., & Chen, P. 2020. Registration of soybean germplasm R11-7999 with high seed protein content and high yield. J. Plant Registrations 14, 82-86.

Food and Agriculture Organization of the United Nations. (2021). FAOSTAT statistical database. https://www.fao.org/faostat/en/#data/QCL/visualize , Retrieved October13, 2021.[Rome] :FAO.

Garcia, M. C., Torre, M., Laborda, F., & Marina, M. L. (1997). Rapid separation of soybean globulins by reversed-phase high-performance liquid chromatography. *Journal of Chromatography A*, *758*(1), 75-83.

Grant, D., Nelson, R. T., Cannon, S. B., & Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic acids research*, *38*(suppl_1), D843-D846.

Jain, R. K., Joshi, A., Chaudhary, H. R., Dashora, A., & Khatik, C. L. (2018). Study on genetic variability, heritability and genetic advance in soybean [Glycine max (L.) Merrill]. *Legume Research-An International Journal*, *41*(4), 532-536.

Jiang, G. L., Katuuramu, D. N., Xu, Y., Ren, S., & Rutto, L. K. (2020). Analysis and comparison of seed protein, oil, and sugars in edamame dried using two oven-drying methods and mature soybeans. *Journal of the Science of Food and Agriculture*, *100*(10), 3987-3994.

Kambhampati, S., Aznar-Moreno, J. A., Hostetler, C., Caso, T., Bailey, S. R., Hubbard, A. H., ... & Allen, D. K. (2020). On the inverse correlation of protein and oil: Examining the effects of altered central carbon metabolism on seed composition using soybean fast neutron mutants. *Metabolites*, *10*(1), 18.

Kim, M., Schultz, S., Nelson, R. L., & Diers, B. W. (2016). Identification and fine mapping of a soybean seed protein QTL from PI 407788A on chromosome 15. *Crop Science*, *56*(1), 219-225.

Lee, S., Van, K., Sung, M., Nelson, R., LaMantia, J., McHale, L. K., & Mian, M. R. (2019). Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. *Theoretical and Applied Genetics*, *132*(6), 1639-1659.

Leffel, R. C. (1992). Registration of high-protein soybean germplasm lines BARC-6, BARC-7, BARC-8, and BARC-9. *Crop science*, *32*(2).

Li, D., Zhao, X., Han, Y., Li, W., & Xie, F. (2019). Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. *Genomics*, *111*(1), 90-95.

Li, X., Wang, P., Zhang, K., Liu, S., Qi, Z., Fang, Y., ... & Ning, H. (2021). Fine mapping QTL and mining genes for protein content in soybean by the combination of linkage and association analysis. *Theoretical and Applied Genetics*, *134*(4), 1095-1122.

Lusas, E. W. (2004). Soybean processing and utilization. *Soybeans: Improvement, Production, and Uses*, (soybeansimprove), 949–1045.

Mao, T., Jiang, Z., Han, Y., Teng, W., Zhao, X., & Li, W. (2013). Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments. *Plant Breeding*, *132*(6), 630-641.

Mertz-Henning, L. M., Ferreira, L. C., Henning, F. A., Mandarino, J. M., Santos, E. D., Oliveira, M. C., ... & Neumaier, N. (2018). Effect of water deficit-induced at vegetative and reproductive stages on protein and oil content in soybean grains. *Agronomy*, *8*(1), 3.

Mourtzinis, S., Gaspar, A. P., Naeve, S. L., & Conley, S. P. (2017). Planting date, maturity, and temperature effects on soybean seed yield and composition. *Agronomy Journal*, *109*(5), 2040-2049.

Nichols, D. M., Glover, K. D., Carlson, S. R., Specht, J. E., & Diers, B. W. (2006). Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. *Crop science*, *46*(2), 834-839.

Novikova, L. Y., Seferova, I. V., Nekrasov, A. Y., Perchuk, I. N., Shelenga, T. V., Samsonova, M. G., & Vishnyakova, M. A. (2018). Impact of weather and climate on seed protein and oil content of soybean in the North Caucasus. *Vavilovskij Zbreve˘urnal Genetiki i Selekcii/Vavilov Journal of Genetics and Breeding*, *22*(6), 708-715.

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*(5), 2015–2022.

Pathan, S. M., Vuong, T., Clark, K., Lee, J. D., Shannon, J. G., Roberts, C. A., ... & Sleper, D. A. (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.

Patil, G., Mian, R., Vuong, T., Pantalone, V., Song, Q., Chen, P., ... & Nguyen, H. T. (2017). Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theoretical and Applied Genetics*, *130*(10), 1975-1991.

Piper, E. L., & Boote, K. I. (1999). Temperature and cultivar effects on soybean seed oil and protein concentrations. *Journal of the American Oil Chemists' Society*, *76*(10), 1233-1241.

Rossi, M. E., Orf, J. H., Liu, L. J., Dong, Z., & Rajcan, I. (2013). Genetic basis of soybean adaptation to North American vs. Asian mega-environments in two independent populations from Canadian× Chinese crosses. *Theoretical and applied genetics*, *126*(7), 1809-1823.

Singh, R. K., Prasad, A., Muthamilarasan, M., Parida, S. K., & Prasad, M. (2020). Breeding and biotechnological interventions for trait improvement: status and prospects. *Planta*, *252*(4), 1-18.

Sobko, O., Stahl, A., Hahn, V., Zikeli, S., Claupein, W., & Gruber, S. (2020). Environmental Effects on Soybean (Glycine Max (L.) Merr) Production in Central and South Germany. *Agronomy*, *10*(12), 1847.

Song Q, Yan L, Quigley C, Fickus E, Wei H, Chen L, Dong F, Araya S, Liu J, Hyten D, Pantalone V, Nelson RL (2020) Soybean BARCSoySNP6K: An assay for soybean genetics and breeding research The Plant J 104:800-811

Specht, J. E., Chase, K., Macrander, M., Graef, G. L., Chung, J., Markwell, J. P., ... & Lark, K. G. (2001). Soybean response to water: a QTL analysis of drought tolerance. *Crop science*, *41*(2), 493-509.

Tian, X., Zhang, K., Liu, S., Sun, X., Li, X., Song, J., ... & Ning, H. (2020). Quantitative trait locus analysis of protein and oil content in response to planting density in soybean (Glycine max [L.] Merri.) seeds based on SNP linkage mapping. *Frontiers in genetics*, *11*, 563.

Wang, S., Basten, C., & Zeng, Z.-B. 2012. Windows QTL cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC.

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*(5), 839-850.

Wilcox, J. R. (2004). World distribution and trade of soybean. *Soybeans: improvement, production, and uses*, *16*, 1-14.

Yao, Y., You, Q., Duan, G., Ren, J., Chu, S., Zhao, J., ... & Jiao, Y. (2020). Quantitative trait loci analysis of seed oil content and composition of wild and cultivated soybean. BMC plant biology, 20(1), 1-13.

Zhang, S., Hao, D., Zhang, S., Zhang, D., Wang, H., Du, H., ... & Yu, D. (2021). Genome-wide association mapping for protein, oil and water-soluble protein contents in soybean. *Molecular Genetics and Genomics*, *296*(1), 91-102.

Table 2.1. Descriptive statistics of seed protein content (% dry matter) in 250 F_2-derived breeding lines each of two different breeding populations when evaluated in four locations and assessed with near-infrared spectroscopy.

Environment†	Population‡	Mean±Sd§	CV (%)⁋	Minimum	Maximum	Range
18CR	Pop1	42.16±1.63	3.88	38.12	46.57	8.45
	Pop2	43.20±1.47	3.41	38.96	47.56	8.6
19MO	Pop1	44.8±1.50	3.36	40.80	49.16	8.36
	Pop2	45.33±1.57	3.47	39.93	49.29	9.36
19RO	Pop1	45.49±1.19	2.62	42.83	49.42	6.59
	Pop2	45.80±1.31	2.87	41.06	49.78	8.72
20FA	Pop1	41.98±1.09	2.59	38.43	45.22	6.69
	Pop2	41.69±1.31	3.16	38.06	45.39	7.33

† 18CR: 2018 Upala, Costa Rica; 19MO: 2019 Portageville, MO; 19RO: 2019 Rohwer, AR; 20FA: 2020 Fayetteville, AR

‡ Pop1: UA 5115C/UA 5814HP; Pop2: R13-532/UA 5814HP

§ Sd: standard deviation

⁋ CV: Coefficient of variation

Table 2.2. Descriptive statistics of seed oil content (% dry matter) in 250 F2-derived breeding lines each of two different breeding populations when evaluated in four locations and assessed with near-infrared spectroscopy.

Environment†	Population‡	Mean±Sd §	CV(%)⁋	Minimum	Maximum	Range
18CR	Pop1	23.27±0.95	4.08	20.44	26.18	5.74
	Pop2	22.12±0.97	4.41	19.2	24.63	5.43
19MO	Pop1	21.29±0.85	4.02	18.4	23.41	5.01
	Pop2	20.28±0.91	4.48	17.4	22.7	5.3
19RO	Pop1	21.38±0.85	4.01	17.82	23.63	5.81
	Pop2	20.90±0.93	4.47	18.14	23.77	5.63
20FA	Pop1	20.47±0.72	3.55	17.99	22.16	4.17
	Pop2	20.10±0.74	3.72	18.03	22.18	4.15

† 18CR: 2018 Upala, Costa Rica; 19MO: 2019 Portageville, MO; 19RO: 2019 Rohwer, AR; 20FA: 2020 Fayetteville, AR

‡ Pop1: UA 5115C/UA 5814HP; Pop2: R13-532/UA 5814HP

§ Sd: standard deviation

⁋ CV: Coefficient of variation

Table 2.3. QTL analysis of seed protein content for two breeding populations each consisting of 250 F2-derived breeding lines. Analysis conducted using composite interval mapping of phenotypic data consisting of least-square means from 4-environment 1-rep trials. Confidence intervals in centimorgan (cM) within each population. Previously reported QTL in the reported region are presented, as available on Soybase.org as of October 10, 2021.

Population	QTL	Chr. †	CI (cM)	LOD at QTL_Peak ‡	Closest Marker to QTL_Peak	Additive effect (%)§	R^2 (%)	Previously reported QTL within CI ⁋
UA 5115C/ UA 5814HP	qPROT_6	Gm06	137-166.7	4.9	Gm06_46078974_G_A	0.42	12%	-
	qPROT_13	Gm13	135.9-140.3	3.06	Gm13_32011502_T_C	-0.36	9%	SeedProtein 36-20, 36-21
	qPROT_18	Gm18	0-14.9	3.68	Gm18_2228646_C_T	0.31	7%	SeedProtein 20-1, 26-12, 28-5, 36-25, 47-7
R13-532/ UA5814HP	qPROT_4	Gm04	155.9-169.8	3.67	Gm04_47092275_T_C	-0.24	5%	-
	qPROT_5	Gm05	46-81.2	4.39	Gm05_3764264_C_T	-0.34	5%	SeedProtein 9-1, 12-1, 41-1
	qPROT_6	Gm06	97.6-126.1	8.65	Gm06_46078974_G_A	0.51	15%	cqSeedProtein-012, SeedProtein 13-2, 21-3, 24-1, 26-7, 28-1, 29-1, 35-1, 35-2, 36-7, 36-8
	qPROT_13	Gm13	49.8-74.5	6.1	Gm13_29524129_A_C	-0.45	11%	SeedProtein 21-6, 33-2, 36-18, 36-20, 36-21, 26-23, 26-24
	qPROT_16	Gm16	78.1-103.5	4.07	Gm16_33360539_T_C	-0.11	1%	SeedProtein 41-6
	qPROT_20	Gm20	118.7-166.6	6.1	Gm20_45327121_C_A	-0.5	18%	-

† Chromosome number

‡ Genome-wide 1000 permutation tests - LOD =3 (a threshold value)

§ Negative value: allele inherited from UA 5814HP; positive value: allele inherited from UA 5115C

⁋ Previously reported QTL in Soybase.org near particular QTL. A dash symbolized a new QTL.

Table 2.4. QTL analysis of seed oil content for two breeding populations each consisting of 250 F2-derived breeding lines. Analysis conducted using composite interval mapping of phenotypic data consisting of least-square means from 4-environment 1-rep trials. Confidence intervals in centimorgan (cM) within each population. Previously reported QTL in the reported region are presented, as available on Soybase.org as of October 10, 2021.

Population	QTL	Chr.†	CI (cM)	LOD at QTL_Peak ‡	Closest Marker to QTL_Peak	Additive effect (%)§	R^2(%)	Previously reported QTL within CI ⁋
UA 5115C/ UA 5814HP	qOIL_6	Gm06	151.3-153.4	2.65	Gm06_46321637_C_T	-0.22	9%	SeedOil 43-7
	qOIL_13	Gm13	175.5-181	2.96	Gm13_37339900_C_T	0.42	15%	-
	qOIL_15	Gm15	71.5-119.9	4.74	Gm15_12531884_A_G	0.34	23%	SeedOil 43-15
R13-532/ UA 5814HP	qOIL_4	Gm04	77.4-87.4	3.15	Gm04_9174100_G_A	-0.27	8%	-
	qOIL_6	Gm06	83.7-108	4.92	Gm06_18447419_G_A	-0.29	12%	SeedOil 23-1, 27-1, 30-5, 33-2
	qOIL_13	Gm13	54.5-74.5	4.68	Gm13_29524129_A_C	0.28	11%	SeedOil 36-5, 37-7
	qOIL_20	Gm20	102.6-163.6	6.87	Gm20_41356542_T_C	0.41	22%	SeedOil 13-5

† Chromosome number

‡ Genome-wide 1000 permutation tests - LOD =3 (a threshold value)

§ Negative value: allele inherited from UA 5814HP; positive value: allele inherited from UA 5115C

⁋ Previously reported QTL in Soybase.org near particular QTL. A dash symbolized a new QTL.

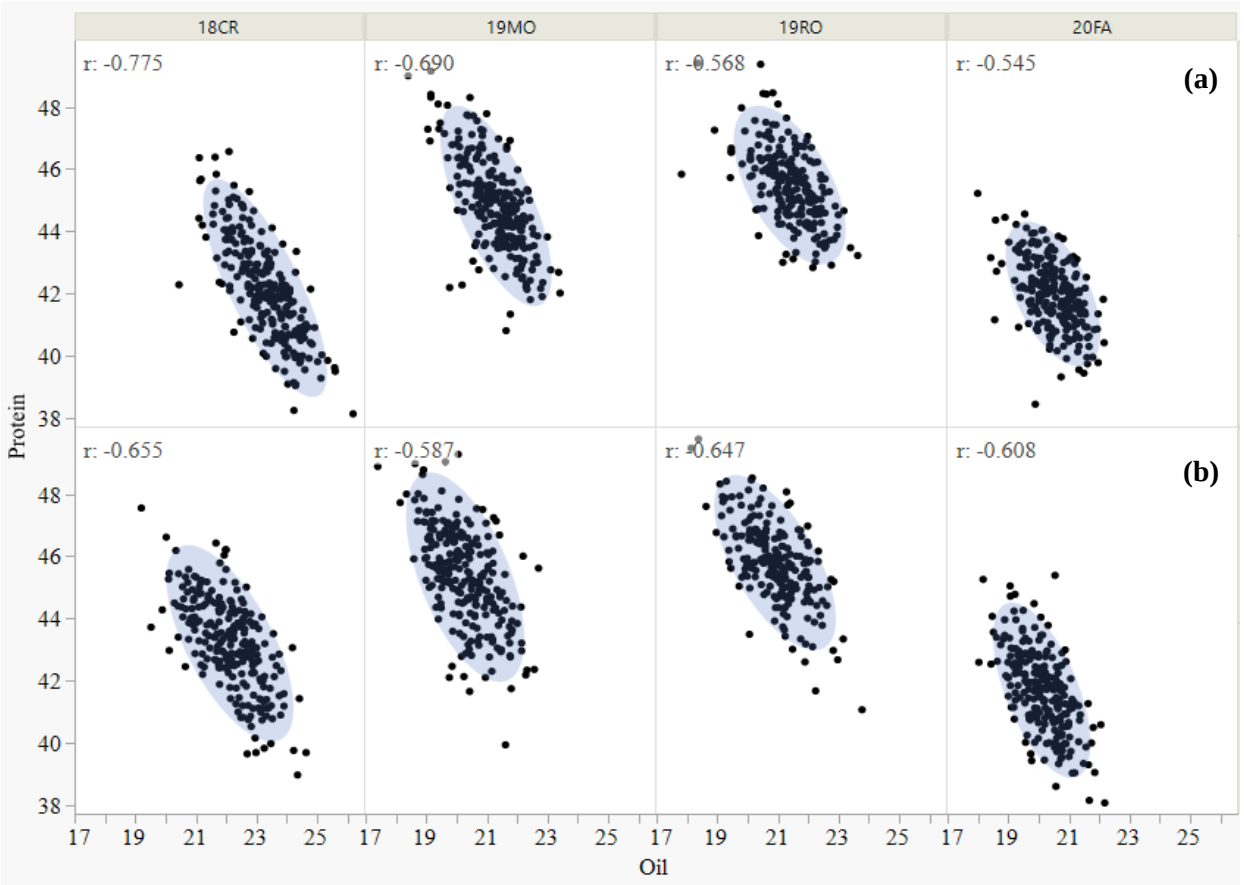

Figure 2.1. Pearson correlation analysis of oil content (%) and protein content (%) for 250 F2-derived lines derived from UA 5115C/UA 5814HP (a) and R13-532/UA 5814HP (b) each. The light blue color circle represents the 95% confidence interval for the correlation; the thinner the circle, the higher the correlation.

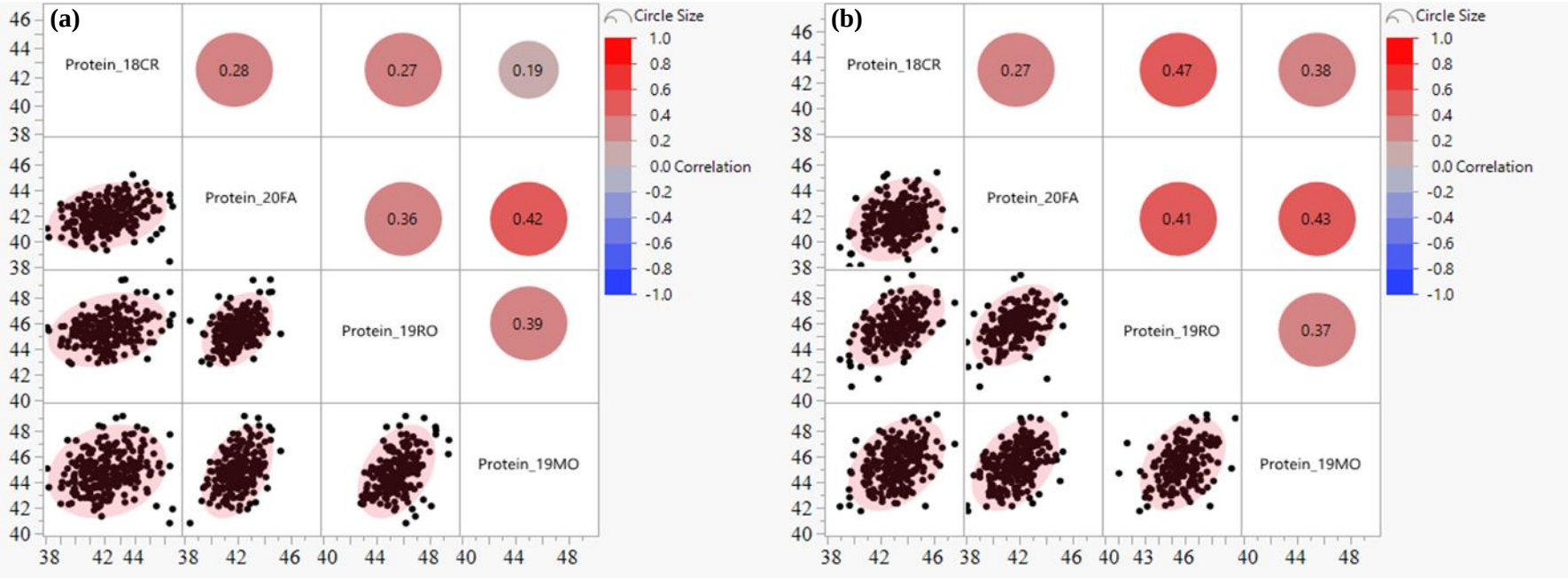

Figure 2.2. Pearson correlation analysis of seed protein among 250 F2-derived lines from UA 5115C/UA 5814HP (a) and R13-532/UA 5814HP (b). The shaded area represents the 95% confidence interval for the correlation.

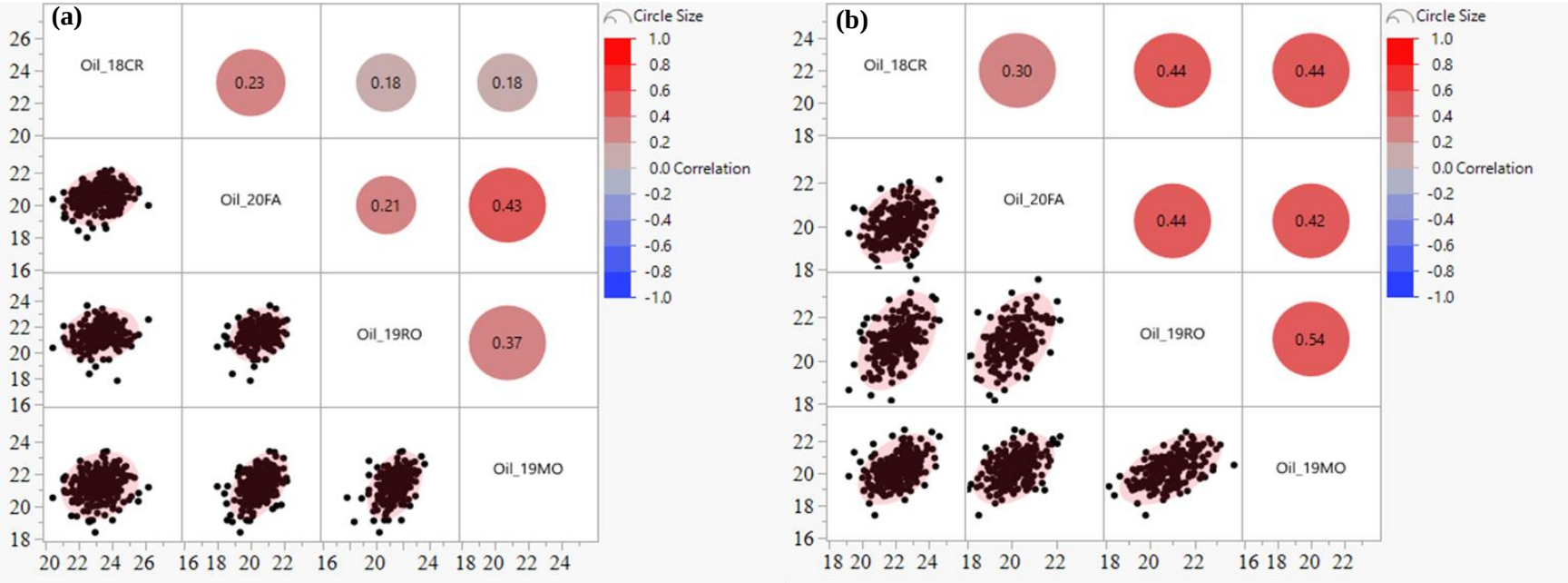

Figure 2.3. Pearson correlation analysis of seed oil among 250 F2-derived lines from UA 5115C/UA 5814HP (a) and R13-532/UA 5814HP (b) each. The shaded area represents the 95% confidence interval for the correlation.

Table S 2.1. ANOVA table for seed protein content for 250 F2-derived breeding lines originated from the cross of UA 5115C/UA 5814HP when grown on non-replicated RCBD planted in four environments.

Source	DF	Sum of Squares	Mean Square	F Ratio	p-value	h^2
Genotype	249	871.79	3.50	2.61	<.0001	91.24%
Environment	3	2291.47	763.82	568.39	<.0001	
Error	708	951.43	1.34			
Total	960	4114.69				

Table S 2.2. ANOVA table for seed protein content for 250 F2-derived breeding lines originated from the cross of R13-532/UA 5814HP when grown on non-replicated RCBD planted in four environments.

Source	DF	Sum of Squares	Mean Square	F Ratio	p-value	h^2
Genotype	249	1083.43	4.35	3.56	<.0001	93.44%
Environment	3	2620.26	873.42	714.90	<.0001	
Error	701	856.44	1.22			
Total	953	4560.13				

Table S 2.3. QTL analysis for seed protein content in 250 F2-derived progenies of UA 5115C/UA 5814HP

Environment	QTL	CI (cM)	LOD at QTL_Peak †	Closest Marker to QTL_Peak	Additive effect (%)‡	R^2 (%)
Across-environment	qPROT_6	137-166.7	4.9	Gm06_46078974_G_A	0.42	12%
	qPROT_13	135.9-140.3	3.06	Gm13_32011502_T_C	-0.36	9%
	qPROT_18	0-14.9	3.68	Gm18_2228646_C_T	0.31	7%
18CR	qPROT_3	28.4-49	5.65	Gm03_5796468_A_G	-0.75	6%
	qPROT_6	137.3-167.6	6.04	Gm06_46078974_G_A	0.95	15%
	qPROT_13	145.1-150.7	3.13	Gm13_33471044_G_A	-0.15	4%
	qPROT_18	0.8-7	3.07	Gm18_1010310_C_T	0.38	11%
20FA	qPROT_3	36.3-48.9	3.13	Gm03_6459920_A_G	-0.58	11%
	qPROT_15	128.6-143.5	3.32	Gm15_48596343_G_A	-0.47	12%
19MO	qPROT_16	35.2-56.1	3.4	Gm16_4351139_A_C	-0.56	7%
19RO	qPROT_13	134.8-144.5	3.13	Gm13_32011502_T_C	-0.56	13%

† Genome-wide 1000 permutation tests - LOD =3 (a threshold value)

‡ Negative value: allele inherited from UA5 814HP; positive value: allele inherited from UA 5115C

Table S 2.4. QTL analysis for seed protein content in 250 F2-derived progenies R13-532/UA 5814HP.

Environment	QTL	CI (cM)	LOD at QTL Peak †	Closest Marker to QTL_Peak	Additive effect (%)‡	R^2(%)
Across-environment	qPROT_4	155.9-169.8	3.67	Gm04_47092275_T_C	-0.24	5%
	qPROT_5	46-81.2	4.39	Gm05_3764264_C_T	-0.34	5%
	qPROT_6	97.6-126.1	8.65	Gm06_46078974_G_A	0.51	15%
	qPROT_13	49.8-74.5	6.1	Gm13_29524129_A_C	-0.45	11%
	qPROT_16	78.1-103.5	4.07	Gm16_33360539_T_C	-0.11	1%
	qPROT_20	118.7-166.6	6.1	Gm20_45327121_C_A	-0.5	18%
18CR	qPROT_6	83.6-108.6	7.47	Gm06_18447419_G_A	0.77	17%
	qPROT_20	165.4-167.2	3.1	Gm20_46056821_A_G	-0.43	6%
20FA	qPROT_4	146.5-170.4	6.71	Gm04_47092275_T_C	-0.63	13%
	qPROT_20	140.7-167.5	4.26	Gm20_45724030_T_C	-0.48	9%
19MO	qPROT_6	108.3-128.6	7.39	Gm06_46271407_G_A	0.81	16%
	qPROT_5	45.5-87	5.35	Gm05_31605772_A_G	-0.75	12%
	qPROT_5	96.3-108.6	3.54	Gm05_33176582_G_A	-0.54	12%
	qPROT_4	92.5-104.9	3.64	Gm04_27207244_A_G	0.15	3%
	qPROT_17	80.8-96.4	4.6	Gm17_13722544_A_G	0.15	5%
	qPROT_13	54.3-75	5.37	Gm13_29677928_G_T	-0.66	11%
19RO	qPROT_16	82.1-88.8	3.54	Gm16_32876100_A_G	-0.11	6%
	qPROT_2	1.7-20.6	4.04	Gm02_3091665_T_G	-0.03	2%
	qPROT_10	20.9-67.7	3.27	Gm10_7074398_A_G	0.54	10%
	qPROT_14	106.8-131	3.42	Gm14_46106800_T_C	-0.56	12%

† Genome-wide 1000 permutation tests - LOD =3 (a threshold value)

‡ Negative value: allele inherited from UA 5814HP; positive value: allele inherited from R13-532

Table S 2.5. QTL analysis for oil protein content in 250 F2-derived progenies of UA 5115C/UA 5814HP.

Location	QTL	CI (cM)	LOD at QTL_Peak †	Closest Marker to QTL_Peak	Additive effect (%)‡	R^2(%)
Combined location	qOIL_13	175.5-181	2.96	Gm13_37339900_C_T	0.42	15%
	qOIL_15	71.5-119.9	4.74	Gm15_12531884_A_G	0.34	23%
	qOIL_6	151.3-153.4	2.65	Gm06_46321637_C_T	-0.22	9%
CR	qOIL_6	39.8-44.9	2.53	Gm06_7323345_T_G	0.15	8%
	qOIL_6	142.7-152.5	3.26	Gm06_45768166_A_G	-0.38	8%
	qOIL_8	17.7-20	2.59	Gm08_14033412_T_G	-0.31	4%
FAY	qOIL_18	1.3-2.6	2.61	Gm18_930251_C_T	-0.22	10%
	qOIL_10	39.9-46.2	3.35	Gm10_4103498_G_T	-0.02	1%
PR	qOIL_13	50-62.1	3.38	Gm13_8529479_G_T	0.1	33%
	qOIL_13	163.1-192.7	5.6	Gm13_36224364_G_A	0.64	33%
ROH	qOIL_6	17.1-19.5	2.62	Gm06_3335673_A_G	0.3	5%
	qOIL_20	0-8.4	3.62	Gm20_827937_T_C	0.5	14%

† Genome-wide 1000 permutation tests - LOD =3 (a threshold value)

‡ Negative value: allele inherited from UA 5814HP; positive value: allele inherited from UA 5115C

Table S 2.6. QTL analysis for oil protein content in 250 F2-derived progenies of R13-532/UA 5814HP

Location	QTL	CI (cM)	LOD at QTL_Peak †	Closest Marker to QTL_Peak	Additive effect (%)‡	R^2 (%)
Combined location	qOIL_4	77.4-87.4	3.15	Gm04_9174100_G_A	-0.27	8%
	qOIL_6	83.7-108	4.92	Gm06_18447419_G_A	-0.29	12%
	qOIL_13	54.5-74.5	4.68	Gm13_29524129_A_C	0.28	11%
	qOIL_20	102.6-163.6	6.87	Gm20_41356542_T_C	0.41	22%
CR	qOIL_6	98.1-106.1	3.63	Gm06_18447419_G_A	-0.37	9%
	qOIL_6	108.8-117.9	3.68	Gm06_45768166_A_G	-0.31	9%
	qOIL_8	73.5-81.4	3.31	Gm08_15573572_T_G	-0.27	17%
	qOIL_8	87.7-89.5	3.18	Gm08_17171212_G_A	-0.27	17%
	qOIL_11	130.2-156.3	3.24	Gm11_38648336_A_G	0.23	6%
	qOIL_13	0-11	3.29	Gm13_20484995_T_G	-0.31	6%
	qOIL_14	100.5-127	3.17	Gm14_46247903_A_G	0.28	5%
FAY	qOIL_6	86.1-105.6	3.91	Gm06_18447419_G_A	-0.25	7%
	qOIL_17	61.1-77.9	3.73	Gm17_11272874_G_A	-0.27	6%
	qOIL_20	107.7-164.2	4.29	Gm20_45327121_C_A	0.38	16%
PR	qOIL_4	75.8-89.2	3.9	Gm04_9174100_G_A	-0.43	10%
	qOIL_13	54.3-74.7	5.25	Gm13_29677928_G_T	0.45	13%
	qOIL_20	102.6-163.3	6.06	Gm20_41356542_T_C	0.52	18%
	qOIL_1	0-5.3	3.14	Gm01_42848317_T_G	0.15	3%
ROH	qOIL_4	0-17.6	4.1	Gm04_1621110_C_A	0.07	10%
	qOIL_4	74.3-88.6	4.66	Gm04_10117285_G_A	-0.45	10%
	qOIL_18	97.8-142.8	4.81	Gm18_54021599_G_T	-0.62	22%
	qOIL_13	54.3-73.3	4.09	Gm13_29677928_G_T	0.37	10%

† Genome-wide 1000 permutation tests - LOD =3 (a threshold value)

‡ Negative value: allele inherited from UA 5814HP; positive value: allele inherited from R13-532.

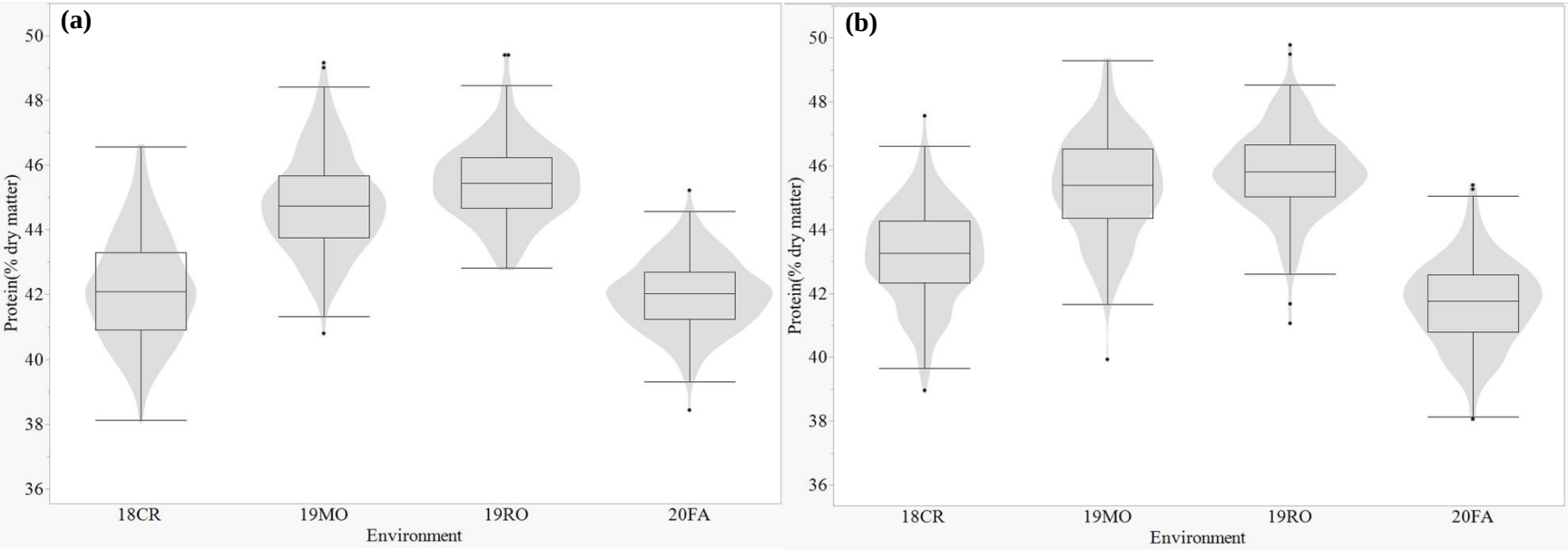

Figure S 2.1. Variation of seed protein (% dry matter) content in each location for the population derived from UA 5115C/UA 5814HP (a) and R13-532/UA 5814HP (b). The black line in the middle of the box shows the median, the white box indicates the range from the lower quartile to the upper quartile, the grey shaded line represents the dispersion and frequency distribution of the phenotypic data. The black dots represent phenotypic data that were extreme in each location

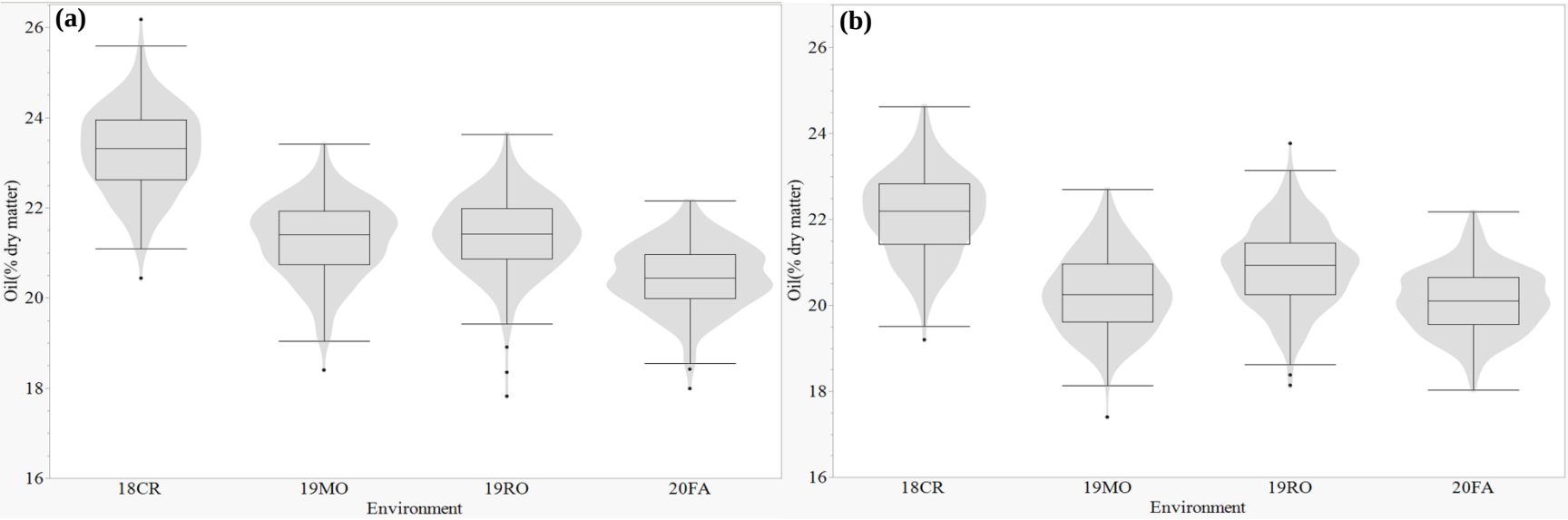

Figure S 2.2. Variation of seed oil (% dry matter) content in each location for the population derived from UA 5115C/UA 5814HP (a) and R13-532/UA 5814HP (b). The black line in the middle of the box shows the median, the white box indicates the range from the lower quartile to the upper quartile, the grey shaded line represents the dispersion and frequency distribution of the phenotypic data. The black dots represent phenotypic data that were extreme in each location.

CHAPTER III

Impact of Delaying Irrigation on Wilting, Seed Yield, and Other Agronomic Traits of Determinate MG5 Soybean

Ravelombola, F.[1], Acuña, A.[1], Florez-Palacios, L.[1], Wu, C.[1], Harrison, D.[1], deOliveira, M.[1], Winter, J.[1], DaSilva, M.P.[2], Roberts, T. [1], Henry, C.G. [1], Grignola, F. [1], Shakiba, E. [1], and Mozzoni, L. [1*]

[1]Department of Crop, Soil, and Environmental Sciences, University of Arkansas, Fayetteville, AR 72701.

[2]Department of Entomology and Plant Pathology, University of Arkansas, Fayetteville, AR, 72701, USA

*Corresponding author (lmozzon@uark.edu)

Ravelombola, F., Acuña, A., Florez-Palacios, L., Wu, C., Harrison, D., deOliveira, M., Winter, J., DaSilva, M.P., Roberts, T. , Henry, C.G. , Grignola, F. , Shakiba, E. , and Mozzoni, L. (2022). Impact of delaying Irrigation on Wilting, Seed Yield, and Other Agronomic Traits of Determinate MG5 Soybean. *Agronomy* (Accepted)

ABSTRACT

Soybean production in the U.S. Mid-South relies heavily on irrigation with 85% of soybean surfaces irrigated in Arkansas. Reduction in irrigation due to water quantity restrictions will significantly affect soybean seed yield, making variety selection increasingly important. The objective of the study was to assess if irrigation onsets at different reproductive stages affect wilting, seed yield and key agronomic traits on determinate maturity group 5 (MG 5) soybean. One-hundred sixty-five F_4-derived populations of recombinant inbred lines with determinate growth habit, similar maturity, and contrasting wilting potential were planted in an augmented strip-plot design in four environments as a single replicate. Four irrigation onsets were applied at R1 (initiation flower), R2 (full bloom), R3 (initiation pod), R4 (full pod) using an atmometer. Results indicated significant differences in wilting and yield but no significant differences in maturity, protein, oil concentration, and 100-seed weight across different irrigation onsets. There was no significant difference between the fast and slow wilting genotypes across different irrigation onsets for each trait. Allowable depletions measured in this study indicated that both fast-and slow-wilting soybean genotypes determinate MG5 can tolerate high allowable depletion with no significant yield penalty at R3 growth stage in silt loam soil.

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.]) is one of the most important worldwide crops, with a cultivated area of 126 million hectares (Mha) and a total production of 353 million tons in 2021, with 113 million tons produced by the United States (U.S.) (FAOSTAT, 2022). Soybean yield increased dramatically as a result of market development, breeding advances, and improved management during the Green Revolution (Pingali, 2012). It is one of the largest row crop in the U.S Mid-South accounting for 5.6 Mha for Missouri (MO), 3.6 Mha for Texas (TX), 3.1 Mha for Arkansas (AR), 2.1 Mha for Mississippi (MS), 1.5 Mha for Tennessee (TN), and 1 Mha for Louisiana (LA) (USDA, 2022).

Soybean production in the U.S. Mid-South relies heavily on irrigation (Reba & Massey, 2020), with 85% of soybean surfaces currently supplemented with water in Arkansas (AFBF, 2022). An average of 575 mm of water is needed during a soybean crop cycle (Dogan et al., 2007). Water is usually the main limiting factor for soybean productivity (Anda et al., 2020; Mekonnen et al., 2020; Ohashi et al., 2006; Sinclair et al., 2014; Xiong et al., 2021). Indeed, in a study of two soybean varieties under three irrigation levels, Anda et al. (2020) observed an impact of water availability on seed yield when a crop water stress index, proportional to the observed versus potential evaporation, was greater than a given threshold at reproductive soybean stages. Moreover, in a study of soybean water productivity under irrigated and dryland conditions, Mekonnen et al. (2020) observed large variations across the state of Nebraska, due to differences in climate, soil, water management, planting date, soybean maturity group/genetics, duration of the growing period.

The National Centers for Environmental Information reported a water shortage in eastern Arkansas between May and July 2018 (Young et al., 2018), causing a reduction in water levels at irrigation reservoirs and generating concern on water availability for crop irrigation during

reproductive stages. Also, some areas in AR, including key soybean growing counties like Lonoke, Prairie and Arkansas, have seen alluvial aquifer depth to water greater than 30 m, increasing irrigation costs and reducing well output (James et al., 2019). A reduction in groundwater availability could result in farmers having to skip or delay irrigation at a certain reproductive stage. Reduction in irrigation due to water restrictions will significantly affect soybean output and state revenue. As crop water availability becomes hard to predict at planting, soybean variety availability and selection become increasingly important for farmers. Yet, advances in genetic improvement for drought resistance in soybean is still limited (Fuganti-Pagliarini et al., 2017; Hufstetler et al., 2007; Yan et al., 2020).

Crops are subject to different abiotic and biotic stress during their growing season. Among abiotic stress, drought has been claimed to be the most devastating, having a drastic effect on productivity in rain-fed areas as it reduces plant growth and seed yield (Toker et al., 2007). According to Clement et al. (2008), as part of the *Fabaceae* family, soybean is one of the most drought-sensitive legumes. One typical feature of legume plants is the presence of nodules resulting from the relationship between plants and *Bradyrhizobium* spp. for biological N_2 fixation. However, this relation is particularly sensitive to drought (Marinković et al., 2019; Sheteiwy et al., 2021; Zahran, 1999). Hence, soybean seed yield might be reduced by 40% under water stress (Baghel et al., 2018; Dogan et al., 2007; J. E. Specht et al., 1999). Flowering (R1 and R2 stages (Fehr & Caviness, 1977)) and subsequent periods (pod setting: R3 and R4, and seed filling: R5 and R6 (Fehr & Caviness, 1977)) were found to be the most critical for water stress in soybeans (Buezo et al., 2019; Manavalan et al., 2009). Plants have different mechanisms to adapt to climatic variations by employing biochemical, molecular, physiological, and morphological changes (Moore et al., 2008). In soybean, drought tolerance mechanisms include:

increased rooting depth, reduced stomatal conductance, leaf rolling/folding, reduced evaporation surface, increased leaf-surface wax accumulation, and enhanced water-storage abilities in specific organs (Carrow, 1996; Fang & Xiong, 2015; Ludlow & Muchow, 1990; O'Toole & Bland, 1987). Canopy wilting is the first visible symptom of water stress, and a number of genotypes have been identified as slow wilting under field conditions (Carter, 1999; Carter et al., 2006). Slow wilting genotype could maintain cell turgor under drought condition (Devi & Sinclair, 2013; Sadok et al., 2012). In soybean, slow canopy wilting and sustained nitrogen fixation under drought have resulted in maximizing yield under water-limited environments (Sinclair et al., 2007).

In the U.S Great Plains, the effect of reduced irrigation on soybean has been well characterized (Kranz & Specht, 2012). However, no information is available on the impact of delaying irrigation in soybean in the Midsouth, and on relative performance of slow wilting versus non-slow wilting genotypes under delayed irrigation practices. Therefore, the research objective of this study was to assess if different irrigation onsets at different reproductive stages affect soybean seed yield and other key agronomic traits, including wilting, maturity, protein, oil concentration, and 100-seed weight, of determinate maturity group 5 (MG 5) soybean genotypes using contrasting wilting-potential populations.

MATERIALS AND METHODS

Plant materials and experimental design

Progeny from two populations were used in this study to magnify the presence of different genotypes with slow- and fast-wilting responses. The choice of maturity for this experiment was based on the availability of genetic material with slow wilting at the University of Arkansas, System-Division of Agriculture soybean breeding program. A total of 165 $F_{4:7}$

breeding lines (73 derived from the cross N07-14753/R11-1057 and 92 derived from R11-2933/R11-1057) were used, all of which had a determinate grow habit and similar maturity group 5 (MG 5), along with two parental checks (PC) (R11-2933 and R11-1057) **(Supplementary Figure S 3.1)** and two commercial checks (CC) based on seed availability (AG55X7, AG56X8, P53AG7X, P55A49X). Trials were grown in four environments (location-year combination) using an augmented strip-plot design under four furrow-irrigation onset treatments as a single replicate. Environments included Stuttgart, AR (silt loam soil) in 2019 and 2020 (19STU and 20STU), and Rohwer, AR (silt loam soil) in 2019 and 2020 (19ROH and 20ROH). The four irrigation onsets were: 1) full irrigation (irrigation initiated at flowering - R1 stage), 2) irrigation initiated at full flowering (R2 stage), 3) irrigation initiated at beginning of pod development (R3 stage), and 4) irrigation initiated when pods were 2 cm at one of the four uppermost nodes (R4 stage). The irrigation at each designated growth stage was triggered using the decision table developed by Henry et al. (2014) for atmometer (water-filled device measuring the actual evaporation of water) measurements based on 50% of the plots reaching the desired stage. The atmometer consists of a green canva cover with a ceramic plate on the top that simulates the transpiration of the leave surface (**Supplementary Figure S 3.2**). Each strip of irrigation onset (R1, R2, R3, and R4) was composed of ten blocks. One block was composed of four checks (two parental checks and two commercial cultivars) and 16 randomly assigned genotypes, including seven and nine genotypes from the first and second population, respectively, where individual lines were a random factor within populations **(Supplementary Figure S 3.3)**. The plots were 4.6 m long with 1.5 m alley, and consisted of two rows 0.97 m apart in 19ROH and 20ROH and 0.91 m apart in 19STU and 20STU. The planting date was 05/30/2019, 05/28/2019, 05/21/2020, and 05/19/2020 for 19STU, 19ROH, 20STU, and 20ROH,

respectively. Standard agronomic practices were used at each location, including fertilization to recommended levels as defined by Slaton et al. (2013).

Weather conditions, soil properties, soil water content and irrigation estimation

Weather data (air temperature and rainfall) and soil samples were collected to characterize the growing conditions. The average minimum temperature, maximum temperature, and rainfall for 19ROH, 19STU, 20ROH and 20STU were accessed from the Southern Regional Climate Center (www.srcc.lsu.edu/station_search), by searching for the respective weather stations where the study was conducted. The estimated cumulative potential evapotranspiration (PET) for each environment was calculated based on the sum of the daily potential evapotranspiration (ET_0). The ET_0 was estimated by the FAO–Penman–Monteith method (Allen et al., 1998) using the following equation:

$$ET_0 = \frac{0.408\,(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma\,(1 + 0.34u_2)}$$

where ET_0 is the daily potential evapotranspiration (mm.day^{-1}), R_n the net radiation at the crop surface (MJ.m-^{2}.day^{-1}), G the soil heat flux density (MJ.m-^{2}.day^{-1}), T the daily mean air temperature (°C), u_2 the wind speed at 2-m height (m.s^{-1}), e_s the saturation vapor pressure (kPa), e_a the actual vapor pressure (kPa), $(e_s - e_a)$ the saturation vapor pressure deficit (kPa), Δ the slope vapor pressure curve (kPa. °C^{-1}), and γ the psychrometric constant (kPa.°C^{-1}).

Soil samples were collected at 15-cm depth before planting and before harvest, for each environment (19ROH, 19STU, 20ROH, and 20STU). At each sampling date, composite samples from five random subsamples were collected on the sides of the planting bed (on each front half and back half of the field) to account for field heterogeneity. Samples were sent to the Agriculture Diagnostic Laboratory (University of Arkansas, Fayetteville) for analysis. Soil pH

and electrical conductivity (EC) were determined using a potentiometer at a ratio of 1:2 (w/v). Nutrients in the soil (P, K, Ca, Mg, S, Na, Fe, Mn, Zn, B, and Cu) were extracted with Mehlich-3 at a ratio of 1:10 (v/v) and their concentrations analyzed by inductively coupled argon-plasma spectrometry (ICAP, Spectro Analytical Instruments, Spectro Arcos ICP, Kleve, Germany). The sand, clay, and silt contents were expressed in percentage to determine the soil texture, based on the USDA soil texture triangle.

A total of 64 sensors (WATERMARK Soil Moisture Sensors – 200 SS, Irrometer, Riverside, California, USA) were installed in 2020 in Stuttgart (20STU) and Rohwer (20ROH) to measure the soil matric potential. For both 20STU and 20ROH, sensors were installed at 15-, 30-, 46-, and 76-cm depths, on the side of the bed at the ¾ of the field for each irrigation onset (R1, R2, R3, and R4). Two sets of 15- and 30-cm depths sensors were placed randomly in each irrigation onset. To quantify the water stress intensity, a manual reading of the watermark sensors was done at the time of the canopy wilting rating using a hand held meter (30 KTCD-NL, Irrometer, Riverside, California, USA). The volumetric water content (VWC) of the soil (%) at the time of canopy wilting rating was calculated using the average of the soil matric potential (SMP) at 15-cm and 30-cm depths (cbar), converted to VWC using soil water retention curve reported by Henry et al. (2020) for a silt loam soil in DeWitt (AR), a location close to the study site. The field capacity (FC) was 35.6% and the wilting point (WP) 8.9%. The available water content (AWC in %) was calculated to be 26.7% (FC-WP) (Henry et al., 2020). The allowable depletion (AD) at the time of canopy wilting rating and prior to each irrigation onset, which indicates the maximum amount of plant available water allowed to be removed from the soil before irrigation refill occurs, was then calculated using the following equation:

$$AD = \left(1 - \frac{(VWC - WP)}{AWC}\right) * 100$$

In addition to the watermark sensors, a time-domain reflectometer sensor (TDR-315L, Acclima) was installed at 10-cm depth on the side of the bed, in the middle of the field for 20ROH and 20STU to monitor the soil moisture in each irrigation onset (R1, R2, R3, and R4). The TDR sensors were coupled with AquaTrac (AgSense, Huron, SD) telemetry units for logging and transmitting the soil moisture data. TDR sensors were installed on 07/20/2022 in Stuttgart (20STU) and later in the season (08/06/2020) in Rohwer (20ROH), due to the unavailability of the equipment. The VWC measured from the TDR sensors before the irrigation was used to compute AD since it has a recording of 30-minute interval.

A flowmeter (McCrometer, California USA) was used in 20STU to measure the flow rate of irrigation. Readings were performed at the beginning and the end of each irrigation. Based on the flow rate and the surface irrigated, the amount of irrigation was calculated as follow:

$$Irr = \frac{(F_R - I_R) * 0.01}{S} * 25.4$$

where *Irr* is the irrigation amount (mm), F_R the flow rate at the end of irrigation (acre-inch), I_R the flow rate at the beginning of the irrigation (acre-inch) and S the surface irrigated (acres).

Since, the flowmeter was not available for 19STU, but an approximated equal surface was used for 20STU for each irrigation onset. The amount of irrigation for 19STU was estimated using the average acre-inch of the flow rate for 20STU. For 19ROH and 20ROH, the amount of irrigation was also an estimation using an average flow rate of 1000 gallon per minute and the duration of irrigation of each irrigation onset, the calculation was based on the following equation:

$$Irr = \frac{Flow\ rate * duration}{27514} * S * 25.4$$

where *Irr* is the irrigation amount (mm) and S the surface irrigated (acres).

Soybean phenotyping

Visual rating for canopy wilting was taken using a 10-point scale, from 0 to 9, where (0) represents no wilting and (9) a dead plant. The canopy wilting was rated one time prior to the onset of irrigation at the designated reproductive stage. Rating evaluations were performed at least six days of dry conditions (no precipitation), and to reduce the impact of diurnal variation in evaporative demand, data was collected between 11:00 am and 3:00 pm. Maturity was recorded when 95% of the pods in a plot had reached mature pod color (Fehr & Caviness, 1977), and was expressed as the number of days after 31st August. Total yield (t.ha^{-1}) was calculated from seed moisture, weight, and plot dimensions. Seed protein and oil estimation concentration in percentage of dry matter was performed for each line using subsamples of 50 seeds via Near-Infrared Spectroscopy DA 7250 NIR analyzer (Perten, Sweden). Lastly, 100-seed weights were evaluated in grams (g).

Data analysis

For the soil properties analysis, a matched paired *t-test* was performed in JMP Pro 16.0 (SAS Institute, Cary, NC) to analyze the difference between pre-planting and pre-harvest samplings within each irrigation onset in each environment. The canopy wilting for the slow wilting R11-2933 and the fast-witling parent R11-1057 at the R4 stage were compared using a t-test in JMP Pro 16.0 (SAS Institute, Cary, NC). Prior to analysis of each trait, lines were classified as fast or slow wilting based on the mean of the canopy wilting of R11-2933 ± σ, where σ is the sample standard deviation. Mean and standard deviation were calculated from a dataset of the four location-year environments under non-irrigated conditions at the R4 stage. Entry values greater than the mean plus the standard deviation (> 2.59) were considered a fast-

wilting genotype, while entries with values equal to or less than the mean plus the standard deviation (≤ 2.59) were considered slow wilting.

Data, including canopy wilting, maturity, yield, protein, oil, and 100 seed-weight, were analyzed as an augmented strip-plot design using PROC GLIMMIX in SAS (SAS Institute, Cary, NC). The statistical model for the analysis was the following:

$$y_{ijkl} = \mu + t_i + g_j + s_k + b_{k(l)} + gt_{ij} + ts_{ik} + gs_{jk} + e_{ijkl}$$

where *i* is the number of irrigation onsets: R1, R2, R3, and R4 (1, 2, 3, 4), *j* the number of witling classes: fast wilting and slow wilting classes (1,2), *k* the number of the environments: 19ROH, 19STU, 20ROH, and 20STU (1,2,3,4) and *l* the number of the blocks (1…10). y_{ijkl} is the mean response of the *ijkl*th observation (canopy wilting, maturity, yield, protein, oil concentration, and 100 seed weight), μ the overall mean response. Fixed effects were the irrigation onset t_i, the wilting class g_j and the interaction between the wilting class and the irrigation onset tg_{ij}. The random effects were the environment s_k, the block nested within the environment $b_{k(l)}$, the interaction between the irrigation onset and the environment ts_{ik}, the interaction between the wilting class and the environment gs_{jk} and the experimental error e_{ijkl}.

RESULTS

Environment characteristics

The four environments presented a silt loam soil texture **(Supplementary Table S 3.2).** Results in the difference between pre- and post-samplings showed an in-crease of 0.3 in pH and an average decrease of 134.3 μmhos.cm^{-1} (p-value < 0.05) in electrical conductivity (EC) for each irrigation onset (**Supplementary Figure S 3.6 and S 3.7**). We also measured a general decrease in macronutrients including P, K, S, Ca, and Mg and micronutrients (Fe, Mn, Zn)

between pre-planting and pre-harvest samplings (**Supplementary Table S 3.3**). However, an increase in Na was observed in the 19ROH environment with an increase of 14.42 $mg.kg^{-1}$ soil DW in R1 stage and 36 $mg.kg^{-1}$ soil DW in the R4 stage (**Supplementary Table S 3.3**). Likewise, a general increase in B was found in each environment except for 20STU in R2, R3, and R4 (**Supplementary Table S 3.3**).

The monthly average temperature from planting to harvest in each environment is given in **Supplementary Table S 3.1**. The period of July–August corresponds to the pod initiation stage, where the average maximum monthly temperature ranged from 30 to 33°C in each environment. On the day of the canopy wilting rating, the air temperature ranged between [20—30], [22—31], [23—36] and [23—33] °C at R1, R2, R3 and R4, respectively, for 19STU and between [20—30], [22—31], [23—36] and [23—34] °C at R1, R2, R3 and R4, respectively, for 19ROH. For 20STU, temperature ranged between [24—33], [23—33], [21—30] and [24—33] °C at R1, R2, R3 and R4 canopy wilting rating, respectively. Similar trend of temperature was recorded for 20ROH: [24—33], [23—31], [20—30] and [22—31] °C at R1, R2, R3 and R4 canopy wilting rating, respectively. The cumulative potential evapotranspiration (PET) indicated a high evaporative demand between June and July (**Supplementary Figure S 3.5**), where the highest cumulative PET was calculated in 20STU (400 mm). Pod initiation (R4 stage) was recorded at the end of July when the highest evaporative demand was recorded in each environment.

The cumulative rainfall from planting to harvest for each environment (19ROH, 19STU, 20ROH, 20STU) is given in **Supplementary Figure S 3.4**. Rainfall was more important in Rohwer (ROH) than in Stuttgart (STU) with respectively 587 mm and 541 mm in 2020 and respectively 471 mm and 417 mm in 2019, in Rohwer (ROH) and Stuttgart (STU). The

distribution of rainfall and irrigation during the growing season in each environment, and the time of canopy wilting rating is presented in **Figure 3.1 and 3.2**.

Prior to canopy wilting rating at R1, there had been no rainfall for 9 to 11 days in each environment (19ROH, 19STU, 20ROH, and 20STU). Rating at R2 was carried out when there had been no rain for 15 to 17 days for 19ROH, 20ROH, and 20STU. In 19STU, 4 days prior to the canopy wilting at R2, there had been 1.5 mm precipitation (**Figure 3.2**). Canopy wilting rating at R3 was recorded when there had been no rainfall for 23 days, 8 days and 7 days in 19ROH, 19STU and 20ROH, respectively. Five days prior to canopy wilting at R3, we recorded 2.79 mm of rainfall in 20STU. There had been no rainfall for 9 to 14 days prior to canopy wilting rating at R4 in 19ROH, 19STU and 20ROH while 3 mm of rainfall was recorded prior to the canopy wilting rating at R4 in 20STU **(Figure 3.2)**.

The hydraulic properties of the soil at the time of canopy wilting rating are given in **Table 3.1**. At the time of the canopy wilting rating, the soil matric potentials (SMP) in 20STU at 46-cm depth were -101, -58, -199 and -199 cbar at R1, R2, R3, and R4, respectively. At 76-cm depth, the SMP were -52, -77, -199 and -199 cbar for R1, R2, R3, and R4, respectively. R3 and R4 showed the lowest value of the average SMP at 15-30 cm depth (-199±0cbar) in 20STU. Similarly, in 20ROH, R3 and R4 presented the lowest SMP (-108 ± 21.1 and -124.5 ± 17.9 cbar) among the irrigation treatment. At R1 and R2, AD was relatively low in 20ROH compared to 20STU. The AD at R3 and R4 reached 55% in 20STU while in 20ROH, AD ranged from 33 to 38%. The lowest VWC was recorded at R3 and R4 for both environments, however, 20ROH displayed relatively higher VWC compared to 20STU at R3 and R4 (**Table 3.1**).

The irrigation was applied based on the atmometer, but the soil moisture content as well as the subsequent allowable depletion (AD) was determined before each irrigation onset (**Table**

3.2). In 20STU, irrigation onset at R4 stage presented the highest average of VWC before the onset of irrigation (19.5%). For both environments, the lowest values of VWC before irrigation were recorded at R1 with 11.1% and 11.4% for 20STU and 20ROH, respectively. The average allowable depletion was 92%, 88%, 75% and 60% when irrigation was onset at R1, R2, R3, and R4, respectively in 20STU. In 20ROH, the average depletion in each different irrigation onset was 91% at R1, 64% at R2, 71% at R3, and 87% at R4. Both environments had experienced water stress since allowable depletion exceeded 50% (assumed stress).

Canopy wilting

The *t*-test across four environments at R4 stage results showed a highly significant difference for canopy wilting (p-value < 0.0001) between the slow-witling parental check R11-2933 (1.59 ± 0.62) and the fast-wilting parental check R11-1057 (3.34 ± 0.87). Furthermore, a highly significant wilting class (Fast-wilting (FW) vs slow-wilting (SW)) effect on the canopy wilting (p-value < 0.001) was observed, results somewhat expected because of the nature wilting classes were constructed nonetheless demonstrating that the group means were statistically different and not just one standard deviation away. Moreover, there was a significant treatment effect (irrigation onset at different reproductive stages) on canopy wilting (p-value < 0.001), with a highly significant interaction effect between wilting class and irrigation treatment (p-value < 0.0001). As irrigation was further delayed, higher canopy wilting was observed. The FW group mean increased from 1.67 ± 0.32 when irrigation was triggered at R1, to 3.17 ± 0.32 when irrigation was triggered at R4. A significant difference in wilting severity between FW and SW occurred when the irrigation was triggered at R3 and R4 **(Figure 3.2)**.

Maturity

There was no significant wilting-class-by-treatment interaction on maturity (*p*-value = 0.53). Moreover, no significant difference was shown for treatment and maturity effect (*p*-value = 0.68), and for wilting class and maturity (*p*-value = 0.13). Delaying irrigation did not affect maturity under our experimental conditions **(Figure 3.3)**.

Seed yield

There was a significant treatment effect on seed yield (*p*-value < 0.05). However, there no was significant wilting class effect on yield (*p*-value = 0.05), and no interaction between the wilting class and the irrigation treatment (*p*-value = 0.33). Results showed that when irrigation was triggered at R4 stage, there was a significant yield reduction (23%) for determinate MG 5 soybeans. Nevertheless, no significant yield difference was reported by triggering irrigation at R1, R2, and R3 stages under our environmental conditions **(Figure 3.4)**. In addition, no yield differences between FW and SW genotypes were found under the delayed-irrigation methods **(Figure 3.4)**.

Seed protein and oil concentration, and 100-seed weight

There was no statistical difference for wilting-class-by-irrigation-treatment interaction in terms of seed protein (*p*-value = 0.6433) or oil content (*p*-value = 0.2603). Moreover, results showed that there was no effect of irrigation treatment on seed protein (*p*-value = 0.7939) and oil content (*p*-value = 0.8571). Likewise, there was no significant difference in terms of wilting class for protein (*p*-value = 0.3711) and oil content (*p*-value = 0.4423). Reduced irrigation did not affect protein and oil content for FW and SW determinate MG 5 even if irrigation was triggered at R4 stage **(Table 3.3)**. The average protein concentration ranged from 39.33 % to 40.08 %; while the oil concentration ranged from 21.28 % to 21 % (**Table 3.3**).

No significant interaction effect between wilting class and irrigation treatment (*p*-value = 0.8127) was found for 100-seed weight. Similarly, there was no statistical difference of irrigation treatment (*p*-value = 0.9885) or witling class (*p*-value = 0.5200) on 100-seed weight. Seed size was not affected if irrigation was delayed at R4 stage under our experimental conditions **(Table 3.3)**. The average 100-seed weight was 14.99 g to 15.43 g (**Table 3.3**).

DISCUSSION

Drought is one of the greatest threats to crop profitability. Thus, circumventing this problem is a priority for farmers (Lauer et al., 2012). When facing precipitation or groundwater shortages, Midsouth soybean famers might skip or delay irrigation at critical stages of soybean. The present investigation aims to appraise if different irrigation onsets at different reproductive stages affect soybean wilting, seed yield, and other key agronomic traits, including maturity, protein, oil, and 100-seed weight for determinate maturity group 5 (MG 5) soybean genotypes using contrasting wilting potential populations.

We observed an increase in soil pH with irrigation, in agreement with Bouaroudj et al. (2019). The increase in pH could be hypothesized of the result of high content of basic cations such as Na^{+}, Ca^{2+} and Mg^{2+} in the irrigation water, which increase the alkaline reserve of the soil and enhances the rate of denitrification thereby producing hydroxyl ions. Unlike Bouaroudj et al. (2019), we observed a decrease in EC with irrigation. In our experiments, pre-harvest soil samplings were done in October. The cumulative precipitation (mm) displayed a steep increase before harvest (September to October) **(Supplementary Figure S 3.4)**, indicating high precipitation right before harvest time, concomitantly decreasing EC as salts move with water. The general decrease in nutrients across environments in our results could be explained by the uptake of nutrients of soybean plants during their growth.

Our results showed an increase in canopy wilting when irrigation was delayed to R3 and R4 stages, and that the fast-wilting (FW) group had a significantly higher canopy wilting than slow-wilting (SW) group. The severity of canopy wilting in response to drought varies among soybean genotypes and the onset time (Carter, 1999; Sloane et al., 1990; Valliyodan et al., 2017). Higher canopy wilting at R1 and R2 stages compared to R3 and R4 stages could be explained by the greater soil water content at R1 (25.5 % and 33.5 %) and R2 stage (27.9 % and 28.8 %) (Table 1). Soybean genotype experienced more water stress at R3 stage and R4 stage in 20STU since the allowable depletion exceeded 50%. Moreover, the greater magnitude of canopy wilting response at R4 compared to earlier growth stages was because the evaporative demand was greater. In fact, the canopy was probably more fully closed (greater transpiration per unit area), and the soil moisture was more depleted at R4 stage. According to Valliyodan et al. (2017) and Charlson et al. (2009), as the soil dries, soybean with SW have delayed leaf wilting compared with FW, which agrees with the results of the current study. Under full soil moisture, plants will absorb water through its roots. This water will be used by the plant or released through transpiration by opening the stomata in the leaves. Photosynthesis will also occur normally with CO_2 and O_2 being absorbed and released through the open stomata. Once soil moisture becomes limited, water loss through transpiration still occurs; therefore, water loss leads to wilting. The first visible symptom is wilting under water stress (Carter, 1999; Carter et al., 2006). SW genotypes maintain cell turgor under drought condition (Devi & Sinclair, 2013; Sadok et al., 2012). SW mechanism, a basis for drought tolerance, has been studied by several researchers. Pantalone et al. (1996) stated that SW appeared to be involved as a better water resource exploration by a larger root system, while Tanaka et al. (2010) reported that the SW trait was due to a lower stomatal conductance. Bellaloui et al. (2013) speculated that the mechanisms were

related to the accumulation of minerals (such as K, Ca, B, Na) or organic compounds (such as sucrose, raffinose and stachyose and oleic acid) under drought stress in SW. Therefore, plants could maintain cell turgor, conserve water, and achieve osmoregulation. The higher leaf water potentials detected in SW genotypes suggested their ability to retain more water through water conservation and nutrient homeostasis (Bellaloui et al., 2013; Kunert & Vorster, 2020). Recently, Ye et al. (2020) confirmed the SW mechanism was linked to the water conservation strategy of limited maximum transpiration rates.

Delaying irrigation until R4 stage did not affect maturity of FW and SW in the current study. Determinate MG 5 soybean genotypes could sustain their development under mild drought. This phenomenon is valuable since it enables delaying irrigation without shortening the cycle of soybeans. Yield reductions in this study (23%) when irrigation was delayed at R4 stage were higher than previously reported by Karam et al. (2005) but lower than the studies carried out by Dogan et al. (2007) under non-irrigated conditions. In our study, yield results indicated that soybean was more sensitive to water stress at R4. At an early reproductive stage R1, R2 and R3, both fast-and slow-wilting genotypes were under water stress as AD was higher than 50% (Table 2). However, soybean genotypes can recover from any effect of moisture stress until R3 stage in silt loam soil as rainfall or irrigation were triggered; thus, it could compensate the deficit of water during R1, R2 and R3 stages. However, when irrigation was withheld at R4 in silt loam soil, both fast-and slow-wilting soybean genotypes experienced water stress (AD greater than 50%) that led to a decrease in seed yield. Indeed, the R4 stage has also been identified as the most critical drought-sensitive stage by Karam et al. (2005) and Smith et al. (2021). At R4, the plant reaches the full-pod stage in which the pod grows rapidly, and seed development begins. As a result of water stress, lower water potential in the leaves reduces the water potential

gradient between leaves and pods, reducing the flow of metabolites to the expanding cells (Westgate & Peterson, 1993). Water stress imposed on soybean throughout the growth stages reduces growth and affects seed yield (Eck et al., 1987; Kanungo et al., 2021; Yang et al., 2020), similar to our finding. In contrast, Sweeney & Granade (2002) and Marais & Bufé (2013) reported that water stress during flowering followed by full irrigations increased yield. This study is also in agreement with Foroud et al. (1993) and Huck et al. (1983), who reported that soybean yield components can recover from any effect of moisture stress at the R2 stage.

The SW phenotype has been used as one of the indicators to screen drought tolerance in the field (Charlson et al., 2009). This trait was predicted using a simulation model to improve yield under drought by > 80% of the growing seasons in most regions of the U.S (Sinclair et al., 2010). However, our investigation under reduced irrigation showed no statistical difference in seed yield of SW versus the FW genotypes. Similar to the current study, Ye et al. (2020) stated that under non-stress (irrigated) conditions, the FW recombinant inbred lines showed no statistically significant seed yield over the SW recombinant lines, but 12.8% to 13.7% yield advantage over the fast-witling lines under non-irrigated conditions. The disagreement of our research results and Ye et al. (2020) could result by the fact that we did not have a non-irrigated treatment and our environment received sufficient rainfall for a successful soybean crop development. In addition, Ye et al. (2020) indicates that when the yield was evaluated separately for each recombinant population, there was no significant difference between FW and SW under either condition (irrigated and rain-fed), which agrees with our results.

The effect of water deficit on soybean protein and oil concentration was evaluated in several studies, and different responses have been observed. Foroud et al. (1993) and Ghassemi-Golezani & Lotfi (2013) detected an increase in protein concentration under well-watered

conditions; contrarily, Specht et al. (2001), Rotundo & Westgate (2009), and Navabpour et al. (2017) found that water stress during soybean seed filling (R5 and R6) increased protein concentration and concluded that the increase in protein content could be due to the stimulation of protein synthesis rather from a concentration effect due to lower biomass production under the stress condition. In our studies, since irrigation was triggered before or at R4 stage (full pod development), we did not see a significant impact of irrigation treatment on seed protein concentration, as expected due to the timing of seed protein accumulation (Saldivar et al., 2011).

Oil has considerable importance to the soybean industry because of its high economic value as a source of edible oil and a major renewable feedstock for biodiesel production (Gashaw & Lakachew, 2014). Previous studies showed that drought stress reduced the oil concentration of seeds at later stages of grain filling in soybean (Ghassemi-Golezani & Lotfi, 2013; Martin et al., 2019). Indeed, Dornbos & Mullen (1992) found that serious water shortages during seed filling (R5 and R6) reduced seed oil concentration by 12.4%. On the contrary, Bellaloui et al. (2012) documented that severe drought can increase soybean oil seed concentration. In the present investigation, there was no significant difference in oil concentration regardless of the irrigation onsets (R1, R2, R3, and R4) and the wilting class. Under different irrigation onsets applied for the current study, the oil concentration was an average of 21%, which is above the minimum value of 20% required by the industry (Wilson, 2004).

Dogan et al. (2007) stated that water stress along with severe climatic conditions during R3 stage in soybeans increased pod numbers, resulting in lower yield and 1000-seed weights. In contrast, McWilliams et al. (1999), Desclaux et al. (2000), Clemente & Cahoon (2009), and Xiong et al. (2021) reported that if soybeans are under severe temperature and soil water stress

conditions, seed size will decrease. In our study we noted a trend towards the reduction in size as irrigation was delayed, but it was not of statistical significance at level of 5%.

CONCLUSIONS

Overall, no yield differences between FW and SW determinate MG 5 soybean genotypes under delayed irrigation were observed in the current study. As irrigation was further delayed, higher wilting severity occurred as water content is lower. Also, delaying irrigation until the R4 stage led to a reduction in seed yield. However, delaying irrigation did not affect maturity, protein, oil concentration, and 100-weight under our experimental conditions. Allowable depletions measured in this study indicated that both fast-and slow-wilting soybean genotypes determinate MG5 can tolerate high allowable depletion up to 90 % with no significant yield penalty at R3 stage in silt loam soil. The study suggests that even if high water deficits are experienced at early stages from delayed or inadequate irrigation that yields will likely not be significantly reduced in a furrow irrigation production system for soybean in silt loam. A deficit irrigation which is a water-saving irrigation strategy without compromising seed yield, could be implemented for farmers in the Mid -South as result of a groundwater shortage.

REFERENCES

AFBF. (2022). *Soybeans | Arkansas Farm Bureau*. https://www.arfb.com/pages/arkansas-agriculture/commodity-corner/soybean/

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome, 300(9), D05109.

Anda, A., Soós, G., Menyhárt, L., Kucserka, T., & Simon, B. (2020). Yield features of two soybean varieties under different water supplies and field conditions. *Field Crops Research*, *245*, 107673. https://doi.org/10.1016/j.fcr.2019.107673

Baghel, L., Kataria, S., & Guruprasad, K. N. (2018). Effect of static magnetic field pretreatment on growth, photosynthetic performance and yield of soybean under water stress. *Photosynthetica*, *56*(2), 718–730. https://doi.org/10.1007/s11099-017-0722-3

Bellaloui, N., Gillen, A., Mengistu, A., Kebede, H., Fisher, D., Smith, J., & Reddy, K. (2013). Responses of nitrogen metabolism and seed nutrition to drought stress in soybean genotypes differing in slow-wilting phenotype. *Frontiers in Plant Science*, *4*. https://www.frontiersin.org/article/10.3389/fpls.2013.00498

Bellaloui, N., Mengistu, A., Fisher, D. K., & Abel, C. A. (2012). Soybean Seed Composition Constituents as Affected by Drought and Phomopsis in Phomopsis Susceptible and Resistant Genotypes. *Journal of Crop Improvement*, *26*(3), 428–453. https://doi.org/10.1080/15427528.2011.651774

Bouaroudj, S., Menad, A., Bounamous, A., Ali-Khodja, H., Gherib, A., Weigel, D. E., & Chenchouni, H. (2019). Assessment of water quality at the largest dam in Algeria (Beni Haroun Dam) and effects of irrigation on soil characteristics of agricultural lands. *Chemosphere*, *219*, 76–88. https://doi.org/10.1016/j.chemosphere.2018.11.193

Buezo, J., Sanz-Saez, Á., Moran, J. F., Soba, D., Aranjuelo, I., & Esteban, R. (2019). Drought tolerance response of high-yielding soybean varieties to mild drought: Physiological and photochemical adjustments. *Physiologia Plantarum*, *166*(1), 88–104. https://doi.org/10.1111/ppl.12864

Carrow, R. N. (1996). Drought Avoidance Characteristics of Diverse Tall Fescue Cultivars. *Crop Science*, *36*(2), cropsci1996.0011183X003600020026x. https://doi.org/10.2135/cropsci1996.0011183X003600020026x

Carter Jr, T. E. (1999). Recent advances in breeding for drought and aluminum resistance in soybean. *Proceedings at the World Soybean Research Conference VI Chicago, IL, 1999*, 106–125.

Carter Jr, T. E., Orf, J., Purcell, L. C., Specht, J., Boerma, H. R., Chen, P., Sinclair, T., & Rufty, T. (2006). Tough times, tough plants-new soybean genes defend against drought and other stresses. *American Seed Trade Association Conference Proceedings*.

Charlson, D. V., Bhatnagar, S., King, C. A., Ray, J. D., Sneller, C. H., Carter, T. E., & Purcell, L. C. (2009). Polygenic inheritance of canopy wilting in soybean [Glycine max (L.) Merr.]. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *119*(4), 587–594. https://doi.org/10.1007/s00122-009-1068-4

Clement, M., Lambert, A., Herouart, D., & Boncompagni, E. (2008). Identification of new up-regulated genes under drought stress in soybean nodules. *Gene*, *426*(1–2), 15–22. https://doi.org/10.1016/j.gene.2008.08.016

Clemente, T. E., & Cahoon, E. B. (2009). Soybean Oil: Genetic Approaches for Modification of Functionality and Total Content. *Plant Physiology*, *151*(3), 1030–1040. https://doi.org/10.1104/pp.109.146282

Desclaux, D., Huynh, T.-T., & Roumet, P. (2000). Identification of Soybean Plant Characteristics That Indicate the Timing of Drought Stress. *Crop Science*, *40*(3), 716–722. https://doi.org/10.2135/cropsci2000.403716x

Devi, M. J., & Sinclair, T. R. (2013). Fixation Drought Tolerance of the Slow-Wilting Soybean PI 471938. *Crop Science*, *53*(5), 2072–2078. https://doi.org/10.2135/cropsci2013.02.0095

Dogan, E., Kirnak, H., & Copur, O. (2007). Deficit irrigations during soybean reproductive stages and CROPGRO-soybean simulations under semi-arid climatic conditions. *Field Crops Research*, *103*(2), 154–159. https://doi.org/10.1016/j.fcr.2007.05.009

Dornbos Jr., D. L., & Mullen, R. E. (1992). Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. *Journal of the American Oil Chemists' Society*, *69*(3), 228–231. https://doi.org/10.1007/BF02635891

Eck, H. V., Mathers, A. C., & Musick, J. T. (1987). Plant water stress at various growth stages and growth and yield of soybeans. *Field Crops Research*, *17*(1), 1–16. https://doi.org/10.1016/0378-4290(87)90077-3

Fang, Y., & Xiong, L. (2015). General mechanisms of drought response and their application in drought resistance improvement in plants. *Cellular and Molecular Life Sciences: CMLS*, *72*(4), 673–689. https://doi.org/10.1007/s00018-014-1767-0

FAOSTAT. (2022). *Crops and livestock products*. https://www.fao.org/faostat/en/#data/QCL/visualize

Fehr, W., & Caviness, C. E. (1977). Stages of soybean development. *Undefined*. https://www.semanticscholar.org/paper/Stages-of-soybean-development-Fehr-Caviness/ba57a53c2d3fae2d4665f17ddfa4a88ad573e0db

Foroud, N., Mündel, H.-H., Saindon, G., & Entz, T. (1993). Effect of level and timing of moisture stress on soybean plant development and yield components. *Irrigation Science*, *13*(4), 149–155. https://doi.org/10.1007/BF00190029

Fuganti-Pagliarini, R., Ferreira, L. C., Rodrigues, F. A., Molinari, H. B. C., Marin, S. R. R., Molinari, M. D. C., Marcolino-Gomes, J., Mertz-Henning, L. M., Farias, J. R. B., de Oliveira, M. C. N., Neumaier, N., Kanamori, N., Fujita, Y., Mizoi, J., Nakashima, K., Yamaguchi-Shinozaki, K., & Nepomuceno, A. L. (2017). Characterization of Soybean Genetically Modified for Drought Tolerance in Field Conditions. *Frontiers in Plant Science, 8*. https://www.frontiersin.org/article/10.3389/fpls.2017.00448

Gashaw, A., & Lakachew, A. (2014). Production of biodiesel from non edible oil and its properties. *Intl. J. Sci. Environ. Technol., 3*, 1544–1562.

Ghassemi-Golezani, K., & Lotfi, R. (2013). Influence of Water Stress and Pod Position on Oil and Protein Accumulation in Soybean Grains. *International Journal of Agronomy and Plant Production, 4*, 2341–2345.

Henry, C. G., Espinoza, L., Ismanov, M., Francis, P., & Ross, J. (2014). *Scheduling Irrigation using an Atmometer (ET Gauge) for Arkansas Soybeans*.

Henry,C.G., Pimentel,J.P., Gahr, P.N., and Clark, T., "Evaluating Irrigation Timing, Depletion, Water-use and Efficiencies in Furrow Irrigated-Rice.," in *B.R. Wells Arkansas Rice Research Studies 2020.*, pp. 255–258.

Huck, M. G., Ishihara, K., Peterson, C. M., & Ushijima, T. (1983). Soybean Adaptation to Water Stress at Selected Stages of Growth. *Plant Physiology*, *73*(2), 422–427.

Hufstetler, E. V., Boerma, H. R., Carter, T. E., & Earl, H. J. (2007). Genotypic Variation for Three Physiological Traits Affecting Drought Tolerance in Soybean. *Crop Science*, *47*(1), 25–35. https://doi.org/10.2135/cropsci2006.04.0243

James, B., Harvey, J., Forrest, B., Barke, C., Broach, J., Stewart, K., & Guess, M. (2019). *Arkansas ground water protection and management report for 2019* (p. 87). Arkansas Department of Agriculture. https://cdm16039.contentdm.oclc.org/digital/collection/p266101coll7/id/46311

Kanungo, M., Guruprasad, K. N., Kataria, S., Asa Dudin, G., Nasser Alyemeni, M., & Ahmad, P. (2021). Foliar application of fungicide-opera alleviates negative impact of water stress in soybean plants. *Saudi Journal of Biological Sciences*, *28*(5), 2626–2633. https://doi.org/10.1016/j.sjbs.2021.02.079

Karam, F., Masaad, R., Sfeir, T., Mounzer, O., & Rouphael, Y. (2005). Evapotranspiration and seed yield of field grown soybean under deficit irrigation conditions. *Agricultural Water Management*, *75*(3), 226–244. https://doi.org/10.1016/j.agwat.2004.12.015

Kranz, W. L., & Specht, J. E. (2012). Irrigating Soybean. *University of Nebraska-Lincoln Extension, Bulletin G1367*, 4.

Kunert, K., & Vorster, B. J. (2020). In search for drought-tolerant soybean: Is the slow-wilting phenotype more than just a curiosity? *Journal of Experimental Botany*, *71*(2), 457–460. https://doi.org/10.1093/jxb/erz235

Lauer, J. G., Bijl, C. G., Grusak, M. A., Baenziger, P. S., Boote, K., Lingle, S., Carter, T., Kaeppler, S., Boerma, R., Eizenga, G., Carter, P., Goodman, M., Nafziger, E., Kidwell, K., Mitchell, R., Edgerton, M. D., Quesenberry, K., & Willcox, M. C. (2012). The Scientific Grand Challenges of the 21st Century for the Crop Science Society of America. *Crop Science*, *52*(3), 1003–1010. https://doi.org/10.2135/cropsci2011.12.0668

Ludlow, M. M., & Muchow, R. C. (1990). A Critical Evaluation of Traits for Improving Crop Yields in Water-Limited Environments11This paper is based substantially on one first published by ICRISAT: Ludlow, M. M., and Muchow, R. C. (1988). Critical evaluation of the possibilities for modifying crops for higher production per unit of precipitation. ICRISAT (International Crops Research Institute for the Semi-Arid Tropics). In: Research on Drought Problems in the Arid and Semi-Arid Tropics. Proceedings of the International Consultants Meeting, 17–20 November 1986, pp. 179–211. ICRISAT Center, India. Patancheru, A.P. 502 324, India: ICRISAT. In N. C. Brady (Ed.), *Advances in Agronomy* (Vol. 43, pp. 107–153). Academic Press. https://doi.org/10.1016/S0065-2113(08)60477-0

Manavalan, L. P., Guttikonda, S. K., Phan Tran, L.-S., & Nguyen, H. T. (2009). Physiological and Molecular Approaches to Improve Drought Resistance in Soybean. *Plant and Cell Physiology*, *50*(7), 1260–1276. https://doi.org/10.1093/pcp/pcp082

Marais, D., & Bufé, M. M. (2013). *A summary of the effects of growth stage scheduled irrigation in soybean*.

Marinković, J., Bjelić, D., Đorđević, V., Balešević-Tubić, S., Jošić, D., & Vucelić-Radović, B. (2019). Performance of different Bradyrhizobium strains in root nodule symbiosis under drought stress. *Acta Physiologiae Plantarum*, *41*(3), 37. https://doi.org/10.1007/s11738-019-2826-9

Martin, A. R., Hayes, F. J., Borden, K. A., Buchanan, S. W., Gordon, A. M., Isaac, M. E., & Thevathasan, N. V. (2019). Integrating nitrogen fixing structures into above- and belowground functional trait spectra in soy (Glycine max). *Plant and Soil*, *440*(1), 53–69. https://doi.org/10.1007/s11104-019-04058-1

McWilliams, D. A., Berglund, D. R., & Endres, G. J. (1999). Soybean growth and management quick guide. *North Dakota State Univ, No. 58105*, 8.

Mekonnen, M. M., Hoekstra, A. Y., Neale, C. M. U., Ray, C., & Yang, H. S. (2020). Water productivity benchmarks: The case of maize and soybean in Nebraska. *Agricultural Water Management*, *234*, 106122. https://doi.org/10.1016/j.agwat.2020.106122

Moore, J. P., Vicré-Gibouin, M., Farrant, J. M., & Driouich, A. (2008). Adaptations of higher plant cell walls to water loss: Drought vs desiccation. *Physiologia Plantarum*, *134*(2), 237–245. https://doi.org/10.1111/j.1399-3054.2008.01134.x

Navabpour, S., Hezarjaribi, E., & Mazandarani, A. (2017). Evaluation of drought stress effects on important agronomic traits, protein and oil content of soybean genotypes.

Environmental Stresses in Crop Sciences, *10*(4), 491–503. https://doi.org/10.22077/escs.2017.61.1021

Ohashi, Y., Nakayama, N., Saneoka, H., & Fujita, K. (2006). Effects of drought stress on photosynthetic gas exchange, chlorophyll fluorescence and stem diameter of soybean plants. *Biologia Plantarum*, *50*(1), 138–141. https://doi.org/10.1007/s10535-005-0089-3

O'Toole, J. C., & Bland, W. L. (1987). Genotypic Variation in Crop Plant Root Systems. In N. C. Brady (Ed.), *Advances in Agronomy* (Vol. 41, pp. 91–145). Academic Press. https://doi.org/10.1016/S0065-2113(08)60803-2

Pantalone, V. R., Rebetzke, G. J., Burton, J. W., & Carter Jr., T. E. (1996). Phenotypic Evaluation of Root Traits in Soybean and Applicability to Plant Breeding. *Crop Science*, *36*(2), cropsci1996.0011183X003600020039x. https://doi.org/10.2135/cropsci1996.0011183X003600020039x

Pingali, P. L. (2012). Green Revolution: Impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(31), 12302–12308. https://doi.org/10.1073/pnas.0912953109

Purcell, L. C., Edwards, J. T., & Brye, K. R. (2007). Soybean yield and biomass responses to cumulative transpiration: Questioning widely held beliefs. *Field Crops Research*, *101*(1), 10–18. https://doi.org/10.1016/j.fcr.2006.09.002

Reba, M. L., & Massey, J. H. (2020). Surface Irrigation in the Lower Mississippi River Basin: Trends and Innovations. *Transactions of the ASABE*, *63*(5), 1305–1314. https://doi.org/10.13031/trans.13970

Rotundo, J. L., & Westgate, M. E. (2009). Meta-analysis of environmental effects on soybean seed composition. *Field Crops Research*, *110*(2), 147–156. https://doi.org/10.1016/j.fcr.2008.07.012

Sadok, W., Gilbert, M. E., Raza, M. A. S., & Sinclair, T. R. (2012). Basis of Slow-Wilting Phenotype in Soybean PI 471938. *Crop Science*, *52*(3), 1261–1269. https://doi.org/10.2135/cropsci2011.11.0622

Saldivar, X., Wang, Y.-J., Chen, P., & Hou, A. (2011). Changes in chemical composition during soybean seed development. *Food Chemistry*, *124*(4), 1369–1375. https://doi.org/10.1016/j.foodchem.2010.07.091

Sheteiwy, M. S., Ali, D. F. I., Xiong, Y.-C., Brestic, M., Skalicky, M., Hamoud, Y. A., Ulhassan, Z., Shaghaleh, H., AbdElgawad, H., Farooq, M., Sharma, A., & El-Sawah, A. M. (2021). Physiological and biochemical responses of soybean plants inoculated with Arbuscular mycorrhizal fungi and Bradyrhizobium under drought stress. *BMC Plant Biology*, *21*(1), 195. https://doi.org/10.1186/s12870-021-02949-z

Sinclair, T. R., Marrou, H., Soltani, A., Vadez, V., & Chandolu, K. C. (2014). Soybean production potential in Africa. *Global Food Security*, *3*(1), 31–40. https://doi.org/10.1016/j.gfs.2013.12.001

Sinclair, T. R., Messina, C. D., Beatty, A., & Samples, M. (2010). Assessment across the United States of the Benefits of Altered Soybean Drought Traits. *Agronomy Journal*, *102*(2), 475–482. https://doi.org/10.2134/agronj2009.0195

Sinclair, T. R., Purcell, L. C., King, C. A., Sneller, C. H., Chen, P., & Vadez, V. (2007). Drought tolerance and yield increase of soybean resulting from improved symbiotic N2 fixation. *Field Crops Research*, *101*(1), 68–71. https://doi.org/10.1016/j.fcr.2006.09.010

Slaton, N., Roberts, T., & Ross, J. (2013). Fertilization and liming practices. *Arkansas Soybean Production Handbook*, 2–16.

Sloane, R. J., Patterson, R. P., & Carter Jr., T. E. (1990). Field Drought Tolerance of a Soybean Plant Introduction. *Crop Science*, *30*(1), 118–123. https://doi.org/10.2135/cropsci1990.0011183X003000010027x

Smith, S. D., Bowling, L. C., Rainey, K. M., & Cherkauer, K. A. (2021). Quantifying Effects of Excess Water Stress at Early Soybean Growth Stages Using Unmanned Aerial Systems. *Remote Sensing*, *13*(15), 2911. https://doi.org/10.3390/rs13152911

Specht, J. e., Chase, K., Macrander, M., Graef, G. l., Chung, J., Markwell, J. p., Germann, M., Orf, J. h., & Lark, K. g. (2001). Soybean Response to Water: A QTL Analysis of Drought Tolerance. *Crop Science*, *41*(2), 493–509. https://doi.org/10.2135/cropsci2001.412493x

Specht, J. E., Hume, D. J., & Kumudini, S. V. (1999). Soybean Yield Potential—A Genetic and Physiological Perspective. *Crop Science*, *39*(6), 1560–1570. https://doi.org/10.2135/cropsci1999.3961560x

Sweeney, D. W., & Granade, G. V. (2002). Effect of a single irrigation at different reproductive growth stages on soybean planted in early and late June. *Irrigation Science*, *21*(2), 69–73. https://doi.org/10.1007/s00271-001-0051-x

Tanaka, Y., Fujii, K., & Shiraiwa, T. (2010). Variability of Leaf Morphology and Stomatal Conductance in Soybean [Glycine max (L.) Merr.] Cultivars. *Crop Science*, *50*(6), 2525–2532. https://doi.org/10.2135/cropsci2010.02.0058

Toker, C., Canci, H., & Yildirim, T. (2007). Evaluation of perennial wild Cicer species for drought resistance. *Genetic Resources and Crop Evolution*, *54*(8), 1781–1786. https://doi.org/10.1007/s10722-006-9197-y

USDA. (2022). *NASS QuickStats Ad-hoc Query Tool*. https://quickstats.nass.usda.gov/results/8D4EC4C4-286C-3654-828E-B497F1CC83EC

Valliyodan, B., Ye, H., Song, L., Murphy, M., Shannon, J. G., & Nguyen, H. T. (2017). Genetic diversity and genomic strategies for improving drought and waterlogging tolerance in

soybeans. *Journal of Experimental Botany*, *68*(8), 1835–1849. https://doi.org/10.1093/jxb/erw433

WESTGATE, M. E., & PETERSON, C. M. (1993). Flower and Pod Development in Water-Deficient Soybeans (Glycine max L. Merr.). *Journal of Experimental Botany*, *44*(258), 109–117.

Wilson, R. F. (2004). Seed Composition. In *Soybeans: Improvement, Production, and Uses* (pp. 621–677). John Wiley & Sons, Ltd. https://doi.org/10.2134/agronmonogr16.3ed.c13

Xiong, R., Liu, S., Considine, M. J., Siddique, K. H. M., Lam, H.-M., & Chen, Y. (2021). Root system architecture, physiological and transcriptional traits of soybean (Glycine max L.) in response to water deficit: A review. *Physiologia Plantarum*, *172*(2), 405–418. https://doi.org/10.1111/ppl.13201

Yan, C., Song, S., Wang, W., Wang, C., Li, H., Wang, F., Li, S., & Sun, X. (2020). Screening diverse soybean genotypes for drought tolerance by membership function value based on multiple traits and drought-tolerant coefficient of yield. *BMC Plant Biology*, *20*(1), 321. https://doi.org/10.1186/s12870-020-02519-9

Yang, M.-H., Jahufer, M. Z. Z., He, J., Dong, R., Hofmann, R., Siddique, K. H. M., & Li, F.-M. (2020). Effect of traditional soybean breeding on water use strategy in arid and semi-arid areas. *European Journal of Agronomy*, *120*, 126128. https://doi.org/10.1016/j.eja.2020.126128

Ye, H., Song, L., Schapaugh, W. T., Ali, M. L., Sinclair, T. R., Riar, M. K., Mutava, R. N., Li, Y., Vuong, T., Valliyodan, B., Pizolato Neto, A., Klepadlo, M., Song, Q., Shannon, J. G., Chen, P., & Nguyen, H. T. (2020). The importance of slow canopy wilting in drought tolerance in soybean. *Journal of Experimental Botany*, *71*(2), 642–652. https://doi.org/10.1093/jxb/erz150

Young, A. H., Knapp, K. R., Inamdar, A., Hankins, W., & Rossow, W. B. (2018). The International Satellite Cloud Climatology Project H-Series climate data record product. *Earth System Science Data*, *10*(1), 583–593. https://doi.org/10.5194/essd-10-583-2018

Zahran, H. H. (1999). Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiology and Molecular Biology Reviews: MMBR*, *63*(4), 968–989, table of contents. https://doi.org/10.1128/MMBR.63.4.968-989.1999

Table 3.1: Average soil matric potential (SMP), soil volumetric water content (VWC) and allowable depletion (AD) at the time of canopy wilting rating for different irrigation onsets in two environments 20STU and 20ROH (15-and 30-cm. depths)

Environment	Irrigation onset	SMP (mean ± se, cbar)	VWC (%)	AD (%)
20STU	R1	-128.2 ± 36.5	25.5	38
	R2	-104.6 ± 46.7	27.9	29
	R3	-199 ± 0	21	55
	R4	-199 ± 0	21	55
20ROH	R1	-55.2 ± 16.4	33.5	8
	R2	-90 ± 10.1	28.8	25
	R3	-108 ± 21.1	26.9	33
	R4	-124.5 ± 17.9	25.5	38

Table 3.2: Average percent of soil volume water content (VWC) and allowable depletion (AD) before irrigation from different irrigation onsets in two environments 20STU and 20ROH

Environment	Irrigation onset	VWC before irrigation (%)	AD before irrigation (%)
20STU	R1	11.1	92
	R2	12	88
	R3	15.7	75
	R4	19.5	60
20ROH	R1	11.4	91
	R2	18.5	64
	R3	16.7	71
	R4	12.4	87

Table 3.3. Soybean seed protein, oil, and 100-seed weight of each wilting class under different onset irrigations (R1, R2, R3, and R4) evaluated in four environments (location-year combination 19ROH, 19STU, 20ROH, and 20STU)

Wilting Class	Irrigation Onset	Protein (%)	Oil (%)	100-seed weight (g)
Fast wilting	R1	39.33	21.39	14.99
Slow wilting	R1	39.34	21.40	15.07
Fast wilting	R2	40.08	21.28	15.16
Slow wilting	R2	39.94	21.44	15.16
Fast wilting	R3	39.85	21.48	15.23
Slow wilting	R3	39.57	21.60	15.25
Fast wilting	R4	39.62	21.52	15.31
Slow wilting	R4	39.36	21.52	15.43

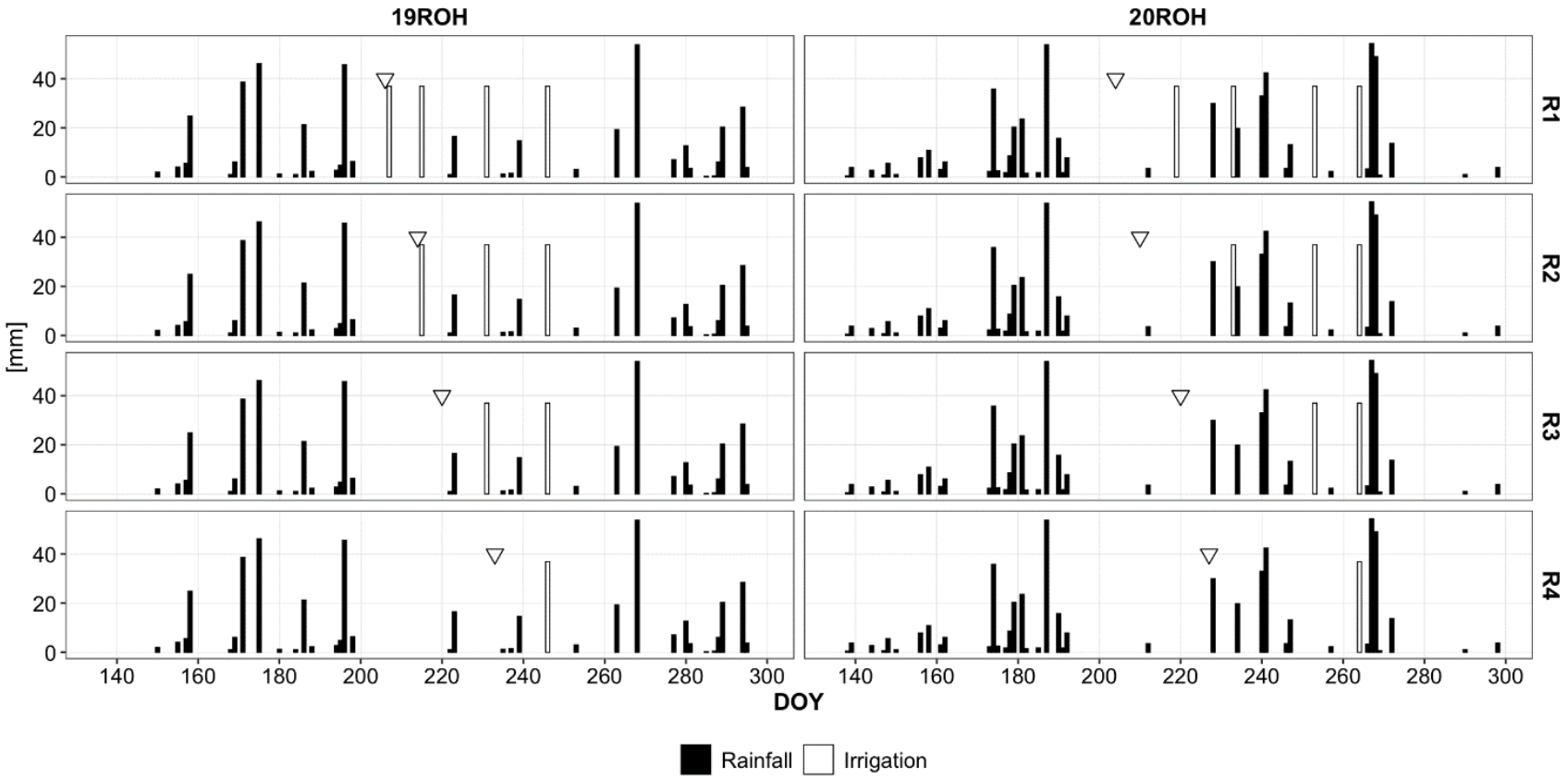

Figure 3.1: Daily distribution of rainfall and irrigation (mm) during the two growing seasons (2019 and 2020) in Rohwer (ROH).

DOY: day of the year. The ▽ symbol indicates the time of canopy wilting rating.

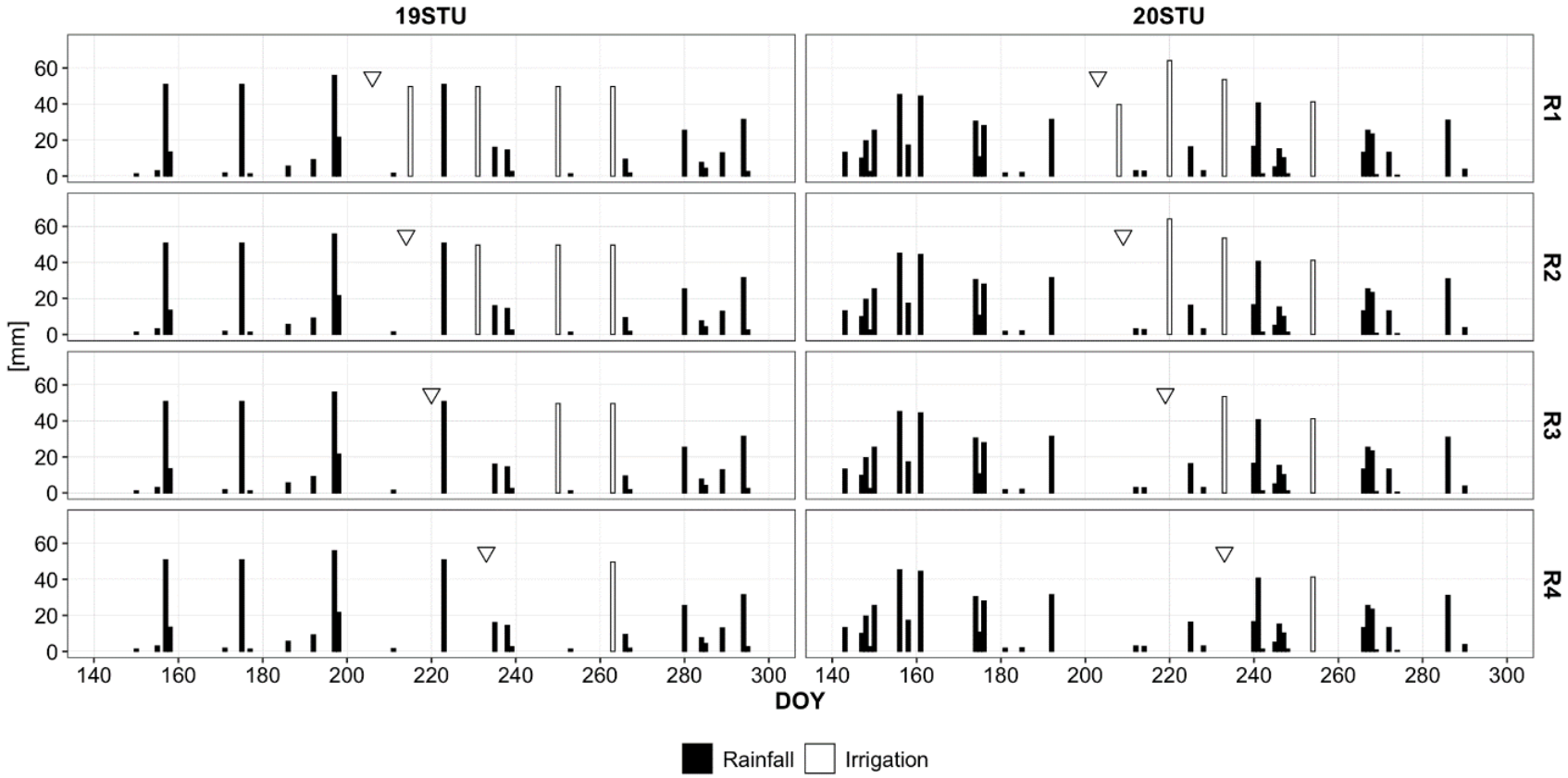

Figure 3.2: Daily distribution of rainfall and irrigation (mm) during the two growing seasons (2019 and 2020) in Stuttgart (STU).

DOY: day of the year. The ▽ symbol indicates the time of canopy wilting rating.

Figure 3.3. Canopy witling of each wilting class under different onset irrigations (R1, R2, R3, and R4). Same letters are not significantly different (Tukey's; $p < 0.05$). Whiskers denote standard error of the mean.

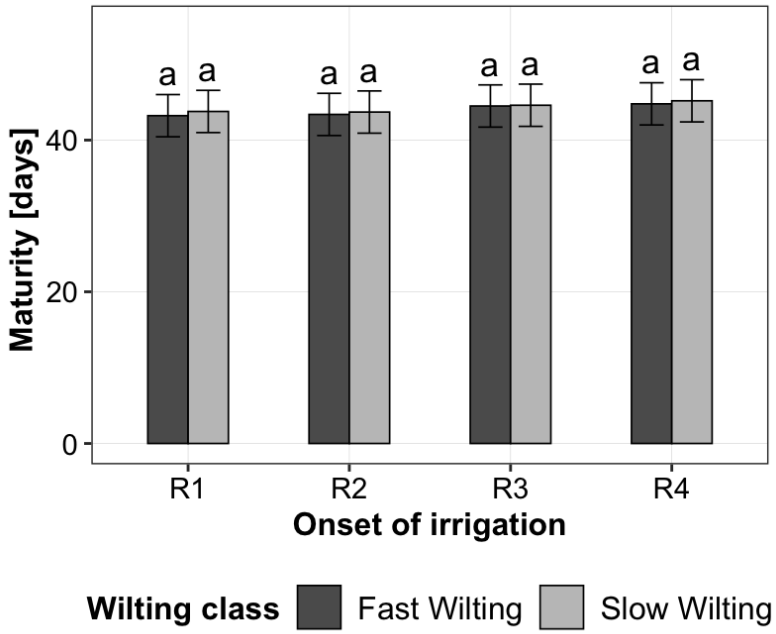

Figure 3.4. Maturity of each wilting class under different onset irrigations (R1, R2, R3, and R4). Same letters are not significantly different (Tukey's; $p < 0.05$). Whiskers denote standard error of the mean.

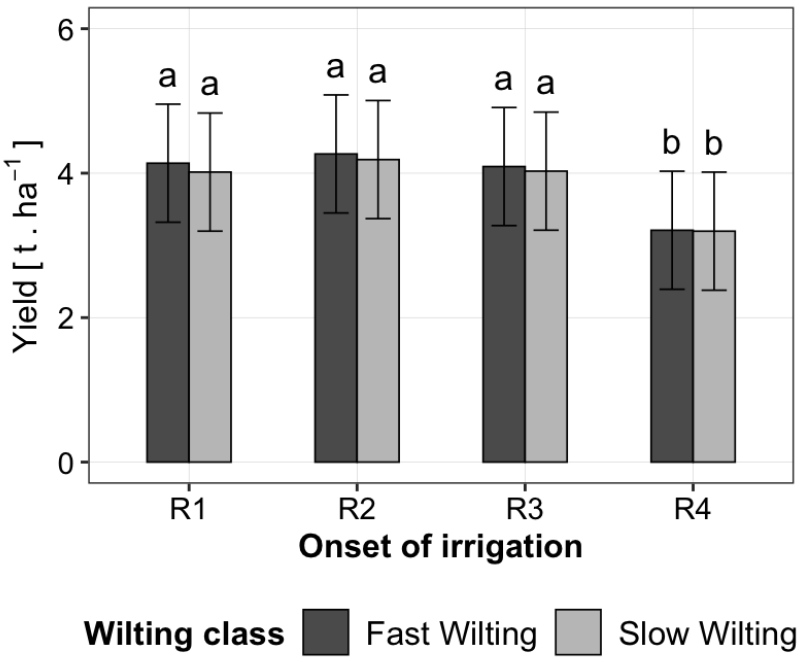

Figure 3.5. Soybean seed yield of each wilting class under different onset irrigations (R1, R2, R3, and R4). Same letters are not significantly different (Tukey's; $p < 0.05$). Whiskers denote standard error of the mean.

Table S 3.1. Average minimum (Min_T), maximum (Max_T), average (Avg_T) temperature in °C monthly from planting to harvest evaluated in four environments (location-year combination 19ROH, 19STU, 20ROH, and 20STU)

	19ROH			19STU			20ROH			20STU		
	Min_T (°C)	Max_T (°C)	Avg_T (°C)	Min_T (°C)	Max_T (°C)	Avg_T (°C)	Min_T (°C)	Max_T (°C)	Avg_T (°C)	Min_T (°C)	Max_T (°C)	Avg_T (°C)
May	20	30	25	19	32	26	18	28	23	19	28	23
June	20	30	25	21	30	25	21	30	25	21	31	26
July	22	31	27	22	32	27	24	32	28	24	33	28
Aug	23	32	27	23	32	28	21	30	26	21	32	27
Sept	22	34	28	22	34	28	19	28	24	19	29	24
Oct	13	24	19	12	24	18	10	21	16	12	23	17

Table S 3.2. Soil type in each environment based on the percentage of sand, silt and clay

Field	Sand (%)	Silt (%)	Clay (%)	Soil Type
19ROH	7.1	78.5	14.4	Silt loam
19STU	1.05	83.8	15.15	Silt loam
20ROH	12.99	69.71	17.3	Silt loam
20STU	2.29	76.09	21.62	Silt loam

Table S 3.3. Characteristic of elements of soils from soil samplings for pre and post samplings for different irrigation onsets (R1, R2, R3, and R4) in four environments (19ROH, 19STU, 20ROH, and 20STU), paired t- test, mean difference, t ratio and *p*-value

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
P ($mg.kg^{-1}$)	19ROH	R1	46.92	30.83	-16.10	-1.979	0.1421
		R2	71.93	34.31	-37.63	-5.983	0.0093
		R3	46.31	41.55	-4.76	-0.445	0.6862
		R4	37.34	22.30	-15.04	-1.299	0.2848
	19STU	R1	45.75	25.89	-19.86	-4.734	0.0179
		R2	36.99	30.24	-6.75	-2.501	0.0876
		R3	40.12	29.18	-10.94	-2.381	0.0975
		R4	40.96	32.06	-8.90	-1.732	0.1817
	20STU	R1	28.39	19.56	-8.83	-2.226	0.1123
		R2	29.01	20.01	-9.01	-15.588	0.0006
		R3	37.10	24.44	-12.66	-14.433	0.0007
		R4	23.98	23.05	-0.92	-0.433	0.6942
	20ROH	R1	49.24	39.30	-9.94	-8.66	0.0032
		R2	68.28	61.00	-7.28	-2.02	0.1366
		R3	67.73	68.40	0.67	0.216	0.8425
		R4	66.23	71.02	4.79	5.196	0.0138

Table S 3.4. (Cont.)

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
K ($mg.kg^{-1}$)	19ROH	R1	161.96	115.94	-46.02	-2.042	0.1337
		R2	187.05	119.71	-67.34	-46.765	<0.0001
		R3	156.56	119.97	-36.59	-2.69	0.0744
		R4	149.42	164.30	14.88	4.566	0.0197
	19STU	R1	157.88	104.58	-53.31	-9.754	0.0023
		R2	161.95	128.30	-33.65	-4.641	0.0188
		R3	176.12	125.98	-50.14	-21.65	0.0002
		R4	167.36	129.10	-38.25	-4.939	0.0159
	20STU	R1	168.82	102.77	-66.05	-22.863	0.0002
		R2	163.43	107.53	-55.90	-12.124	0.0012
		R3	160.45	114.85	-45.60	-2.57	0.0825
		R4	176.57	111.95	-64.62	-20.312	0.0003
	20ROH	R1	173.93	112.38	-61.55	-10.738	0.0017
		R2	173.53	141.23	-32.30	-3.411	0.0421
		R3	173.73	140.03	-33.70	-6.826	0.0064
		R4	136.48	145.91	9.43	0.645	0.5647

Table S 3.5. (Cont.)

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
Ca (mg.kg^{-1})	19ROH	R1	1056.78	978.13	-78.65	-1.378	0.2618
		R2	1205.34	1034.90	-170.44	-2.534	0.0851
		R3	1060.30	1166.54	106.24	4.145	0.0255
		R4	1379.12	1813.85	434.73	3.965	0.0287
	19STU	R1	1122.19	1193.60	71.42	1.45	0.2429
		R2	1004.78	1060.78	55.99	1.732	0.1817
		R3	1023.49	965.52	-57.97	-2.79	0.0684
		R4	1193.12	1082.05	-111.07	-5.059	0.0149
	20STU	R1	1169.34	1248.88	79.53	1.385	0.259
		R2	1073.21	969.45	-103.75	-1.552	0.2184
		R3	1122.21	1068.33	-53.88	-1.6126	0.2052
		R4	1200.12	1228.29	28.18	3.73	0.0336
	20ROH	R1	1336.47	1332.06	-4.41	-0.065	0.9521
		R2	1567.39	1513.60	-53.79	-2.353	0.1
		R3	1768.38	1583.96	-184.43	-11.62	0.0014
		R4	1522.84	1557.73	34.89	2.02	0.1366

Table S 3.6. (Cont.)

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
Mg ($mg.kg^{-1}$)	19ROH	R1	186.33	192.37	6.04	-0.115	0.9154
		R2	146.40	202.02	55.62	-2.209	0.1141
		R3	168.15	170.90	2.75	-0.346	0.7519
		R4	445.51	256.78	-188.73	4.091	0.0264
	19STU	R1	105.85	100.34	-5.51	-2.886	0.0632
		R2	97.08	94.61	-2.47	-8.66	0.0032
		R3	95.36	89.12	-6.23	-7.505	0.0049
		R4	123.38	98.57	-24.81	-7.715	0.0045
	20STU	R1	118.92	111.45	-7.48	-6.062	0.009
		R2	106.15	93.39	-12.76	-2.501	0.0876
		R3	110.86	93.77	-17.09	-6.735	0.0067
		R4	111.09	104.06	-7.03	-4.041	0.0273
	20ROH	R1	219.39	208.55	-10.84	-0.866	0.4502
		R2	257.27	242.35	-14.91	-7.175	0.0056
		R3	277.29	251.03	-26.26	-3.399	0.0425
		R4	238.58	246.00	7.42	1.367	0.2649

Table S 3.7. (Cont.)

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
S ($mg.kg^{-1}$)	19ROH	R1	8.05	9.36	1.30	3.752	0.0331
		R2	9.25	8.51	-0.74	-0.44	0.6895
		R3	9.32	6.77	-2.55	-2.676	0.0753
		R4	7.15	9.05	1.90	4.701	0.0182
	19STU	R1	10.89	6.15	-4.74	-7.154	0.0056
		R2	10.75	6.88	-3.88	-14.818	0.0007
		R3	11.10	8.37	-2.73	-8.66	0.0032
		R4	10.24	7.90	-2.34	-9.045	0.0029
	20STU	R1	10.48	6.21	-4.27	-74.478	<0.0001
		R2	10.35	8.50	-1.86	-1.307	0.2821
		R3	11.53	7.19	-4.34	-4.86	0.0166
		R4	9.86	7.97	-1.89	-1.218	0.31
	20ROH	R1	14.36	8.70	-5.67	-1.829	0.1648
		R2	15.35	9.40	-5.95	-5.027	0.0152
		R3	14.52	8.70	-5.82	-202.65	<0.0001
		R4	18.18	8.74	-9.44	-4.251	0.0239

Table S 3.8. (Cont.)

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
Na (mg.kg^{-1})	19ROH	R1	20.87	35.29	14.42	2.027	0.1356
		R2	19.08	20.46	1.37	0.21	0.8465
		R3	22.74	23.00	0.25	0.509	0.6456
		R4	30.14	66.14	36.01	2.808	0.0674
	19STU	R1	37.74	32.56	-5.18	-1.957	0.1451
		R2	31.96	31.41	-0.55	-0.141	0.8967
		R3	31.74	28.57	-3.17	-3.464	0.0405
		R4	41.01	29.69	-11.32	-8.896	0.003
	20STU	R1	36.14	33.34	-2.81	-2.407	0.0952
		R2	28.50	20.12	-8.38	-6.427	0.0076
		R3	29.13	20.17	-8.96	-11.991	0.0012
		R4	27.91	20.54	-7.37	-4.992	0.0155
	20ROH	R1	28.79	23.42	-5.36	-0.677	0.5465
		R2	27.81	23.35	-4.46	-15.588	0.0006
		R3	30.62	19.49	-11.13	-96.128	<0.0001
		R4	34.16	18.25	-15.91	-2.673	0.0754

Table S 3.9. (Cont.)

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
Fe(mg.kg^{-1})	19ROH	R1	316.15	251.84	-64.31	-2.771	0.0695
		R2	293.57	270.64	-22.93	-1.285	0.289
		R3	310.25	237.47	-72.77	-2.536	0.0849
		R4	242.63	180.58	-62.05	-2.825	0.0664
	19STU	R1	592.73	461.22	-131.51	-13.347	0.0009
		R2	711.37	541.74	-169.62	-14.321	0.0007
		R3	640.83	582.04	-58.79	-2.761	0.07
		R4	619.34	569.66	-49.68	-1.884	0.156
	20STU	R1	620.82	560.81	-60.01	-6.113	0.0088
		R2	620.80	626.74	5.93	0.266	0.8071
		R3	557.64	517.83	-39.81	-20.042	0.0003
		R4	475.73	403.05	-72.69	-4.214	0.0244
	20ROH	R1	304.44	259.86	-44.58	-38.105	<0.0001
		R2	349.78	321.58	-28.20	-1.102	0.3509
		R3	327.94	334.31	6.37	0.776	0.4941
		R4	376.17	355.38	-20.78	-1.651	0.1972

Table S 3.10. (Cont.)

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
Mn (mg.kg^{-1})	19ROH	R1	90.98	56.02	-34.95	-2.331	0.102
		R2	135.65	59.70	-75.96	-8.227	0.0038
		R3	81.14	77.83	-3.31	-0.023	0.8322
		R4	95.78	72.14	-23.64	-1.979	0.1421
	19STU	R1	101.44	47.64	-53.81	-93.53	<0.0001
		R2	80.69	41.65	-39.04	-19.547	0.0003
		R3	87.78	31.78	-56.00	-96.994	<0.0001
		R4	96.94	38.24	-58.71	-34.063	<0.0001
	20STU	R1	135.06	63.00	-72.06	-24.9415	0.0001
		R2	129.75	65.42	-64.33	-110.851	<0.0001
		R3	160.87	69.53	-91.34	-105.655	<0.0001
		R4	212.67	129.47	-83.20	-9.584	0.0024
	20ROH	R1	142.83	116.57	-26.26	-9.006	0.0029
		R2	136.90	103.76	-33.14	-6.35	0.0079
		R3	172.14	119.48	-52.66	-12.124	0.0012
		R4	151.41	124.67	-26.73	-1.798	0.169

Table S 3.11. Cont.

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
Zn (mg.kg^{-1})	19ROH	R1	2.67	2.37	-0.30	-1.299	0.2848
		R2	4.17	2.48	-1.69	-4.663	0.0186
		R3	2.85	2.51	-0.34	-0.932	0.419
		R4	2.63	2.29	-0.35	-1.347	0.2707
	19STU	R1	0.93	0.57	-0.37	-1.732	0.1817
		R2	0.81	0.62	-0.20	-3.464	0.0405
		R3	0.88	0.56	-0.31	.	.
		R4	0.72	0.60	-0.12	-5.196	0.0138
	20STU	R1	0.96	0.75	-0.22	-8.66	0.0032
		R2	0.90	0.73	-0.17	.	.
		R3	0.84	0.65	-0.19	-5.196	0.0138
		R4	0.95	0.76	-0.19	-0.866	0.4502
	20ROH	R1	5.05	4.26	-0.79	-8.66	0.0032
		R2	5.39	4.97	-0.42	-2.309	0.1041
		R3	5.00	4.58	-0.42	-15.588	0.0006
		R4	4.73	4.44	-0.29	-8.66	0.0032

Table S 3.12. (Cont.)

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
Cu (mg.kg^{-1})	19ROH	R1	1.04	0.91	-0.12	-5.196	0.0138
		R2	1.28	0.96	-0.32	-12.124	0.0012
		R3	0.85	1.14	0.29	4.041	0.0273
		R4	1.19	1.60	0.40	.	.
	19STU	R1	0.77	0.61	-0.16	3x10-5	<0.0001
		R2	0.58	0.60	0.03	.	.
		R3	0.65	0.57	-0.08	.	.
		R4	0.67	0.60	-0.07	-1.732	0.1817
	20STU	R1	0.95	0.73	-0.23	-3.464	0.0405
		R2	0.89	0.74	-0.15	-0.866	0.4502
		R3	1.12	0.76	-0.36	-12.124	0.0012
		R4	1.35	1.17	-0.18	-5.196	0.0138
	20ROH	R1	1.66	1.54	-0.12	-1.732	0.1817
		R2	1.77	1.78	0.02	.	.
		R3	1.86	1.68	-0.18	-3.464	0.0405
		R4	1.62	1.56	-0.06	-0.346	0.7519

Table S 3.13. (Cont.)

Elements	Env	Irrigation level	Pre-Sampling	Post-Sampling	Mean Difference (Post-Pre)	t ratio	*p*-value
B ($mg.kg^{-1}$)	19ROH	R1	0.41	0.70	0.29	.	.
		R2	0.53	0.96	0.43	1.732	0.1817
		R3	0.44	0.95	0.50	8.66	0.0032
		R4	0.48	1.10	0.62	5.196	0.0138
	19STU	R1	0.37	0.51	0.14	1.732	0.1817
		R2	0.41	0.43	0.02	1.732	0.1817
		R3	0.40	0.98	0.58	3.81	0.0318
		R4	0.37	0.81	0.44	3.117	0.0526
	20STU	R1	1.16	1.48	0.32	12.124	0.0012
		R2	1.33	1.32	-0.02	-0.577	0.6042
		R3	1.45	1.33	-0.12	-1.732	0.1817
		R4	1.27	1.25	-0.02	.	.
	20ROH	R1	1.37	1.59	0.21	.	.
		R2	1.52	1.59	0.07	1.732	0.1817
		R3	1.55	1.60	0.04	1.732	0.1817
		R4	1.62	1.58	-0.04	.	.

Figure S 3.1. Picture of R11-2933 (slow witling parent) and R11-1057 (fast wilting parent) at R4 stage

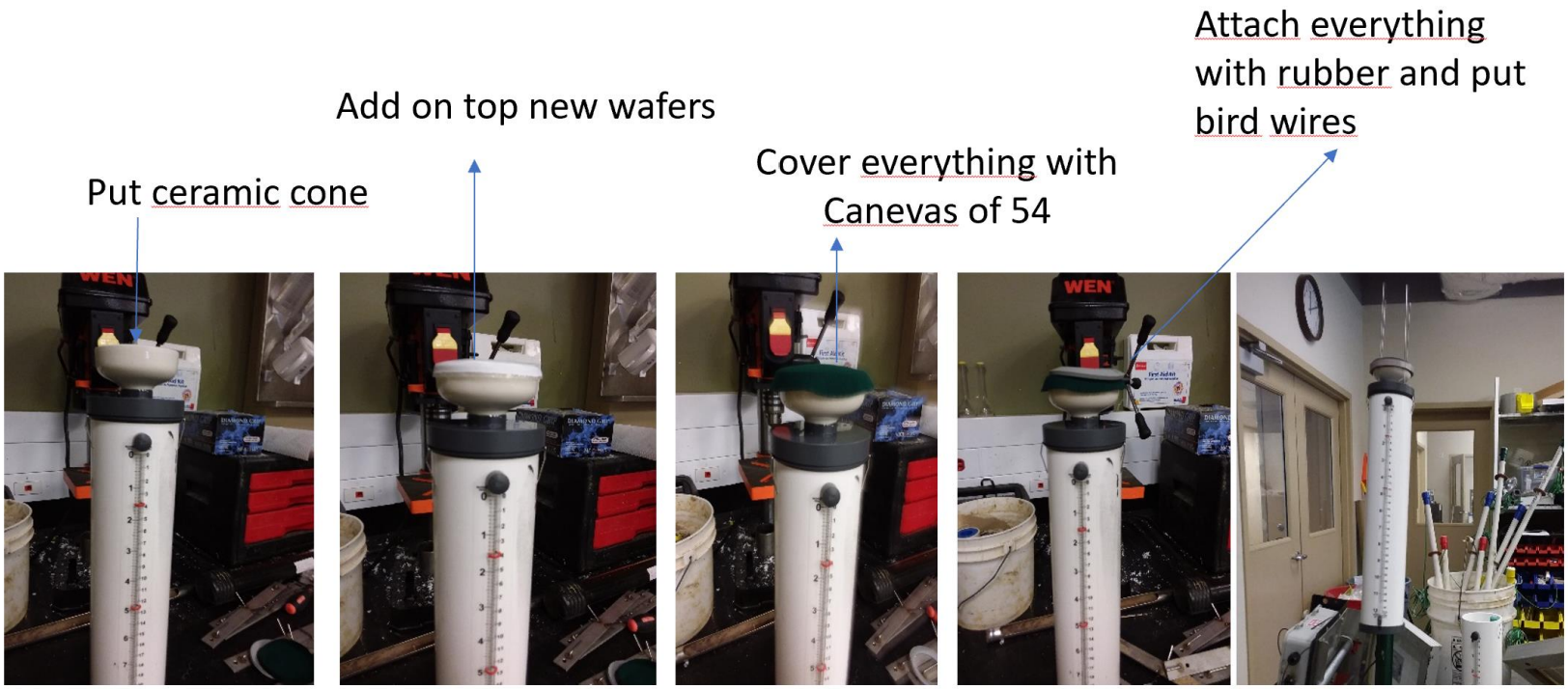

Figure S 3.2: Atmometer setting up and installation

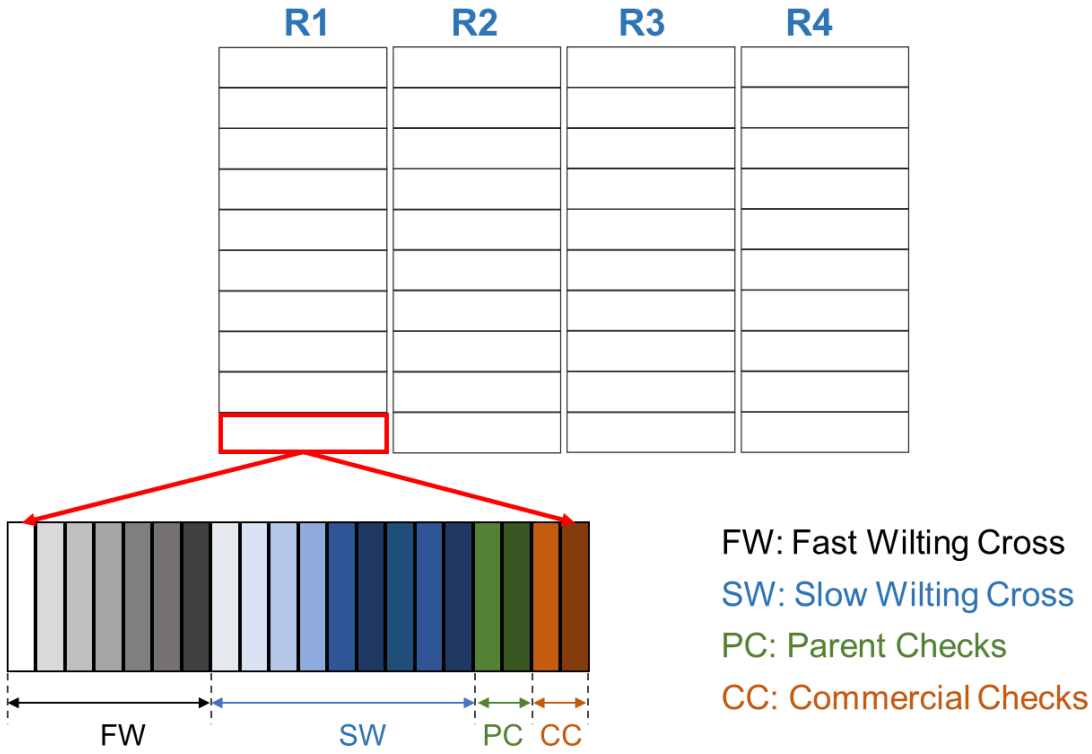

Figure S 3.3. The layout of the experimental design. Each strip (R1, R2, R3, and R4) represents a different irrigation onset composed by 10 blocks. Each block was composed of 2 PC, 2 CC, 7 FW, and 9 SW that were randomized. The order of the picture is just for the presentation, but lines were randomized.

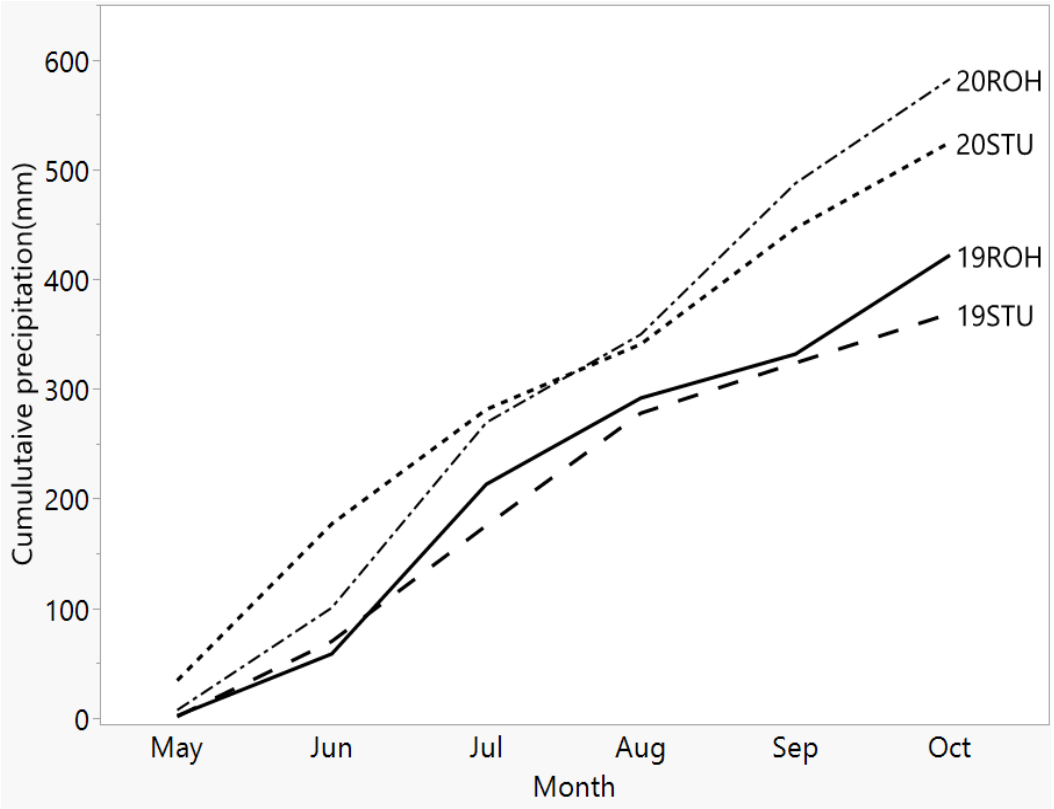

Figure S 3.4: Cumulative precipitation in mm from emergence to harvest time (May 30 to October 31) from each environment 19ROH, 19STU, 20ROH, and 20STU

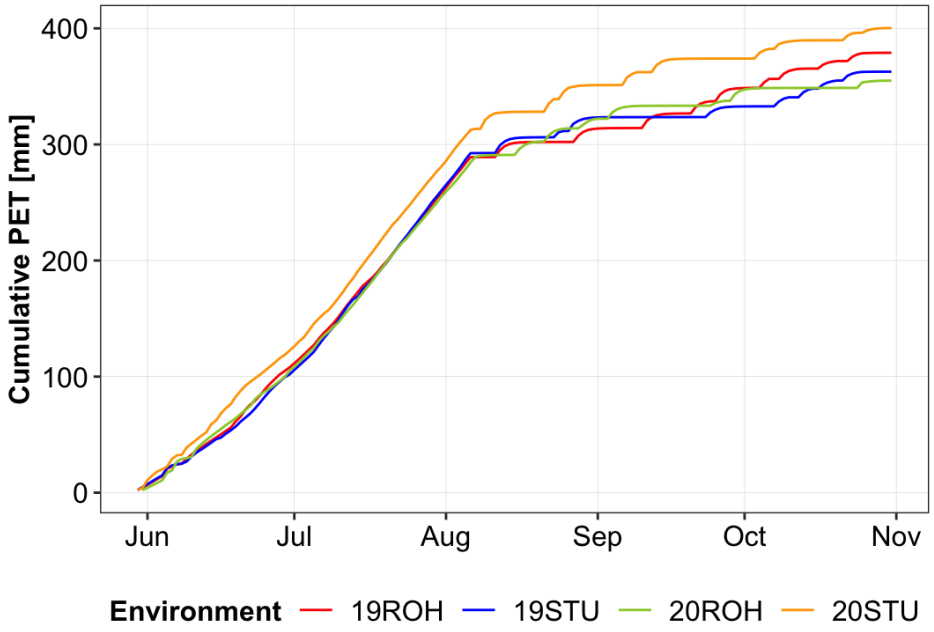

Figure S 3.5: Cumulative potential evapotranspiration in mm from emergence to harvest time. from each of the environments evaluated. The calculation in May was for 2 days (May 30 and May 31)

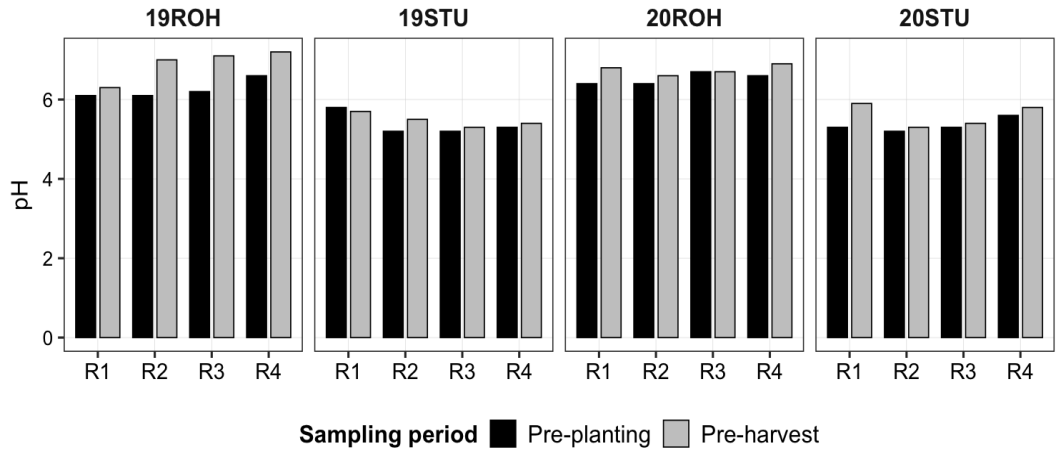

Figure S 3.6. Soil pH evaluation of different irrigation onsets (R1, R2, R3, and R4) in four different environments at either pre-planting or pre-harvest.

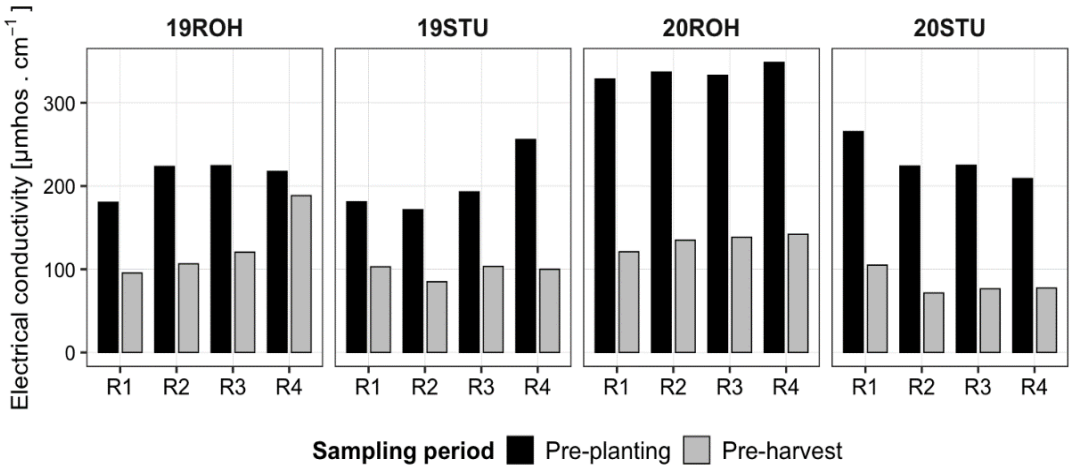

***Figure S* 3.7.** Soil Electrical conductivity evaluation of different irrigation onsets (R1, R2, R3, and R4) in four different environments at either pre-planting or pre-harvest.

CHAPTER IV

Nested association mapping for wilting, maturity, seed yield and seed yield genomic selection under reduced irrigation in two RILs soybean populations

ABSTRACT

Soybean [*Glycine max* (L.) Merr.] is a diploid legume species. It provides affordable source of protein to animals and humans. However, exploring molecular approaches to increase yield genetic gain has been one of the main challenges for soybean breeders and geneticists due to climate change. Therefore, the objectives of this study were to conduct a nested association mapping (NAM) for wilting, maturity, and seed yield and to identify superior individuals in seed yield using genomic approach under different irrigation onsets. A total of 167 genotypes along with commercial checks were evaluated under four different irrigation onsets. Soybean was genotyped using the Infinium Soy6KSNP Beadchips. The results indicated that: 1) a total of 4, 39, and 7 SNPs were found to be significantly associated with canopy wilting, maturity, and seed yield, respectively, using the combined data under different irrigation onsets obtained over four environments (location-year combination) 2) overall genomic selection accuracy was moderate ranging from 0.39 to 0.44, and genomic selection was efficient to select superior soybean lines under reduced irrigation.

INTRODUCTION

Climate change is a major threat to food insecurity in the 21st century (Gowdy, 2020). The food and feed sectors are dependent on nutrient-rich crops, such as soybean [*Glycine max* (L.) Merr.]. Soybean a diploid legume (2n= 2x= 20), is one of the most important legumes worldwide by providing oil and being a source of vegetable protein. It provides about 60% of the vegetable-derived proteins and more than 57% of oilseed (USDA FAS, 2018). Developing soybean-derived biofuel has been recently increasing, with an estimated value exceeding $30 billion in the United States (U.S.) (SoyStats, 2022).

As soybean cultivation is usually rain-dependent, water is usually the main factor for soybean productivity (Ohashi et al., 2006; Sinclair et al., 2014; Anda et al., 2020; Mekonnen et al., 2020; Xiong et al., 2021). In the U.S. Mid-South, soybean production depends heavily on irrigation (Reba & Massey, 2020), with 85% of soybean acres currently supplemented with water in Arkansas (AFBF, 2022). Water shortage during soybean developmental and growth stages could be detrimental to soybean production. Evidence of the negative effects of drought stress on soybean has been reported (Specht et al., 1999; Baghel et al., 2018; Dogan et al., 2007). Developing soybean varieties that meet the needs of end users is critical.

Therefore, soybean breeding programs aiming at improving drought tolerance is still required. Breeding for drought tolerance can make use of good understanding of the genetic control for drought tolerance. According to Clement et al. (2008), as part of the *Fabaceae* family, soybean is one of the most drought-sensitive legumes. Canopy wilting is the first visible symptom of water stress, and a number of genotypes have been identified as slow wilting under field conditions (Carter et al.; 1999, 2006).

Several QTL mappings were conducted for canopy witling to detect genomic regions associated with that trait. Research conducted by Charlson et al. (2009) found seven regions on chromosomes Gm08, Gm13, Gm14, and Gm17 using 93 Recombinants Inbred Lines (RILs). Genome-wide association studies (GWAS) are a recent tool developed to identify quantitative trait loci for various traits. GWAS studies on soybean slow wilting were conducted by Kaler et al. (2017), reporting 21 single nucleotide polymorphisms (SNPs) associated with the trait located within a gene or very close to genes that had a reported biological connection to transpiration or water transport.

The use of molecular markers through marker-assisted selection (MAS) in soybean breeding program has been thriving. Tools such as quantitative trait loci (QTL), genome-wide association mapping (GWAS), nested association mapping (NAM) have increasingly become popular in efforts towards uncovering the genetic basis of traits of interest in agriculture and identifying new markers. Genome-wide association studies (GWAS) have more recently been extensively used to compensate for drawbacks of the conventional QTL mapping approach. In comparison, GWAS is able to survey numerous historical recombination events in collections of landraces, varieties and breeding lines (Flint-Garcia et al., 2003), and achieve a higher resolution of QTL mapping than solely on biparental segregating populations. GWAS has been particularly successful in interpreting the associations between molecular markers and traits (Si et al., 2016; Yano et al., 2016), although the filter of rare variants and inherent population structure in natural populations tends to reduce statistical power (Xiao et al., 2017). Rare variants may be the cause of the phenotypic variant of interest and thus a source of the missing heritability (Eichler et al., 2010). Moreover, it is hard to detect variants underlying traits of interest correctly if they are significantly correlated to population structure (Flint-Garcia et al., 2005).

To overcome spurious associations and increase the detection power of rare alleles in crop species, multi-parent cross populations using different cross designs have been developed, such as Nested Association Mapping (NAM), Multi-parent Advanced Generation Inter-Crosses (MAGIC), and Random-open-parent Association Mapping (ROAM) (Yu & Buckler, 2006; Dell'Acqua et al., 2015; Xiao et al., 2016). The NAM population design was first proposed in maize (Yu and Buckler, 2006; Yu et al., 2008), and the first example in maize consisted of 5,000 RILs derived from 25 segregating families generated from crossing the homozygous B73 line with 25 lines representing a wide coverage of the domesticated maize genepool (McMullen et al., 2009).

Quantitative traits have proven difficult to select for using MAS based on the fact that they are polygenic and loci responsible for variation in these traits often have small effects. Meuwissen et al. (2001) introduced the concept of genomic selection (GS) to take advantage of genotypic data to predict the performance of genotypes for complex traits. The main difference between MAS and GS, is that GS utilizes all markers across the genome to predict the performance of traits of interest, while MAS relies on a few markers to select specific QTL often associated with qualitative traits. Moreover, Heffner et al. (2011) reported that GS provided threefold and twofold genetic gain per year compared to MAS for maize and winter wheat respectively, when costs were equivalent. With the advent of new genotyping platforms, such as single nucleotide polymorphism (SNP) beadchip arrays, Diversity array Technology (DArT), and genotyping-by sequencing (GBS), high-throughput genotyping has made GS more affordable and efficient (Rasheed et al., 2017). Genomic selection has been frequently used to achieve faster genetic gain in plant breeding. There have been several studies examining the potential for GS in soybean but relatively few compared to maize and wheat.

The research objectives of this study were to conduct a nested association mapping study to identify QTL associated with wilting, maturity, and seed yield in soybean under reduced irrigation and to carry out a genomic selection study to select superior high yielding lines that are potentially drought tolerant.

MATERIALS AND METHODS

Plant materials and experimental design

A Nested Association Mapping (NAM) population composed of two populations of recombinant inbred lines (RILs) was constructed from the crosses of N07-14753/R11-1057 (Pop1) and R11-2933/R11-1057 (Pop2). The common parent was R11-1057, a high-yielding line with maturity group 5 (MG 5); while the other parent has MG V: N07-14753 (high yielding line) and R11-2933 (drought tolerant: slow wilting line). As the R11-1057 was the common parent of the two populations, the NAM was also a half-sib population. A total of 165 $F_{4:7}$ breeding lines (73 from Pop1 and 92 from Pop2) along with two parents checks (PC) R11-2933 and R11-1057 and two commercial checks (CC) based on seed availability (AG55X7, AG56X8, P53AG7X, P55A49X) were grown in four environments (location year-combination) using an augmented strip-plot design under four furrow-irrigation onsets. Environments included Stuttgart, AR (silt loam) in 2019 and 2020 (19STU and 20STU); and Rowher (silt loam) in 2019 and 2020 (19ROH and 20ROH), AR. The four irrigation onsets were: 1) full irrigation (irrigation initiated at initiation of flowering (noted as R1; Fehr & Caviness, 1977), 2) irrigation initiated at full flowering (noted as R2), 3) irrigation initiated at beginning of pod development (noted as R3), and 4) irrigation initiated when pods were 2 cm at one of the four uppermost nodes (noted as R4). The irrigation at each designated growth stage was triggered using the decision table developed by Henry et al. (2014) for atmometer measurements based on 50% of the plots

reaching the desired stage. Each strip of irrigation onset (R1, R2, R3, and R4) was composed of ten blocks. One block was composed of four checks (two PC and two CC) and 16 randomly-assigned genotypes, including seven Pop1 and nine Pop2 genotypes where individual lines were a random factor within populations. The plots consisted of two rows 0.76 m apart, 4.6 m long with 1.5 m alley. Standard agronomic practices were used at each location, including fertilization to recommended levels as defined by Slaton et al. (2013).

Phenotyping

To reduce impact of diurnal variation in evaporative demand, rating of canopy wilting was conducted between 11: 00 am and 3:00 pm. Plots were rated prior to the onset time of triggering irrigation using a score of 0 (no wilting) to 9 (plant death). When 95 % of the plots reached mature pod color (Fehr & Caviness, 1977), maturity was recorded and expressed as the number of days after 31st August. Seed yield (kg/ha) was calculated from plot moisture, weight, and dimensions.

Genotyping and quality control

DNA was extracted from young leaves of each lines using the hexadecymethylammonium bromide CTAB protocol (Doyle, 1990). Soybean lines were genotyped using the Infinium Soy6KSNP Beadchips (Song et al., 2020) in the Soybean Improvement Laboratory USDA-ARS, Beltsville. Of the 6,000 single nucleotide polymorphism (SNPs), a total of 3,733 were maintained after SNPs filter (missing data <20 %, heterozygosity <10%, minor allele frequency>5%). Those SNPs were used for further analysis in the nested association (NAM) and the genomic selection (GS) analysis.

Best linear unbiased prediction (BLUP)

Data were run in R using the lme4 package (Bates et al., 2015). Each irrigation onset (R1, R2, R3, and R4) was separately analyzed using the following statistical model:

$$Y_{ijk} = \mu + G_i + E_j + G_i + B_{k(j)} + GB_{ik(j)} + e_{ijk}$$

where *i* was the number of genotypes classes (1…169), *j* number of the environments 19ROH, 19STU, 20ROH, and 20STU (1,2,3,4), and *k:* number of the blocks (1…10)

where y_{ijk} was the mean response of the *ijk*-th observation (canopy wilting, maturity, seed yield), μ the overall mean response; G_j the genotype, E_j the environment, $B_{k(j)}$ the block nested within the environment, $GB_{ik(j)}$ was the interaction between the *i*-th genotype and the *kj*-th block, and e_{ijk} is the experimental error. All factors were put as random factors to generate the best linear unbiased prediction (BLUP) to account for variation resulting from environmental factors. NAM and GS analysis were performed using the BLUP values.

Nested association mapping and candidate gene discovery

Nested association mapping was conducted using the R package NAM (Xavier et al., 2015) The mixed linear model designed for multiple parent intercross populations was used for the SNP and haplotype-based association (Wei &Xu, 2016):

$$Y = \mu + X\alpha + g + e$$

where Y the observed trait value (wilting, maturity, seed yield), μ was the intercept, X was the allele matrix from SNP/haplotype data and family information, α was the SNP/haplotype effects, g was the population structure effect, and e was the residual effect. A logarithm of the

odds (LOD) threshold at 3 was used to declare SNP significant in the nested association mapping.

Candidate gene(s) discovery

Significant SNPs were used for candidate gene(s) discovery. The 40-kb region harboring the significant SNP was considered for candidate gene search using Soybase (https://www.soybase.org/) based on the SNP density. Functional annotation pertaining to candidate gene(s) was investigated using Soybase database as well.

Genomic estimated breeding values (GEBVS)

Genomic selection was carried out using 3,733 SNPs. Genomic estimated breeding values (GEBVs) were computed under ridge regression best linear unbiased predictor (rrBLUP) (Meuwissen et al. 2001). The rrBLUP model was:

$$Y = WG\beta + \varepsilon$$

where Y was the vector phenotype (BLUP yield), W corresponded to the incidence matrix relating the genotype to the phenotype, G denoted the genetic matrix, β indicated the marker effect with $\beta \sim N(0, I\sigma^2_\beta)$, and ε was the random error. The package "rrBLUP" was used in R v.3.6.1 to perform the genomic selection model (Endelman, 2011). GEBVs were estimated using a random training population chosen from the NAM population. Since the commercial checks were not used of the selection, they were removed prior to the genomic selection, leaving a total of 165 soybean breeding lines for the analysis. We performed five-fold cross-validation corresponding to a training population/testing population size of 132/33. The training population was used to fit the model and the testing population was used to assess the accuracy of the model. A total of 100 replications was conducted at each level of cross-validation. Genomic

selection accuracy corresponded to the Pearson's correlation coefficient between the GEBVs and the observed phenotypic values in the testing set (Shikha et al., 2017). The Spearman correlation run in JMP was done to assess GEBVs ranking of each line in each irrigation onset. And a principal component analysis (PCA)was performed in JMP related to each GEBVs in each irrigation onset.

Lines advancement and selection based on Yield GEBVs

A total of 25 lines were selected based on the yield GEBVs. The selection was composed of 12 lines that were from the top 25% across irrigation onsets, top five lines just at irrigation onset at R1, top five lines just at irrigation onset at R4, and three lines than performed the worst (the lowest GEBVs) at irrigation onset at R1 and R4, and one line that from the cross of Pop1. The 25 selected lines previously mentioned, along with four commercial checks (AG48X9, S49-F5X, AG52XF0, AG53X0, and AG56X8) were grown into two experiments during the growing season of 2021: fully irrigated and non-irrigated in Stuttgart, AR. The experimental design was a randomized complete block design (RCBD) with three replications. The plots consisted of two rows 0.76 m apart, 4.6 m long with 1.5 m alley. Yield (kg/ha) was collected at the end of season. The analysis of variance (ANOVA) was done using PROC MXED in SAS (SAS Institute, Cary, NC), and the mean separation was assessed using Least Significant Difference (LSD) at level of alpha = 0.05. LSD procedure was defined as

$$LSD = t_{\frac{\alpha}{2}}\sqrt{\frac{2MSError}{n}}$$

with $t_{\frac{\alpha}{2}}$ being the critical value from the t-table and having a degree of freedom [df(Sum of Square of the Error)] corresponding to the difference between the number of observations and

the number of replications, and n being the number of replications, $MSError$ is the mean square of the error.

RESULTS

Nested association mapping (NAM)

NAM was conducted to identify SNP makers associated with wilting, maturity, and seed yield. There were no QTLs detected at R1 and R2 for canopy wilting that exceeded the LOD threshold (3); however, a total of four QTLs were found to be associated with canopy wilting at R3 and R4. QTLs found at R3 were located on Gm06, Gm09, and two QTLs were located on Gm16 at R4 **(Table 4.1)**. The results indicated a total of 12, 25, 2 SNPs associated with maturity when irrigation was withheld at R2, R3, and R4, respectively. No SNPs having an LOD greater than the threshold (3) for maturity was detected when irrigation was withheld at R1. All SNPs were found on Gm10 except one at R2 located on Gm15 **(Table 4.1)**. Results did not show any significant SNPs having LOD greater than the threshold (3) for seed yield when irrigation was withheld at R1. But a total of 4, 2, 1 SNPs were detected when irrigation was withheld at R2, R3, and R4, respectively. These SNPs were located in Gm04, Gm05, Gm13, and Gm18 **(Table 4.1)**. Among all traits evaluated in this study, maturity had the highest number of significant SNPs. In addition, there is a lack of overlap between the significant SNPs across different irrigation onsets, and also across different traits, indicating that selection using a marker would be difficult as they are complex traits.

Candidate genes

A total of four candidate genes were found for canopy wilting under reduced irrigation at R3 and R4. These candidate genes consisted of *Glyma06g21495*, *Glyma09g04220*,

Glyma16g162500, and *Glyma16g164500* that encode for basic helix-loop-helix/leucine zipper transcription factor, DNA-directed RNA polymerase, alanine aminotransferase, and Zinc finger, DHHC-type, palmitoyltransferase, respectively **(Table 4.1)**.

Out of 39 SNPs found to be associated with maturity at different irrigation onsets (R2, R3, R4), 27 had annoted genes in their vicinity. As some SNPs were found at irrigation at R2, R3, and R4 at the same time, the genes found close to the top five (highest LOD) SNPs associated with maturity were *Glyma.10g205500, Glyma.10g210000, Glyma.10g214600, Glyma.10g210500, Glyma.10g209600.* The annotated gene *Glyma.10g205500* encodes for a phosphoenolpyruvate carboxylase 4. *Glyma.10g210000* and *Glyma.10g214600* encode for Calcium-dependent lipid-binding and glycine-rich cell wall structural protein 2-like, respectively. A probable lysine-specific demethylase JMJ14-like isoform X1 and a protein kinase superfamily protein; IPR011009 are encoded by *Glyma.10g210500* and *Glyma.10g209600* **(Table 4.1)**.

NAM suggested a total of 7 SNPs associated with seed yield under reduced irrigation (R2, R3, and R4) **(Table 4.1)**. Of which, five were mapped in the vicinity of annotated genes. *Glyma.18g026200, Glyma.18g028900, Glyma.18g004200, Glyma.13g181200, Glyma.04g003300* were the candidate genes that encode for protein YLS7-like anthranilate synthase 2, MACPF domain protein, mediator of RNA polymerase II transcription subunit 16, and short-chain dehydrogenase-reductase B, respectively **(Table 4.1)**.

Genomic selection

Overall, genomic selection for seed yield was moderate. The average genomic selection accuracy for seed yield was 0.44, 0.39, 0.39, and 0.41 for irrigation onset at R1, R2, R3 and R4,

respectively. Spearman's correlation of the GEBVs across the different irrigations were evaluated. Overall, correlations were positively moderate and high. The lowest Spearman's correlation coefficient was found between GEBVs at R1 and GEBVs at R3 (ρ=0.51), while the highest was between GEBVs at R2 and GEBVs at R4 (ρ=0.81). A principal component analysis (PCA) was used to assess the relationship across the yield GEBVs across different irrigation onsets. The PCA of the GEBVs identified two distinct components. There was a deviation from the GEBVs at R1 from the GEBVs at R2, R3, and R4 **(Figure 4.1)**.

Lines advancement and selection

Results showed a highly significant difference among genotypes (*p*-value <0.0001) in fully irrigated and non-irrigated conditions. Under irrigated conditions, AG53X0 had the highest yield (6166.02 kg/ha); however, it was not significant different from R18-7427 with 5672.85 kg/ha based on an LSD of 550.29 kg/ha **(Table 4.2)**. Moreover, yield of R18-7427 was not significantly different from the other commercial checks AG52XF0 (4335.77 kg/ha), S49-F5X (5406.78 kg/ha), AG48X9 (5489.42 kg/ha), AG56X8 (5185.72 kg/ha) **(Table 4.2)**. There was an average decrease in yield of 25% under non- irrigated condition. The highest yield under non-irrigated condition was R18-7427 with 4640.81 kg/ha was significantly different from the commercial checks based on the LSD of 418.59 kg/ha **(Table 4.2)**. By selecting the top ten highest yielding under irrigated and non-irrigated conditions, a total of four breeding lines remained in the top ten for both conditions: R18-7427, R18-7456, R18-7389, and R18-7467 **(Table 4.2)**.

DISCUSSION

Nested Association Mapping (NAM) was conducted to identify SNP markers associated with the canopy wilting, maturity, and seed yield. A total of 4, 39, and 7 SNPs were found to be significantly associated with canopy wilting, maturity, and seed yield, respectively, using the combined data under different irrigation onsets obtained over four environments (location-year combination). Diers et al. (2018) reported a total of 19 and 23 SNPs to be associated with maturity and seed yield, respectively using a nested association mapping (NAM) soybean population. The number of significant SNPs varied across different irrigation onsets for each trait.

SNPs related to canopy wilting and seed yield were distributed across the soybean genome and appeared not to be stable as the irrigation was delayed. However, SNPs related to the maturity were stable across different irrigations onsets. This could be explained that maturity is more heritable trait compared to canopy wilting and seed yield. Dutta et al. (2021) reported that heritability for maturity in soybean was 90.41 %. A recent study on genotypic and phenotypic parameters associated with early maturity in soybean carried by Silva et al. (2022) showed a heritability of 90.37 %. The most significant SNPs for maturity were found on chromosome 10. A total of 10 loci on chromosome 10 were reported to be associated with maturity in Soybase (https://www.soybase.org/).

For the canopy-related SNPs, the 4 SNPs mapped in our study were not reported in the Soybase. More than 75 SNPs was identified to be related to canopy wilting in Soybase using a bi-parental populations. For seed yield, previous reports showed that SNP markers associated with yield were scattered across the soybean genome. To date, more than 170 loci have been associated with yield in soybean in Soybase (https://www.soybase.org/). Zatybekov et al. (2017)

mapped SNP markers associated with soybean yield on chromosomes 14, 17, and 20 while our study SNPs yield were detected on chromosome 04, 05, 13, and 18. Diers et al. (2018) reported 23 loci affecting soybean yield on chromosome 16 alone. This suggests that seed yield is a complex trait that is trait controlled by a large number of loci (Assefa et al. 2019).

Genomic selection (GS) for seed yield was conducted using a ridge regression best linear unbiased predictor model. One of the primary goals of GS is to increase genetic gain for economically important traits within breeding programs by reducing the breeding cycle and by increasing the accuracy of selection (Asoro et al., 2013; Rutkoski et al., 2015). As a first step in GS, defining a training population, which consists of breeding lines phenotyped for a target trait and genotyped with genome-wide markers, is crucial. Once trained, the model is used to calculate GEBVs to predict performance on a testing population based simply on genotypic information. The correlation between the GEBVs and the estimated genetic values is used to calculate the prediction accuracy. The GS method is considered cost effective when prediction accuracy values are high enough (Combs & Bernardo, 2013). Moreover, GS has been proven to be effective when dealing with complex traits (Heffner et al., 2009). In this study, GS accuracy was moderate ranging from 0.39 to 0.44 across different irrigation onsets. Many factors affecting variability in prediction accuracy values have been reported including prediction models, breeding schemes, training population size, the relationship between the training and the prediction populations, trait complexities, marker densities, and genotyping platforms (Bernardo, 2016).

Even though the genomic selection accuracy was moderate, it can still supplement the phenotypic selection and would increase the genetic gain by at least 10% (Lozada et al. 2019). Based on the GEBVs ranking and the PCA analysis, breeders should perform independent

selection experiments for soybean under full irrigation as opposed to those targeted to withstand any level of water restriction. In fact, by selecting superior breeding lines based on their GEBVs, and testing them in irrigated and non- irrigated conditions, our results indicated that some lines such R18-7427, R18-7456, R18-7389, and R18-7467 were high yielding and stable under irrigated and non- irrigated. This demonstrates the efficiency of genomic selection for a complex trait like seed yield.

CONCLUSIONS

This study reported the variation in canopy wilting, maturity, seed yield based on NAM population soybean genotypes. To the best of our knowledge, this is one the few reports investigating the genetics of canopy wilting, maturity, seed yield trait under reduced irrigation. In addition, we showed that genomic selection was efficient to select superior individuals. The results from this investigation will contribute to a better understanding of genetic architecture of soybean lines under reduced irrigation. Also, that breeders should perform independent selection experiments for soybean under full irrigation as opposed to those targeted to withstand any level of water restriction.

REFERENCES

Anda, A., Soos, G., Menyhart, L., Kucserka, T., & Simon, B. (2020). Yield features of two soybean varieties under different water supplies and field conditions. *Field crops research*, *245*, 107673.

Arkansas Farm Bureau Federation (AFBF). (2022). Soybeans. Retrieved February, 2022 from https://www.arfb.com/pages/arkansas-agriculture/commodity-corner/soybean/

Asoro, F. G., Newell, M. A., Beavis, W. D., Scott, M. P., Tinker, N. A., & Jannink, J. L. (2013). Genomic, marker-assisted, and pedigree-BLUP selection methods for β-glucan concentration in elite oat. *Crop Science*, *53*(5), 1894-1906.

Assefa, T., Otyama, P. I., Brown, A. V., Kalberer, S. R., Kulkarni, R. S., & Cannon, S. B. (2019). Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. *BMC genomics*, *20*(1), 1-12.

Baghel, L., Kataria, S., & Guruprasad, K. N. (2018). Effect of static magnetic field pretreatment on growth, photosynthetic performance and yield of soybean under water stress. *Photosynthetica*, *56*(2), 718-730.

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. *R package version*, *2*(1), 74.

Bernardo, R. (2016). Bandwagons I, too, have known. *Theoretical and applied genetics*, *129*(12), 2323-2332.

Carter Jr, T. E. (1999). Recent advances in breeding for drought and aluminum resistance in soybean. In *Proceedings at the World Soybean Research Conference VI Chicago, IL, 1999* (pp. 106-125). Superior Printing.

Carter Jr, T. E., Orf, J., Purcell, L. C., Specht, J., Boerma, H. R., Chen, P., ... & Rufty, T. (2006). Tough times, tough plants-new soybean genes defend against drought and other stresses. In *American Seed Trade Association Conference Proceedings*.

Charlson, D. V., Bhatnagar, S., King, C. A., Ray, J. D., Sneller, C. H., Carter, T. E., & Purcell, L. C. (2009). Polygenic inheritance of canopy wilting in soybean [Glycine max (L.) Merr.]. *Theoretical and Applied Genetics*, *119*(4), 587-594.

Combs, E., & Bernardo, R. (2013). Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *The Plant Genome*, *6*(1), plantgenome2012-11.

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., ... & Pè, M. E. (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. *Genome biology*, *16*(1), 1-23.

Diers, B. W., Specht, J., Rainey, K. M., Cregan, P., Song, Q., Ramasubramanian, V., ... & Beavis, W. D. (2018). Genetic architecture of soybean yield and agronomic traits. *G3: Genes, Genomes, Genetics*, *8*(10), 3367-3375.

Dogan, E., Kirnak, H., & Copur, O. (2007). Deficit irrigations during soybean reproductive stages and CROPGRO-soybean simulations under semi-arid climatic conditions. *Field Crops Research*, *103*(2), 154-159.

Doyle, J. J. (1990). Isolation of plant DNA from fresh tissue. *Focus*, *12*, 13-15.

Dutta, P., Goswami, P. K., & Borah, M. (2021). Assessment of genetic variability, heritability and genetic advance in soybean genotypes. *Electronic Journal of Plant Breeding*, *12*(4), 1461-1465.

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics*, *11*(6), 446-450.

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The plant genome*, *4*(3).

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

Flint-Garcia, S. A., Thornsberry, J. M., & Buckler IV, E. S. (2003). Structure of linkage disequilibrium in plants. *Annual review of plant biology*, *54*(1), 357-374.

Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., ... & Buckler, E. S. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. *The plant journal*, *44*(6), 1054-1064.

Gowdy, J. (2020). Our hunter-gatherer future: Climate change, agriculture and uncivilization. *Futures*, *115*, 102488.

Grant, D., Nelson, R. T., Cannon, S. B., & Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic acids research*, *38*(suppl_1), D843-D846.

Heffner, E. L., Jannink, J. L., Iwata, H., Souza, E., & Sorrells, M. E. (2011). Genomic selection accuracy for grain quality traits in biparental wheat populations.

Henry, C. G., Espinoza, L., Ismanov, M., Francis, P., & Ross, J. (2014). Scheduling Irrigation using an Atmometer (ET Gauge) for Arkansas Soybeans. Retrieved April, 2022, from https://www.uaex.edu/environment-nature/water/docs/irrig-ET-Gage-Sheet-Soybean.pdf

Kaler, A. S., Ray, J. D., Schapaugh, W. T., King, C. A., & Purcell, L. C. (2017). Genome-wide association mapping of canopy wilting in diverse soybean genotypes. *Theoretical and Applied Genetics*, *130*(10), 2203-2217.

Lozada, D. N., Mason, R. E., Sarinelli, J. M., & Brown-Guedira, G. (2019). Accuracy of genomic selection for grain yield and agronomic traits in soft red winter wheat. *BMC genetics*, *20*(1), 1-12.

McMullen, M. D., Kresovich, S., Villeda, H. S., Bradbury, P., Li, H., Sun, Q., ... & Buckler, E. S. (2009). Genetic properties of the maize nested association mapping population. *Science*, *325*(5941), 737-740.

Mekonnen, M. M., Hoekstra, A. Y., Neale, C. M., Ray, C., & Yang, H. S. (2020). Water productivity benchmarks: The case of maize and soybean in Nebraska. *Agricultural water management*, *234*, 106122.

Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819-1829.

Ohashi, Y., Nakayama, N., Saneoka, H., & Fujita, K. (2006). Effects of drought stress on photosynthetic gas exchange, chlorophyll fluorescence and stem diameter of soybean plants. *Biologia Plantarum*, *50*(1), 138-141.

Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., & He, Z. (2017). Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Molecular plant*, *10*(8), 1047-1064.

Reba, M. L., & Massey, J. H. (2020). Surface irrigation in the lower Mississippi river basin: Trends and innovations. *Transactions of the ASABE*, *63*(5), 1305-1314.

Rutkoski, J., Singh, R. P., Huerta-Espino, J., Bhavani, S., Poland, J., Jannink, J. L., & Sorrells, M. E. (2015). Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat.

Shikha, M., Kanika, A., Rao, A. R., Mallikarjuna, M. G., Gupta, H. S., & Nepolean, T. (2017). Genomic selection for drought tolerance using genome-wide SNPs in maize. *Frontiers in plant science*, *8*, 550.

Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., ... & Han, B. (2016). OsSPL13 controls grain size in cultivated rice. *Nature genetics*, *48*(4), 447-456.

Silva, E. V. V., Bruzi, A. T., Silva, F. D. S., & Marques, F. S. (2022). Genotypic and phenotypic parameters associated with early maturity in soybean. *Pesquisa Agropecuária Brasileira*, *57*.

Sinclair, T. R., Marrou, H., Soltani, A., Vadez, V., & Chandolu, K. C. (2014). Soybean production potential in Africa. *Global Food Security*, *3*(1), 31-40.

Slaton, N., Roberts, T., & Ross, J. (2013). Fertilization and liming practices. *Arkansas soybean production handbook*, 2-16.

Song, Q., Yan, L., Quigley, C., Fickus, E., Wei, H., Chen, L., ... & Nelson, R. L. (2020). Soybean BARCSoySNP6K: An assay for soybean genetics and breeding research. *The Plant Journal*, *104*(3), 800-811.

SoyStats. (2022). http://soystats.com/value-history/. Accessed 2022-03-11.

Specht, J. E., Hume, D. J., & Kumudini, S. V. (1999). Soybean yield potential—a genetic and physiological perspective. *Crop science*, *39*(6), 1560-1570.

USDA Foreign Agricultural Service. (2018). USDA Foreign Agricultural Service Market and Trade Data. USDA Foreign Agricultural Service.https://data.nal.usda.gov/dataset/usda-foreign-agricultural-service-market-and-trade-data. Accessed 2022-03-11.

Wei, J., & Xu, S. (2016). A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. *Genetics*, *202*(2), 471-486.

Xavier, A., Xu, S., Muir, W. M., & Rainey, K. M. (2015). NAM: association studies in multiple populations. *Bioinformatics*, *31*(23), 3862-3864.

Xiao, Y., Liu, H., Wu, L., Warburton, M., & Yan, J. (2017). Genome-wide association studies in maize: praise and stargaze. *Molecular plant*, *10*(3), 359-374.

Xiao, Y., Tong, H., Yang, X., Xu, S., Pan, Q., Qiao, F., ... & Yan, J. (2016). Genome-wide dissection of the maize ear genetic architecture using multiple populations. *New Phytologist*, *210*(3), 1095-1106.

Xiong, R., Liu, S., Considine, M. J., Siddique, K. H., Lam, H. M., & Chen, Y. (2021). Root system architecture, physiological and transcriptional traits of soybean (Glycine max L.) in response to water deficit: A review. *Physiologia Plantarum*, *172*(2), 405-418.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., ... & Matsuoka, M. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature genetics*, *48*(8), 927-934.

Yu, J., Holland, J. B., McMullen, M. D., & Buckler, E. S. (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics*, *178*(1), 539-551.

Zatybekov, A., Abugalieva, S., Didorenko, S., Gerasimova, Y., Sidorik, I., Anuarbek, S., & Turuspekov, Y. (2017). GWAS of agronomic traits in soybean collection included in breeding pool in Kazakhstan. *BMC plant biology*, *17*(1), 1-8.

Table 4.1. Significant SNPs associated with canopy wilting, maturity, seed yield with their respective LOD (-log10(p_value)) value, annotated gene found within a 40-kb genomic region flanking the significant SNP, and functional annotation corresponding to the candidate gene.

Traits	Irrigation onset	SNP	Chr.	Position (bp)	LOD	Gene name	Functional annotation
Canopy wilting	R3	Gm06_18033759_A_G	Gm06	18033759	3.09	Glyma.06g21495	basic helix-loop-helix/leucine zipper transcription factor
	R4	Gm09_3101824_A_G	Gm09	3101824	3.48	Glyma.09g04220	DNA-directed RNA polymerase
		Gm16_32340079_G_A	Gm16	32340079	3.37	Glyma.16g162500	alanine aminotransferase
		Gm16_32534697_A_G	Gm16	32534697	3.32	Glyma.16g164500	Zinc finger, DHHC-type, palmitoyltransferase
Maturity	R2	Gm10_44744804_A_C	Gm10	44744804	4.82	Glyma.10g214600	glycine-rich cell wall structural protein 2-like
		Gm10_46008769_G_A	Gm10	46008769	3.95	NA	NA
		Gm10_44274964_T_G	Gm10	44274964	3.56	Glyma.10g210000	Calcium-dependent lipid-binding
		Gm10_43783265_G_A	Gm10	43783265	3.4	Glyma.10g205500	phosphoenolpyruvate carboxylase 4
		Gm10_44342605_G_A	Gm10	44342605	3.33	Glyma.10g210500	GATA transcription factor 9
		Gm10_44972284_T_C	Gm10	44972284	3.19	Glyma.10g217300	ATP-binding ABC transporter
		Gm10_44091618_T_C	Gm10	44091618	3.15	NA	NA
		Gm10_46119255_G_A	Gm10	46119255	3.11	Glyma.10g230400, Glyma.10g230300, Glyma.10g230500, Glyma.10g230600	uncharacterized protein LOC100804894, ornithine cyclodeaminase/mu-crystallin, Heavy metal transport/detoxification superfamily protein, Heavy metal transport/detoxification superfamily protein
		Gm10_45863169_A_G	Gm10	45863169	3.1	Glyma.10g227400	ATP-binding cassette protein
		Gm15_20639275_T_C	Gm15	20639275	3.08	NA	NA
		Gm10_45245539_G_A	Gm10	45245539	3.02	Glyma.10g220200	F-box and associated interaction domains-containing protein
		Gm10_46069887_A_G	Gm10	46069887	3.02	Glyma.10g229800	ABC transporter G family member 8-like

Table 4.2. Cont.

Traits	Irrigation onset	SNP	Chr.	Position (bp)	LOD	Gene name	Functional annotation
		Gm10_43783265_G_A	Gm10	43783265	7.23	Glyma.10g205500	phosphoenolpyruvate carboxylase 4
		Gm10_44091618_T_C	Gm10	44091618	6.28	NA	NA
		Gm10_44274964_T_G	Gm10	44274964	6.11	Glyma.10g210000	Calcium-dependent lipid-binding
		Gm10_44744804_A_C	Gm10	44744804	5.91	Glyma.10g214600	glycine-rich cell wall structural protein 2-like
		Gm10_44342605_G_A	Gm10	44342605	5.67	Glyma.10g210500	GATA transcription factor 9
		Gm10_44137020_T_C	Gm10	44137020	5.66	NA	NA
		Gm10_43984045_G_A	Gm10	43984045	5.21	NA	NA
		Gm10_44227652_C_T	Gm10	44227652	5	Glyma.10g209600, Glyma.10g209500	probable lysine-specific demethylase JMJ14-like isoform X1, Protein kinase superfamily protein; IPR011009
		Gm10_44972284_T_C	Gm10	44972284	4.84	Glyma.10g217300	ATP-binding ABC transporter
		Gm10_43894668_A_G	Gm10	43894668	4.26	Glyma.10g206500	myb transcription factor
		Gm10_43612052_T_G	Gm10	43612052	4.08	Glyma.10g204200	two-component response regulator-like APRR2-like isoform X2
Maturity	R3	Gm10_45245539_G_A	Gm10	45245539	4.04	Glyma.10g220200	F-box and associated interaction domains-containing protein
		Gm10_43178809_G_T	Gm10	43178809	3.8	Glyma.10g200100	dual specificity protein phosphatase (DsPTP1) family protein
		Gm10_45890133_C_T	Gm10	45890133	3.75	Glyma.10g227700	chitinase A
		Gm10_44563220_A_G	Gm10	44563220	3.72	NA	NA
		Gm10_44042822_A_G	Gm10	44042822	3.71	NA	NA
		Gm10_46008769_G_A	Gm10	46008769	3.69	NA	NA
		Gm10_44553009_T_C	Gm10	44553009	3.66	Glyma.10g212100, Glyma.10g212200, Glyma.10g212300	Eukaryotic integral membrane protein, ubiquitin-conjugating enzyme 20, MLO-like protein 12-like
		Gm10_46177554_C_T	Gm10	46177554	3.6	Glyma.10g231300	unknown protein
		Gm10_45863169_A_G	Gm10	45863169	3.55	Glyma.10g227400	ABCC subfamily ATP-binding cassette protein
		Gm10_46069887_A_G	Gm10	46069887	3.48	Glyma.10g229800	ABC transporter G family member 8-like
		Gm10_43107961_A_G	Gm10	43107961	3.4	NA	NA

Table 4.3. Cont.

Traits	Irrigation onset	SNP	Chr.	Position (bp)	LOD	Gene name	Functional annotation
Maturity	R3	Gm10_46119255_G_A	Gm10	46119255	3.38	Glyma.10g230500	Heavy metal transport/detoxification superfamily protein
		Gm10_41445184_T_C	Gm10	41445184	3.34	Glyma.10g180000	Auxin-responsive protein
		Gm10_44574663_C_T	Gm10	44574663	3.25	Glyma.10g212400, Glyma.10g212500	tetraspanin-6, serine carboxypeptidase-like 27
	R4	Gm10_44744804_A_C	Gm10	44744804	3.4	Glyma.10g214600	glycine-rich cell wall structural protein 2-like
		Gm10_45245539_G_A	Gm10	45245539	3.06	Glyma.10g220200	F-box and associated interaction domains-containing protein
Seed yield	R2	Gm18_1957770_T_C	Gm18	1957770	3.7	Glyma.18g026200	protein YLS7-like
		Gm18_2178121_C_T	Gm18	2178121	3.58	Glyma.18g028900	anthranilate synthase 2
		Gm05_32327497_T_C	Gm05	32327497	3.1	NA	NA
		Gm18_347275_C_A	Gm18	347275	3.07	Glyma.18g004200	MACPF domain protein
	R3	Gm13_28868130_A_C	Gm13	28868130	3.37	Glyma.13g181200	mediator of RNA polymerase II transcription subunit 16
		Gm04_282832_A_G	Gm04	282832	3	Glyma.04g003300	short-chain dehydrogenase-reductase B
	R4	Gm18_9886770_G_A	Gm18	9886770	3.2	NA	NA

Table 4.4. Yield of different selected soybean lines based on GEBVs under irrigated and non-irrigated conditions during the summer of 2021

Genotype	Yield irrigated (kg.ha^{-1})	Yield non-irrigated (kg.ha^{-1})
R18-7427	5672.85	4640.81
R18-7416	3782.80	4334.43
R18-7456	5344.96	4230.28
R18-7393	4277.32	4198.03
R18-7463	4988.19	4115.39
R18-7389	5287.85	4095.23
AG52XF0	4335.77	4039.46
R18-7414	5027.16	4017.96
R18-7467	5500.17	3909.11
R18-7466	5110.47	3843.94
R18-7479	5220.66	3809.67
S49-F5X	5406.78	3787.50
R18-347	4742.27	3743.15
R18-7391	4444.62	3636.99
AG53X0	6166.03	3618.85
R18-7462	5045.97	3607.43
R18-7455	4931.07	3593.32
AG48X9	5489.42	3546.96
R18-300	4830.96	3523.44
R18-7447	5467.92	3473.05
R18-7395	4731.52	3454.24
R18-7483	4295.46	3431.39
R18-7477	4791.99	3378.31
R18-336	4541.37	3363.53
R18-7461	3863.43	3361.52
R18-7475	4806.77	3295.00
AG56X8	5185.72	3283.58
R18-7422	5236.12	3023.55
R18-7453	3475.74	2914.03
R18-7397	4692.55	2885.81
Grand mean	4889.80	3671.87
LSD 0.05	550.29	418.59

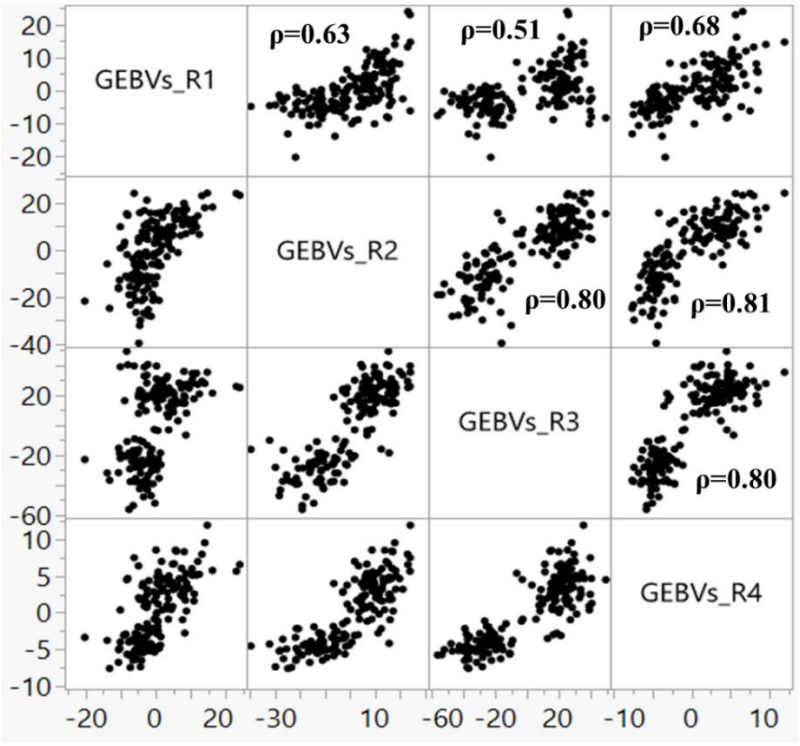

Figure 4.1. *Spearman's correlation among the GEBVs in each irrigation treatment*

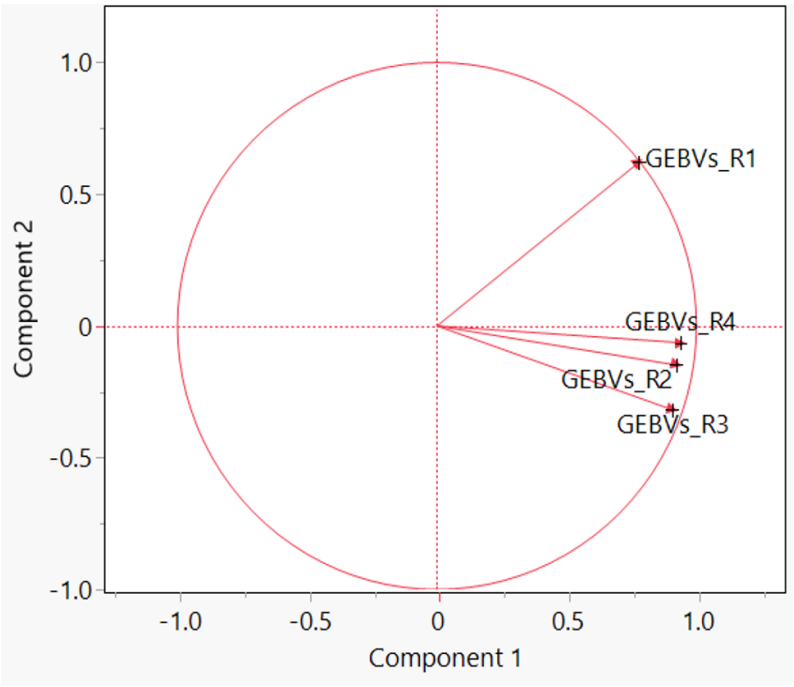

Figure 4.2. Principal component analysis among the GEBVs for R1, R2, R3, R4 stages

CHAPTER V

Spatial models for seed yield, wilting, and maturity in furrow-irrigated soybean plots

Ravelombola, F.[1], Acuña, A.[1], Florez-Palacios, L.[1], Wu, C.[1], Harrison, D.[1], de Oliveira, M.[1], Winter, J.[1], Da Silva, M.P.[2], and Mozzoni, L[1,*]

[1]Department of Crop, Soil, and Environmental Sciences, University of Arkansas, Fayetteville, AR 72701.

[2]Department of Entomology and Plant Pathology, University of Arkansas, Fayetteville, AR, 72701, USA

*Corresponding author (lmozzon@uark.edu)

Ravelombola, F., Acuña, A., Florez-Palacios, L., Wu, C., Harrison, D., de Oliveira, M., Winter, J., Da Silva, M.P., and Mozzoni, L. (2022). Spatial models for seed yield, wilting, and maturity in furrow-irrigated soybean plots. *Journal of Crop Improvement.* (Accepted)

ABSTRACT

Field experiments are subjected to spatial variability due to factors such as soil moisture, fertility, pH, and structure, as well as the pressure of diseases and pests. Soybean yields have been shown to be highly variable across fields. Controlling spatial variability could decrease the risk of erroneous inferences in breeding trials. This study aims at evaluating the spatial variability of furrow-irrigated soybean for seed yield, wilting, and maturity under four different irrigation levels. The field experiment was conducted in four environments (location-year combination). A total of 165 soybean lines of similar relative maturity (maturity group 5) along with commercial checks were planted in an augmented strip plot design. Irrigation treatment decisions were triggered using an atmometer based on a threshold at a designated growth stage. Data were analyzed via Analysis of Variance as a linear mixed model using a blocking structure (block model) and spatial covariances using range and column. Two different spatial models were used: exponential and gaussian. Results showed that the spatial models displayed better data fitting (lower AIC and/or BIC) than the block model in each different irrigation level across different environments and traits. Indeed, genotype ranking for seed yield was different between the block model and the best spatial model, suggesting that spatial adjustment may be necessary for soybean breeding operations under furrow irrigation. Further validation in a breeding yield trial demonstrated similar results of the effectiveness in terms of AIC and/or BIC of the spatial model compared to the block model for soybean seed yield.

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.]) is one of the most important worldwide crops, with a cultivated area of 126 million hectares (Mha) and a total production of 353 million tons in 2021 with 113 million tons produced by the United States (U.S.) (FAOSTAT, 2022). It is one of the most economically important crops in Arkansas, providing a substantial source of revenue to growers, as Arkansas ranks among the top ten producers in the U.S. (USDA NASS, 2021). A total of 4.84 million tons of soybean are produced by more than 6,800 Arkansans farmers in 1.4 million hectares (Mha), making it the largest crop by acreage in the state (USDA NASS, 2021).

Arkansas is the third largest irrigated state (2.0 Mha) in the U.S. after Nebraska (2.4 Mha) and California (3 Mha) (USDA NASS, 2021). Most of the soybean acres in Arkansas are irrigated, with only 15% of Arkansas soybeans rainfed (AFBF, 2021). The most common irrigation practice in the U.S. Mid-South is furrow irrigation (Bryant et al., 2017; Maupin et al., 2014), which consists of creating a channel where the water can flow and percolate by gravity. It is a cost-efficient method as it lessens water loss by gravity and offers a quick massive irrigation (Brouwer et al., 1990; Bryant et al., 2017; Massey et al., 2017). The effectiveness of the furrow irrigation resides on a positive and continuous row grade that requires precision land grading (Quintana-Ashwell et al., 2020). Since most Arkansas' soybean acres are irrigated, and the most common irrigation method is furrow-irrigation, the soybean breeding program at the University of Arkansas System – Division of Agriculture (UADA-SBP) heavily relies on said irrigation practices for its breeding trials, in order to have prediction environments that closely resemble target environments.

In the early yield trial stages of a breeding program, a large number of new genotypes has to be evaluated (usually 2,000 to 5,000 new entries per year for the UADA-SBP) in non-or limited-replication trials. A proposed experimental design to reduce error is to use a block design

called the augmented experimental design which allows the adjustment of the breeding line means for environmental effects estimated on the repeated checks (Federer, 1956; Patterson et al., 1978; Morsy & Fares, 2016, Kumar et al., 2020). Moreover, field experiments are subjected to spatial variability which includes soil texture, fertility, and pH, among others (Wibawa et al., 1993; Gaston et al., 2001; Tola et al., 2017). Furrow irrigation could add an extra dimension of variation because of water gradients on the front and back of field, and potentially unequal flows between rows. With large experiment sizes resulting from augmented block designs grouped by breeding and maturity cohorts, it is very likely to expose soybean plots to various water regimes, soil texture, structure, fertility, and salts (Bautista & Wallender, 1985; Bali & Wallender, 1987; Drewry et al., 2021; Haghnazari et al., 2015; Zaman et al., 2018). The assessment of genotypic effects may be inaccurate, hence lowering the precision of the breeding program's selection. Therefore, controlling spatial variability in field experiments is necessary to reduce the error in the statistical model such as the estimation of genotype values, and the risk of misleading or erroneous inferences (Mo & Si, 1986; Stroup, 2002). In fact, studies have shown that randomized block designs including complete and incomplete blocks and lattices are often not optimal and in consequence results in poor analysis efficiency (Casler, 2015; Yang et al., 2004).

There are two methods for controlling spatial variation: the first uses spatial variance–covariance structures, while the second uses smoothing techniques. In the first case, the model includes correlation related to the rows and columns in the field. Cullis & Gleeson (1991) used an autoregressive model. Gilmour et al. (1997) extended it using a linear variance model, Piepho & Williams (2010) discussed it. Smoothing approaches, on the other hand, were the first applied in agriculture by Green et al. (1985). Smoothing methods have shown to be effective for modeling large-scale dependency; they have the drawback of not always being able to capture

small-scale dependence. Our study aims at evaluating the model effectiveness of the spatial model using the row-column adjustment when used to control for spatial variability in furrow-irrigated soybean research trials as compared to a linear mixed model using blocking structure.

MATERIALS AND METHODS

Plant materials and experimental design

A total of 165 F_4-derived breeding lines (73 derived from the cross N07-14753/R11-1057 and 92 derived from R11-2933/R11-1057) along with four commercial checks (AG55X7, AG56X8, P53AG7X, P55A49X) were grown during the summer of 2019 and 2020 in Stuttgart (19STU and 20STU) (silt loam soil) and Rowher (19ROH and 20ROH) (silt loam soil), Arkansas. All lines in both progenies of the two crosses have similar relative maturity group (maturity group 5). The experimental design was an augmented strip-plot design under four furrow-irrigation conditions. The irrigation levels were: 1) full irrigation (irrigation initiated at initiation of flowering (R1 stage; Fehr & Caviness, 1977), 2) irrigation initiated at full flowering (R2 stage), 3) irrigation initiated at beginning of pod development (R3 stage), and 4) irrigation initiated when pods were 2 cm at one of the four uppermost nodes (R4 stage). The irrigation at each designated growth stage was triggered using the decision table developed by Henry et al. (2014) for atmometer measurements based on 50% of the plots reaching the desired stage. Each strip of irrigation onset (R1, R2, R3, and R4) was composed of ten blocks. One block was composed of four checks (two parental checks and two commercial checks) and 16 randomly assigned genotypes, including seven to eight and nine to ten genotypes from the first and second progeny of the two crosses, respectively, where individual lines were a random factor within populations **(Figure 5.1)**. The plots were 4.6 m long with 1.5 m alley, and consisted of two rows 0.97 m apart in 19ROH and 20ROH and 0.91 m apart in 19STU and 20STU. The plots consisted

of two rows 0.76 m apart, 4.6 m long with 1.5 m alley. The planting date was 05/30/2019, 05/28/2019, 05/21/2020, and 05/19/2020 for 19STU, 19ROH, 20STU, and 20ROH, respectively. Each plot in the field was identified by a unique range and column within each environment (**Figure 5.1**). Standard agronomic practices were used at each location, including fertilization to recommended levels as defined by Slaton et al. (2013).

Phenotyping

Visual wilting severity was taken using a ten-point scale: 0 (no wilting) to 9 (plant death). Wilting score was rated prior to the time of triggering irrigation for each irrigation treatment. Rating was conducted between 11:00 am and 3:00 pm to reduce the impact of diurnal variation in evaporative demand. Maturity was recorded as the day when 95% of the pods in a plot had reached mature pod color (R8; Fehr & Caviness, 1977), and expressed as the number of days after 31st August. Plot seed yield (kg.ha^{-1}) was calculated based on the moisture at harvest, the weight at harvest, the length of the plot, and the row spacing of the two-row plots in each environment. The combine harvester was a plot combine ZÜRN 150 (Zürn Harvesting GmbH & Co. KG, Germany).

Data analysis

Statistical analysis was run using PROC MIXED in SAS v. 9.4 (SAS Institute, Cary, NC) using three different models: linear mixed model with blocking structure; and spatial analyses (range and column adjustment) (Hu & Spilke, 2009) including exponential and gaussian models. The exponential model and the gaussian model are defined by (Hu & Spilke, 2009) as follows:

EXP model $f(d_{ij}) = \exp(-d_{ij}/\theta)$ and GAU model $f(d_{ij}) = \exp(-d_{ij}^2/\rho^2)$

where d_{ij} is the distances of observations *i* and *j* in direction row and column, θ and ρ^2 are the covariances parameters.

The linear mixed model with blocking structure was defined as

$$y_{ijk} = \mu + g_i + t_j + gt_{ij} + b_k + e_{ijk}$$

where i was number of genotypes (1…169), j the number of irrigation onsets R1, R2, R3, and R4 (1, 2, 3, 4) and k the number of blocks (1…10)

The spatial designs model with spatial residuals was defined as:

$$y_{ijkl} = \mu + g_i + t_j + gt_{ij} + w_{(kl)} + e_{ijkl}$$

where *i* was number of genotypes (1…169), *j* the number of irrigation onsets R1, R2, R3, and R4 (1, 2, 3, 4), *k* the number of range (1…10), and *l* the number of column (1…80)

y_{ijk} and y_{ijkl} were the mean response (wilting severity score, maturity, and yield), μ the overall mean response in each model. The fixed effects were the genotype (g_i), the irrigation level (t_j), and the interaction irrigation level and genotype (gt_{ij}). The random effects included block (b_k) for the block design and the range and column (w_{kl}) for the exponential, and the gaussian models. The experimental errors associated with *ijkth* and *ijklth* observation is e_{ijk} and were e_{ijkl} in each model.

In each analysis, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were output to evaluate how well each model fitted the data (Burnham & Anderson, 2004). The studentized residuals plots and the quantile-quantile (QQ) plots were also generated to assess the pattern of the model in each environment (19ROH, 19STU, 20ROH, and 20STU).

The yield Best Linear Unbiased Prediction (BLUP) for each breeding line under different irrigations was computed. Based on the yield BLUP for linear mixed model with blocking

structure and the best spatial analysis model, the top 15% (approximately 25 lines) was selected and the ranking of the selected lines was evaluated for consistency across models. Also, a Kendall correlation was also run in JMP Pro 16.0 (SAS Institute, Cary, NC) to assess the non-parametric correlation between of the ranking of the selected lines in the linear mixed model with blocking structure and the spatial models in each irrigation level.

Validation of spatial adjustment in breeding trial datasets

To validate the efficiency of the spatial adjustment in a breeding program, an unrelated seed yield dataset from the 2019 Arkansas Final yield trials maturity Group 4 Early noted as 19AF4E and 2020 Arkansas Final yield trials maturity group 4 Early noted as 20AF4E of UADA-SBP were used. The 19AF4E were tested in four different Arkansas furrow-irrigated environments arranged as a randomized complete block (RCBD) with two replications in 2019 (Keiser, Pine Tree, Rohwer, and Stuttgart) (19KEI, 19PTR, 19ROH, 19STU, respectively). The 20AF4E trial was evaluated in 2020 in two Arkansas furrow-irrigated environments (Rohwer and Shoffner) (20ROH and 20SHO) as a RCBD with two replications. PROC MIXED procedure was run in SAS v 9.4 with three models (linear mixed model with blocking structure, exponential, and gaussian models) as previously described. Models were compared using AIC and/or BIC. In addition, seed yield BLUPs from the blocking structure and the best spatial model were used to evaluate the change of ranking in each line for the 19AF4E and 20AF4E yield trials. The narrow sense heritability was estimated using the following equation:

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_e^2}{r}}$$

where h^2 the is narrow sense heritability, σ_G^2 is the variance of the genotype, σ_e^2 is the variance of error, and r is the number of replications. Because F_5-derived RILs were used in this study, σ_G^2

was composed entirely of additive variance and additive × additive epistasis variance, with negligible variance associated with other components of dominance variance. As the result, this heritability should be considered as narrow sense heritability.

RESULTS

Comparison of the different models for wilting, maturity, and seed yield

Results showed that AIC and/or BIC of the exponential and gaussian models were smaller than those of the linear mixed model with blocking structure and very similar between each other. Moreover, we observed that genotype was a significant factor for the linear model with blocking structure (p-value < 0.05); however, there was no significant difference in terms of the wilting severity for the environment 19STU and 19ROH (p-value > 0.05) for the exponential and the gaussian models **(Table 5.1)**. Results showed that there was a highly significant difference (p-value < 0.01) of the irrigation treatment on the wilting severity for the mixed model with blocking structure, exponential and gaussian models, except for 19ROH environment for the gaussian model. Statistical tests on the effect of the interaction between genotype and irrigation showed no statistically significant differences for 19ROH and 19STU environments for the spatial models, while there was a highly significance (p-value < 0.0001) for the environment 20ROH and 20STU across all models **(Table 5.1)**. The QQ plots for the wilting severity **(Figure 5.2)** indicated an improvement of the spatial models compared to the mixed model with blocking structure. Non- spatially adjusted mixed model with blocking structure showed significant deviation of the predicted from the observed values. However, the exponential model showed best fit as the points approximately followed a straight line. The studentized residuals plots **(Figure 5.3)** showed a heteroscedasticity for all the models. The residual values followed straight lines, as result of the discrete categorical scale used to rate canopy wilting.

Analysis on maturity displayed akin results as the wilting severity in terms of the efficiency of the spatial model. In fact, a generally smaller BIC were found for the exponential and the gaussian models **(Table 5.2)** as compared to the linear mixed model with blocking structure. However, AIC value for spatial models were slightly lower than the linear mixed model with blocking structure. Significant (p-value < 0.05) maturity differences among the genotypes were detected by the three models across all locations, except for the 19STU for the exponential model, and for 19ROH, 19STU, and 20STU for the gaussian model **(Table 5.2)**. The effect of irrigation on maturity was found to be statistically significant for all environments and all models, except for 19STU for the exponential and 19ROH for the gaussian model. Regarding the interaction between irrigation and genotype on maturity, there was a discrepancy of results between the two spatial models. No significant interaction was discovered across all environments for the gaussian model except for 20ROH; however, the exponential and the linear mixed model with blocking structure models both showed significant interactions between genotypes and irrigation except 19STU and 20STU and **(Table 5.2)**. When comparing the spatial models to the blocking structure models, the QQ plots for maturity **(Figure 5.4)** showed an enhancement in terms of the linearity of plots. Also, all of the models demonstrated heteroscedasticity for the studentized residuals plots **(Figure 5.5)**.

According to the AIC and/or BIC, the spatial models were found to be the best-fitting models in all environments for the seed yield **(Table 5.3)**. The exponential model was the best-fit spatial model for all environments. Variation among genotypes was not significant for two environments (19ROH and 19STU; p-value > 0.05) in the linear mixed model with blocking structure model but was significant in all environments with the exponential model **(Table 5.3)**. Also, a significant irrigation effect on seed yield in all four environments was observed for the

linear mixed model with blocking structure and the gaussian models; however, the exponential model showed no significance for the 19STU environment. Finally, all models were consistent in terms of the significance of the interaction terms across all four environments. The Studentized residuals plots **(Figure 5.7)** showed a homoscedasticity for the spatial model and heteroscedasticity for the linear mixed model with blocking structure model. The spatial models outperformed the linear mixed model with blocking structure model as observed at the QQ plots for yield data **(Figure 5.6)**, for the non-spatially adjusted design, the predicted values were considerably different from the observed values. The exponential model, on the other hand, had the best match, with the points roughly following a straight line.

Results of BLUP and ranking of the augmented design and the exponential models are summarized in the Supplementary **Table S 5.1**. We observed a reduced rank correlation between mixed model with blocking structure and the exponential models in our study. Indeed, in our study the spatial model (exponential model) showed better fit than the linear mixed model with blocking structure model. The Kendall's correlation between the yield BLUPs of genotypes between the two models was calculated for each irrigation treatment, and we observed that the correlation between the augmented design model and the exponential model was 0.72, 0.65, 0.74, and 0.58 for R1, R2, R3, and R4 irrigation treatment, respectively (**Figure 5.8**, p-value < 0.0001).

Selection of the top 25 of the highest yielding breeding lines revealed a coincidence of 70% of selected breeding lines between the two models (the linear mixed model with blocking structure and exponential models) for the results of the irrigation at R1. However, some breeding lines that ranked top five in the linear mixed model with blocking structure model, for instance R18-7479, was poorly ranked in the exponential model (33rd). On the other hand, the line

R18-367 ranked 7th in the exponential model and placed 29th in the linear mixed model with blocking structure model. The results of irrigation at R2 showed that there was a total of 56% of the overlapping of the selected breeding lines between the blocking structure and the exponential model selections. Nonetheless, big differences were observed, for instance genotype R18-7411 that ranked second place in the exponential model was positioned 43rd in the blocking structure model. On the other side, the genotype R18-7463 placed second place in the blocking structure model was ranked 149th in the exponential. For the results of the irrigation at R3, we observed 68 % of overlap, while the R4 had 56 % of overlap between the blocking structure and the exponential model selections. Despite being chosen in the top 15% in the blocking structure model for the onset at R3, R18-7453 ranked low in the exponential model (42nd). For genotype R18-323, the results indicated a ranking of sixth in the exponential model at irrigation R4; but, when the yield data was analyzed as a linear mixed model with a blocking structure model, R18-323 ranked 27th.

Validation of spatial adjustment in breeding trial datasets

Similar analysis comparing the two spatial models and blocking structure model was run from the 2019 and 2020 Arkansas Final yield trials maturity Group IV Early (19AF4 and 20AF4E) in the Soybean Breeding Program of Arkansas. Our results stated that the spatial models outperformed the blocking structure model for seed yield in fully-irrigated breeding trials. Lower AIC and/or BIC were reported for the exponential model **(Table 5.4)** for seed yield in each environment than the blocking structure model. The effect of genotype on seed yield was significant for 19KEI in the blocking structure model from the 19AF4E **(Table 5.4)**; however, no statistical difference among genotypes was found when the spatial model was applied. In the same way for the 20AF4E, there was a statistical difference (p-value < 0.05) of genotype on

yield in 20ROH, but the same environment analyzed with the exponential and gaussian models showed no statistical genotypic effects.

The 19AFE4 trial consisted of 34 entries. The top 10 high yielding genotypes selected by the exponential model included two genotypes (R13-1463 and R17-2000) that were not selected by the blocking structure model **(Supplementary Table S 5.2)**; since these analyses are retrospective, there is no information on the performance of said genotypes in subsequent breeding stages. Finally, up to the seventh ranked genotype, the same genotypes were selected for both block and exponential models. Unlike the 19AF4E, the 20AF4E trial (composed of 19 entries) revealed a similar ranking among the genotypes for the exponential and block models **(Supplementary Table S 5.3)**. Heritabilities in each location **(Table 5.5)** for the blocking structure model and the exponential model showed that there was an average improvement of 6%.

DISCUSSION

One of the most critical aspects of agricultural experimentation is the proper choice of experimental design to control field heterogeneity, especially for large experiments (Casler, 2015; Piepho et al., 2015). According to Fisher (1937), well-designed experiments are based on the three principles: randomization, replication, and local control. Local control or blocking and randomization are associated with controlling spatial variation. The purpose of blocking is to decrease the variation between plots by defining a homogeneous block. Therefore, blocking is an effective way to control experimental error (Casler, 2015), but it is not enough in a situation where field heterogeneity and the size of the experiment are large (Brownie et al., 1993). Furthermore, designs including blocks without considering the real spatial variation among experimental units can strongly decrease the effectiveness of an experiment (Casler, 2015).

Our results showed that, across different environments and traits, the exponential model had the lower AIC and/or BIC, resulting in a greater efficiency compared to the linear mixed model with blocking structure. The smaller AIC and/or BIC, the better the model because of a lower test error (Hoefler et al., 2020). Range–column designs (Williams et al., 2006) were specifically created to control field heterogeneity. Therefore, models with more covariance parameters (range and column vs block) fit better than those with a simpler structure (Duarte & Vencovsky, 2005). When there is a large number of treatments where spatial variability resides, spatial models better suit the data (Müller et al., 2010). In our study, an exemplary data at irrigation R1 in 20STU of the spatial trend of unadjusted seed yield (raw data), block, exponential, and gaussian could be visualized in **Figure 5.9**.

Müller et al. (2010) found that in barley (*Hordeum vulgare* L.) and sugar beet (*Beta vulgaris* L.) trials, the best fit according to AIC for most cases was a block design. Kravchenko et al. (2006) stated that spatial analyses were not always superior to block designs, especially when the spatial structure of the variable could not be clearly defined. It is worth mentioning that in our study, we do not have soil sample analysis in each plot. However, because of the consistency of the results both in the designed experiment, and in the post-hoc reanalysis of breeding datasets, it is expected that the field variability resulting from furrow-irrigation soybeans is sufficient to justify the use of spatial adjustments in breeding plots.

The results reported in the current study showed heteroscedasticity on the student residuals plots, and lack of a straight line for the QQ plot, especially for the wilting and maturity data, for all the three models. However, our results showed that the QQ and the student residuals indicated an improvement of the spatial models compared to the mixed model with blocking structure for the seed yield data. This improvement in error control indeed varied according to

the trait studied. Dependence between efficiencies and different traits has also been detected in other studies (Chen et al., 2018; Dutkowski et al., 2006; Paget et al., 2015).

In yield trials, breeders need to separate genotypes based on yield differences (Casler & Undersander, 2000). Also, the choice of experimental designs and analysis models are needed to control spatial variation. Spatial variability often occurs gradually, and sometimes it is not captured well enough by the experimental design (Grondona & Cressie, 1991). Based on AIC and/or BIC, and the shape of the QQ and studentized residuals plots in our study, we selected the exponential model to spatially-adjust yield values of genotypes. These results agree with Duarte & Vencovsky (2005) that identified an autocorrelation in soybean research plot data that was explained using an exponential model. In their study, the authors also observed a reduction in *p*-value of genotype effects, and a change on genotype rankings after spatial adjustment, as compared to the non-adjusted augmented design model. In our study, an overlap of selected genotypes as low as 56% was detected, meaning that approximately half of the genotypes or lines selected from the blocking structure model was not chosen in the exponential model and vice versa. Stroup et al. (1994) stated that the variability in the blocks may result in inaccurate estimations, lowering the capacity to differentiate the best genotypes. In an actual and simulated wheat crop data, Borges et al. (2019) demonstrated with a large number of genotypes and a significative spatial variability, the right design is critical for achieving more precision and better estimations of genetic effects. Using the best design also helps the difficulties of differentiating genotypes, error effects, and improve heritabilities (Dutkowski et al., 2006). The outcomes of our study support other authors' assertions that spatial models can improve efficiency (Qiao et al., 2000; Sarker, Singh, & Erskine, 2001; Borges et al., 2019; Dutkowski et al., 2006; Gezan et al., 2010).

As breeding stages advanced, the number of entries in testing, and the concurrent trial field footprint decreases. The 19AF4E and the 20AF4E tests had a lower number of entries as compared to the designed experiment in our irrigation treatment study. Even though fewer entries and smaller trial footprint could have been hypothesized to be subject to lesser field variation, the exponential model showed a better fit to the data compared to the blocking structure model. The ranking of genotypes did differ in the 19AF4E trial, although it did not in the 20AF4E. Changes in line rankings affect selections and advancements. Plant breeding's major purpose is to choose the best genotypes; hence, the phenotypic data collection and interpretation are crucial in this process. As a result, the spatial analyses may have an influence on plant breeding selections.

CONCLUSION

In variety trials with large numbers of treatments, spatial analysis allowed better discrimination among genotypes and increased heritabilities. The spatial analysis led to a different ranking of the genetic materials in comparison with the non-spatial analysis, and selections could have been less influenced by local variation. Such differences in selections may have significant consequences for the outcome of plant breeding programs.

REFERENCES

Arkansas Farm Bureau Federation (AFBF). (2021). Soybeans. Retrieved October, 2021, from https://www.arfb.com/pages/arkansas-agriculture/commodity-corner/soybean/

Bali, K., & Wallender, W. W. (1987). Water application under varying soil and intake opportunity time. *Transactions of the ASAE*, *30*(2), 442-0448.

Bautista, E., & Wallender, W. W. (1985). Spatial variability of infiltration in furrows. *Transactions of the ASAE*, *28*(6), 1846-1851.

Borges, A., González-Reymundez, A., Ernst, O., Cadenazzi, M., Terra, J., & Gutiérrez, L. (2019). Can spatial modeling substitute for experimental design in agricultural experiments? *Crop Science*, *59*(1), 44-53.

Brouwer, C., Prins, K., Kay, M., & Heibloem, M. (1990). Irrigation methods. *Irrigation Water Management Training Manual (ILRI/FAO).*, (5).

Brownie, C., Bowman, D. T., & Burton, J. W. (1993). Estimating spatial variation in analysis of data from yield trials: a comparison of methods. *Agronomy Journal*, *85*(6), 1244-1253.

Bryant, C. J., Krutz, L. J., Falconer, L., Irby, J. T., Henry, C. G., Pringle III, H. C., ... & Wood, C. W. (2017). Irrigation water management practices that reduce water requirements for Mid-South furrow-irrigated soybean. *Crop, Forage & Turfgrass Management*, *3*(1), 1-7.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.

Burt, C. M., Clemmens, A. J., Strelkoff, T. S., Solomon, K. H., Bliesner, R. D., Hardy, L. A., ... & Eisenhauer, D. E. (1997). Irrigation performance measures: efficiency and uniformity. *Journal of irrigation and drainage engineering*, *123*(6), 423-442.

Casler, M. D. (2015). Fundamentals of experimental design: Guidelines for designing successful experiments. *Agronomy Journal*, *107*(2), 692-705.

Casler, Michael D., and Daniel J. Undersander. "Forage yield precision, experimental design, and cultivar mean separation for alfalfa cultivar trials." *Agronomy Journal* 92.6 (2000): 1064-1071.

Chen, Z., Helmersson, A., Westin, J., Karlsson, B., & Wu, H. X. (2018). Efficiency of using spatial analysis for Norway spruce progeny tests in Sweden. *Annals of forest science*, *75*(1), 1-13.

Cullis, B. R., & Gleeson, A. C. (1991). Spatial analysis of field experiments-an extension to two dimensions. *Biometrics*, 1449-1460.

Drewry, J. J., Carrick, S., Penny, V., Houlbrooke, D. J., Laurenson, S., & Mesman, N. L. (2021). Effects of irrigation on soil physical properties in predominantly pastoral farming systems: A review. *New Zealand Journal of Agricultural Research*, *64*(4), 483-507.

Dutkowski, G. W., Costa e Silva, J., Gilmour, A. R., Wellendorf, H., & Aguiar, A. (2006). Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. *Canadian Journal of Forest Research*, *36*(7), 1851-1870.

Federer, W. (1956). Augmented (or hoonuiaku) designs Hawaiian Planters Record.

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

Fisher, R. A. (1937). The design of experiments. *The design of experiments.*, (2nd Ed).

Gaston, L. A., Locke, M. A., Zablotowicz, R. M., & Reddy, K. N. (2001). Spatial variability of soil properties and weed populations in the Mississippi Delta. *Soil Science Society of America Journal*, *65*(2), 449-459.

Gezan, S. A., White, T. L., & Huber, D. A. (2010). Accounting for spatial variability in breeding trials: A simulation study. *Agronomy journal*, *102*(6), 1562-1571.

Gilmour, A. R., Cullis, B. R., & Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics*, 269-293.

Haghnazari, F., Shahgholi, H., & Feizi, M. (2015). Factors affecting the infiltration of agricultural soils. *International Journal of Agronomy and Agricultural Research*, *6*(5), 21-35.

Henry, C. G., Espinoza, L., Ismanov, M., Francis, P., & Ross, J. (2014). Scheduling Irrigation using an Atmometer (ET Gauge) for Arkansas Soybeans. https://www.uaex.edu/environment-nature/water/docs/irrig-ET-Gage-Sheet-Soybean.pdf

Hoefler, R., González-Barrios, P., Bhatta, M., Nunes, J. A., Berro, I., Nalin, R. S., ... & Gutierrez, L. (2020). Do spatial designs outperform classic experimental designs? *Journal of Agricultural, Biological and Environmental Statistics*, *25*(4), 523-552.

Hu, X., & Spilke, J. (2009). Comparison of various spatial models for the analysis of cultivar trials. *New Zealand Journal of Agricultural Research*, *52*(3), 277-287.

Kravchenko, A. N., Robertson, G. P., Snap, S. S., & Smucker, A. J. M. (2006). Using information about spatial variability to improve estimates of total soil carbon. *Agronomy Journal*, *98*(3), 823-829.

Kumar, A., Bharti, B., Kumar, J., Bhatia, D., Singh, G. P., Jaiswal, J. P., & Prasad, R. (2020). Improving the efficiency of wheat breeding experiments using alpha lattice design over randomised complete block design. *Cereal Research Communications*, *48*(1), 95-101.

Massey, J. H., Stiles, C. M., Epting, J. W., Powers, R. S., Kelly, D. B., Bowling, T. H., ... & Pennington, D. A. (2017). Long-term measurements of agronomic crop irrigation made in the Mississippi delta portion of the lower Mississippi River Valley. *Irrigation Science*, *35*(4), 297-313.

Maupin, M. A., Kenny, J. F., Hutson, S. S., Lovelace, J. K., Barber, N. L., & Linsey, K. S. (2014). *Estimated use of water in the United States in 2010* (No. 1405). US Geological Survey.

Mo, H., & Si, Y. (1986). Trend variation and it's control in field experiment. *Acta Agr Sinnica*, *12*(4), 233-240.

Morsy, A. R., & Fares, W. M. (2016). Utilization of alpha lattice design and trend analysis for controlling the experimental error in soybean variety trials. *Alex. J. Agric. Sci*, *61*(4), 399-407.

Müller, B. U., Kleinknecht, K., Möhring, J., & Piepho, H. P. (2010). Comparison of spatial models for sugar beet and barley trials. *Crop science*, *50*(3), 794-802.

Paget, M. F., Alspach, P. A., Anderson, J. A., Genet, R. A., & Apiolaza, L. A. (2015). Trial heterogeneity and variance models in the genetic evaluation of potato tuber yield. *Plant Breeding*, *134*(2), 203-211.

Patterson, H. D., Williams, E. R., & Hunter, E. A. (1978). Block designs for variety trials. *The Journal of Agricultural Science*, *90*(2), 395-400.

Piepho, H. P., Möhring, J., Pflugfelder, M., Hermann, W., & Williams, E. R. (2015). Problems in parameter estimation for power and AR (1) models of spatial correlation in designed field experiments. *Communications in Biometry & Crop Science*, *10*(1).

Piepho, H. P., Richter, C., & Williams, E. (2008). Nearest neighbour adjustment and linear variance models in plant breeding trials. *Biometrical Journal*, *50*(2), 164-189.

Qiao, C. G., Basford, K. E., DeLacy, I. H., & Cooper, M. (2000). Evaluation of experimental designs and spatial analyses in wheat breeding trials. *Theoretical and Applied Genetics*, *100*(1), 9-16.

Quintana-Ashwell, N., Gholson, D. M., Krutz, L. J., Henry, C. G., & Cooke, T. (2020). Adoption of water-conserving irrigation practices among row-crop growers in Mississippi, USA. *Agronomy*, *10*(8), 1083.

Sarker, A., Singh, M., & Erskine, W. (2001). Efficiency of spatial methods in yield trials in lentil (Lens culinaris ssp. culinaris). *The Journal of Agricultural Science*, *137*(4), 427-438.

Slaton, N., Roberts, T., & Ross, J. (2013). Fertilization and liming practices. *Arkansas Soybean Production Handbook, University of Arkansas Research and Extension, Little Rock. http://www. uaex. edu/publications/pdf/mp197/chapter5. pdf.*

Stroup, W. W. (2002). Power analysis based on spatial effects mixed models: A tool for comparing design and analysis strategies in the presence of spatial variability. *Journal of Agricultural, Biological, and Environmental Statistics*, *7*(4), 491-511.

Stroup, W. W., Baenziger, P. S., & Mulitze, D. K. (1994). Removing spatial variation from wheat yield trials: a comparison of methods. *Crop Science*, *34*(1), 62-66.

Tacker, Phil, Earl Vories, and Gary Huitink. "Drainage and irrigation." *Corn production handbook* (2006): 95.

Tola, E., Al-Gaadi, K. A., Madugundu, R., Zeyada, A. M., Kayad, A. G., & Biradar, C. M. (2017). Characterization of spatial variability of soil physicochemical properties and its impact on Rhodes grass productivity. *Saudi journal of biological sciences*, *24*(2), 421-429.

United States. (2021) *U.S. National Agricultural Statistics Service NASS*. United States. Retrieved October, 2021 from https://www.nass.usda.gov/Quick_Stats/Lite/result.php?38CD054E-2CD5-3E63-BC3F-58E9C127C12D.

Wibawa, W. D., Dludlu, D. L., Swenson, L. J., Hopkins, D. G., & Dahnke, W. C. (1993). Variable fertilizer application based on yield goal, soil fertility, and soil map unit. *Journal of Production Agriculture*, *6*(2), 255-261.

Yang, R. C., Ye, T. Z., Blade, S. F., & Bandara, M. (2004). Efficiency of spatial analyses of field pea variety trials. *Crop Science*, *44*(1), 49-55.

Zaman, M., Shahid, S. A., & Heng, L. (2018). Irrigation systems and zones of salinity development. In *Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques* (pp. 91-111). Springer, Cham.

Table 5.1. Comparison of three different models (block model, exponential model, and gaussian model) in four environments (19ROH, 19STU, 20ROH, and 20STU) in terms of AIC (Akaike's Information Criteria), BIC (Bayesian Information Criteria), and significant of genotype effects, irrigation effects, and the interaction of the irrigation and genotype effect on soybean wilting.

Model	**Environment**	**AIC**	**BIC**	***p*-value Genotype**	***p*-value Irrigation**	***p*-value Irrigation*Genotype**
Block	19ROH	307.1	307.7	<0.0001	<0.0001	0.0349
	19STU	318.6	319.2	0.0006	<0.0001	0.0038
	20ROH	246.8	247.7	<0.0001	<0.0001	<0.0001
	20STU	206.6	207.2	<0.0001	<0.0001	<0.0001
Exponential	19ROH	305.3	297.3	0.0515	0.0013	0.5207
	19STU	316.1	308.1	0.1157	<0.0001	0.5648
	20ROH	247.1	243.1	<0.0001	<0.0001	<0.0001
	20STU	206.3	202.3	<0.0001	<0.0001	<0.0001
Gaussian	19ROH	306	298	-	-	0.9729
	19STU	316	308	0.0622	<0.0001	0.5458
	20ROH	247.1	243.1	<0.0001	<0.0001	<0.0001
	20STU	206.3	202.3	<0.0001	<0.0001	<0.0001

(-) the convergence criteria were not met and the model could not be run

Table 5.2. Comparison of three different models (block model, exponential model, and gaussian model) in four environments (19ROH, 19STU, 20ROH, and 20STU) in terms of AIC (Akaike's Information Criteria), BIC (Bayesian Information Criteria), and significant of genotype effects, irrigation effects, and the interaction of the irrigation and genotype effect on soybean maturity.

Model	**Environment**	**AIC**	**BIC**	***p*-value Genotype**	***p*-value Irrigation**	***p*-value Irrigation*Genotype**
Augmented	19ROH	621.2	621.5	<0.0001	<0.0001	0.0009
	19STU	830.1	830.7	0.0056	<0.0001	0.8032
	20ROH	355.7	356.4	<0.0001	<0.0001	<0.0001
	20STU	504.1	504.7	<0.0001	<0.0001	0.1612
Exponential	19ROH	622.7	618.7	<0.0001	<0.0001	0.0007
	19STU	381	823	.	.	0.9909
	20ROH	350.2	346.2	<0.0001	<0.0001	<0.0001
	20STU	503.8	497.8	<0.0001	<0.0001	0.6087
Gaussian	19ROH	624.7	618.7	.	.	0.6209
	19STU	830.8	822.8	0.0656	0.0039	0.6973
	20ROH	352.1	346.1	0.004	<0.0001	0.0037
	20STU	505.8	497.8	.	<0.0001	0.6124

Table 5.3. Comparison of three different models (block model, exponential model, and gaussian model) in four different environments (19ROH, 19STU, 20ROH, and 20STU) in terms of AIC (Akaike's Information Criteria), BIC (Bayesian Information Criteria), and significant of genotype effects, irrigation effects and the interaction of the irrigation and genotype effect on soybean seed yield.

Model	**Environment**	**AIC**	**BIC**	***p*-value Genotype**	***p*-value Irrigation**	***p*-value Irrigation*Genotype**
Block	19ROH	2113.1	2113.7	0.6196	<0.0001	0.1332
	19STU	2269.3	2269.9	0.3968	<0.0001	0.6875
	20ROH	2034.9	2035.6	<0.0001	<0.0001	0.1157
	20STU	1751.2	1751.9	<0.0001	<0.0001	0.9971
Exponential	19ROH	2094.5	2088.5	0.0027	0.0132	0.0988
	19STU	2240.3	2234	0.0007	0.1361	0.0757
	20ROH	2037.1	2031.1	0.0005	0.0046	0.6576
	20STU	1732.6	1724.6	0.0002	0.014	0.0545
Gaussian	19ROH	2113.2	2105.2	0.0577	<0.0001	0.1458
	19STU	2254.4	2246.4	0.4485	<0.0001	0.3969
	20ROH	2039.1	2031.1	0.0005	0.0046	0.6576
	20STU	1738.9	1730.9	0.1142	<0.0001	0.2118

Table 5.4. Comparison of three different models (block model, exponential model, and gaussian model) for the 2019 and 2020 Final Yield Trials (19AF4E and 20AF4E) grown in four environments in terms of AIC (Akaike's Information Criteria), BIC (Bayesian Information Criteria), and significant of genotype effects on soybean seed yield.

Yield Trials	Models	Environment	AIC	BIC	*p*-value genotype
19AF4E	Block	19KEI	468.2	466.9	0.0402
		19PTR	495.2	492.6	<.0001
		19ROH	582.3	579.7	0.0165
		19STU	666.6	664	<.0001
	Exponential	19KEI	469.3	465.3	0.069
		19PTR	494.7	490.7	<.0001
		19ROH	578.2	574.2	0.0026
		19STU	664	658	<.0001
	Gaussian	19KEI	471.3	465.3	0.0615
		19PTR	496.3	490.3	0.0051
		19ROH	579.8	575.8	0.0637
		19STU	666	658	<.0001
20AF4E	Block	20ROH	290.2	288.9	0.0008
		20SHO	333.2	331.9	<.0001
	Exponential	20ROH	292.7	286.7	0.0726
		20SHO	331.5	325.5	<.0001
	Gaussian	20ROH	289.9	283.9	0.129
		20SHO	331.5	325.5	<.0001

Table 5.5. Heritability for seed yield for block model and exponential model for each environment for the 2019 Arkansas Final yield trials maturity Group IV Early (19AF4E) and 2020 Arkansas Final yield trials maturity group IV Early (20AF4E)

Environment	Block Model Heritability	Exponential Model Heritability
19KEI	0.49	0.52
19PTR	0.51	0.52
19ROH	0.48	0.52
19STU	0.77	0.84
20ROH	0.70	0.81
20SHO	0.85	0.87

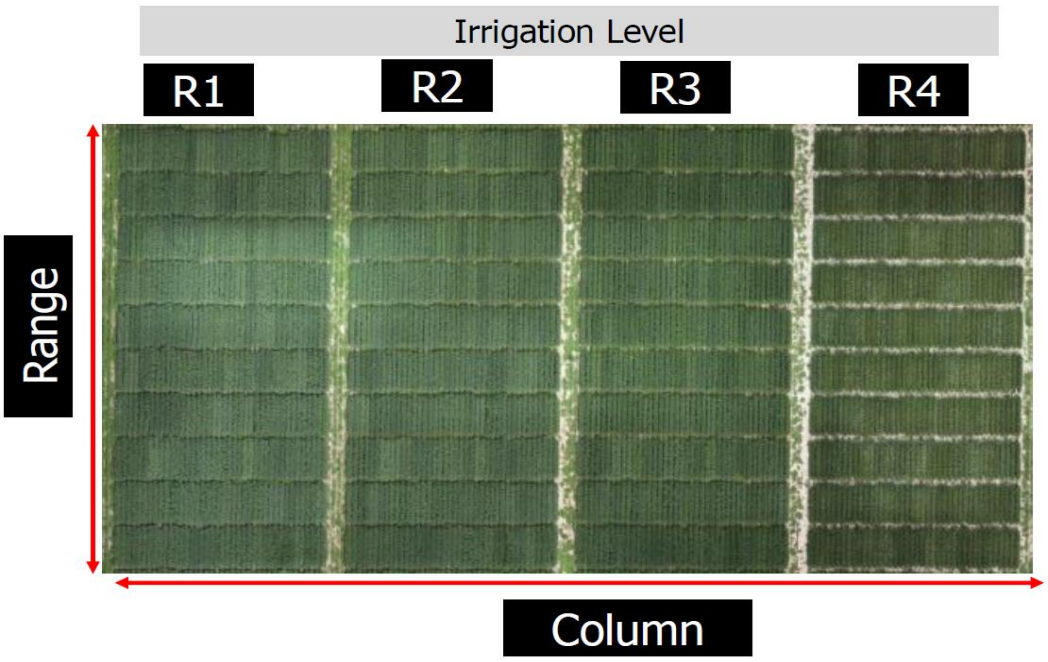

Figure 5.1. Aerial image of the layout of the experimental design in the 20STU environment. Each strip represents a different irrigation level (R1, R2, R3, R4 growth stages). Each irrigation level is separated with four rows of borders.

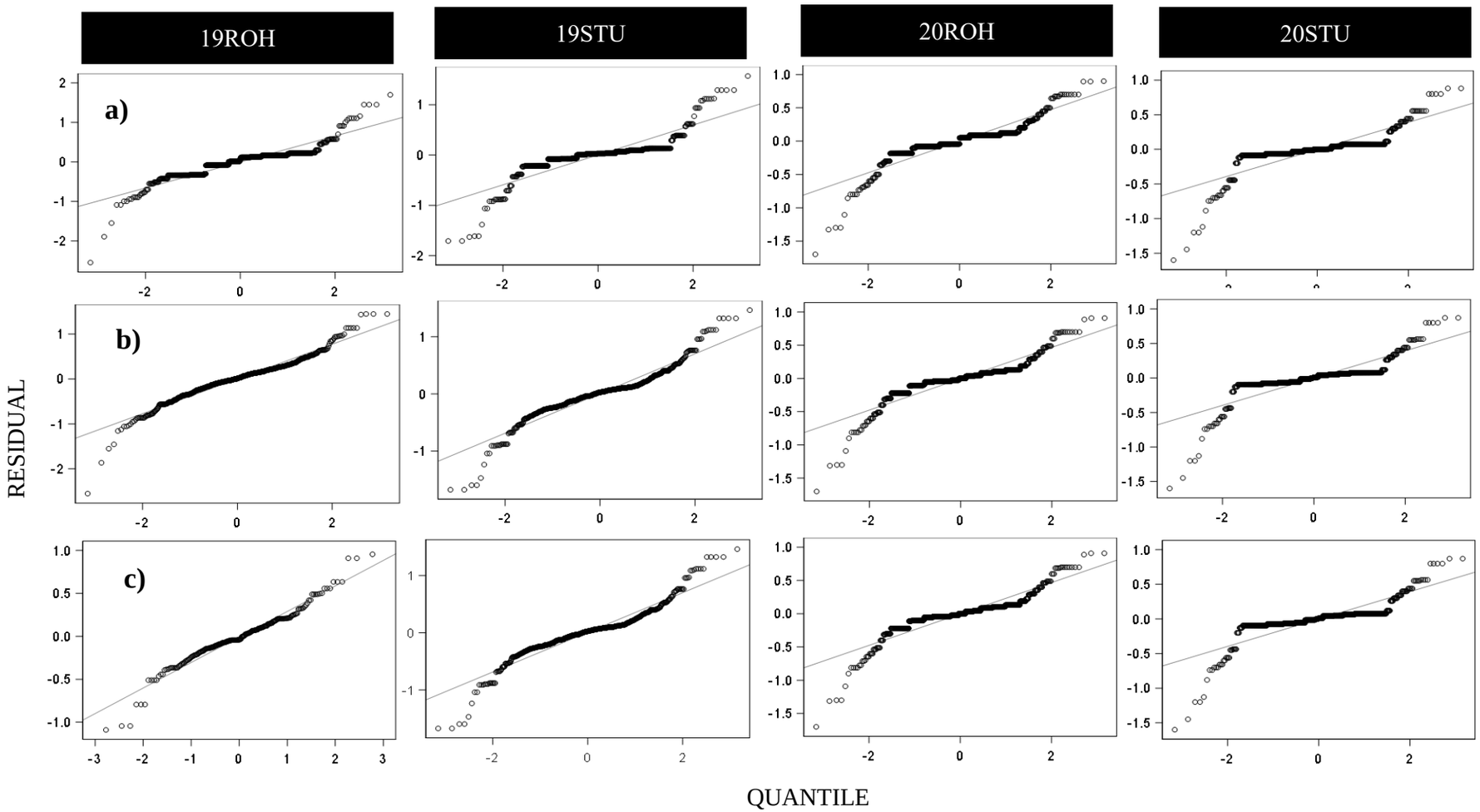

Figure 5.2. QQ plots for block model, exponential model, and gaussian model in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean wilting.

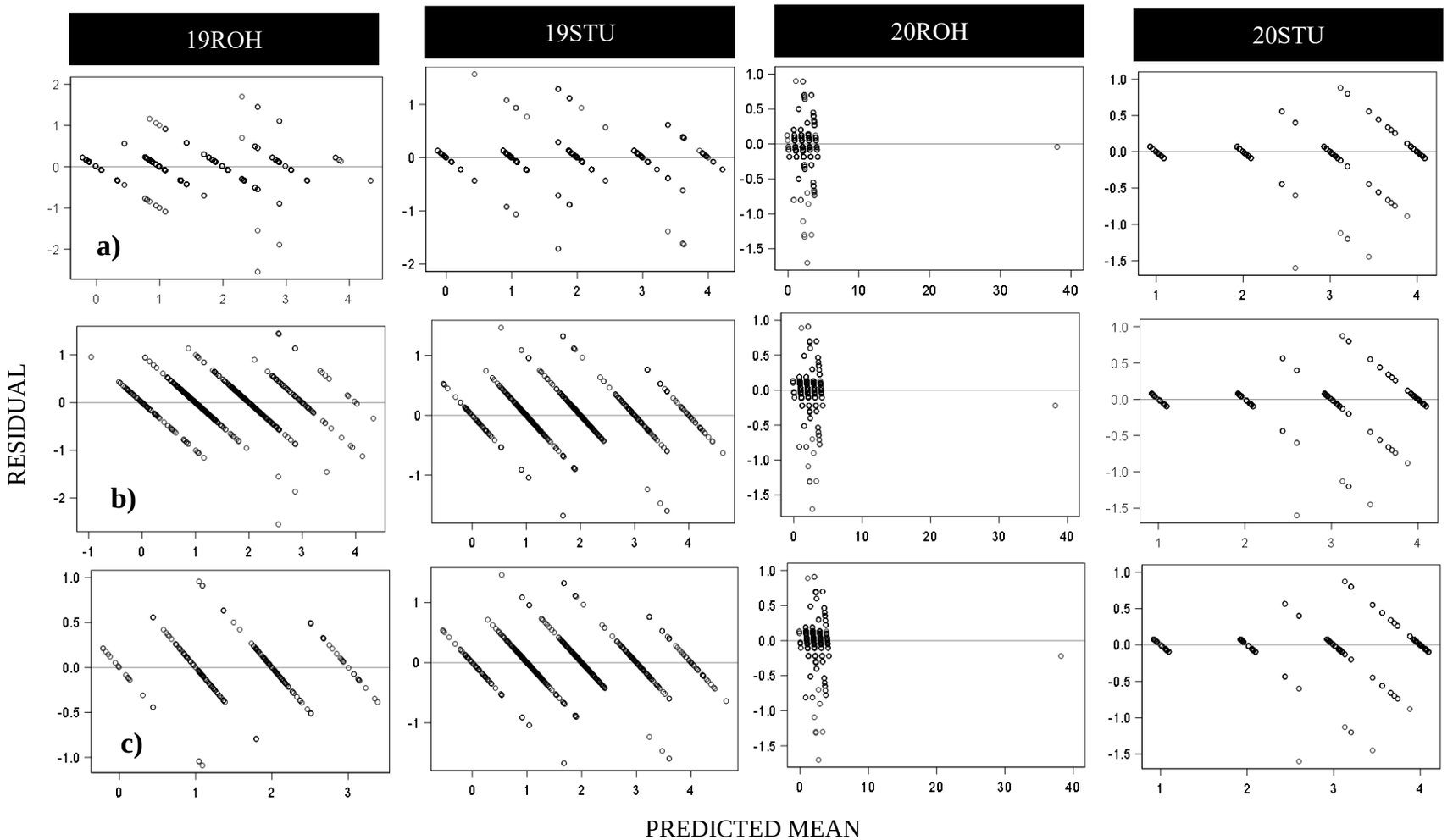

Figure 5.3. Student residuals for block model, exponential model, and gaussian model in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean wilting.

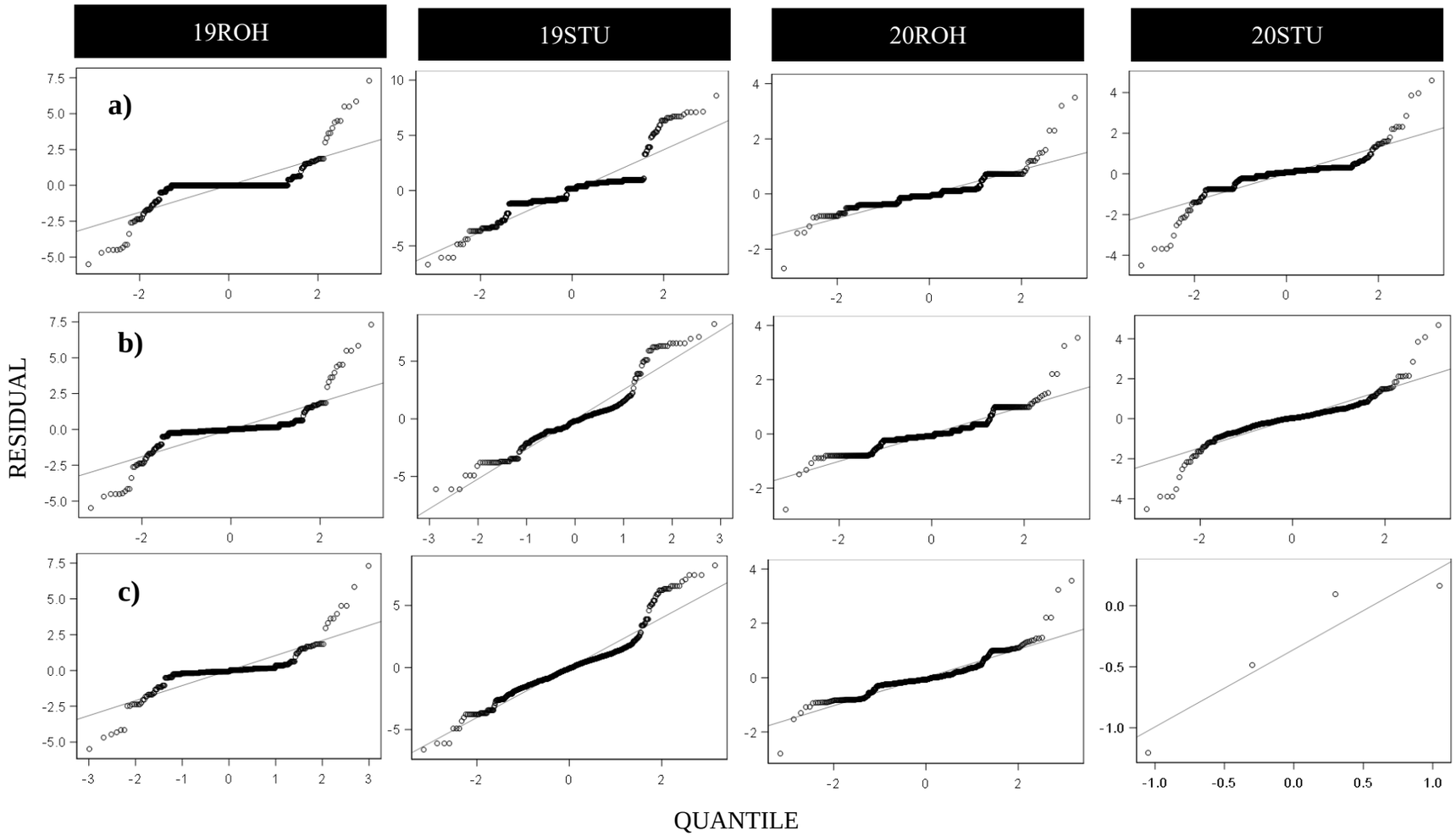

Figure 5.4. QQ plots for block model, exponential model, and gaussian model in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean maturity.

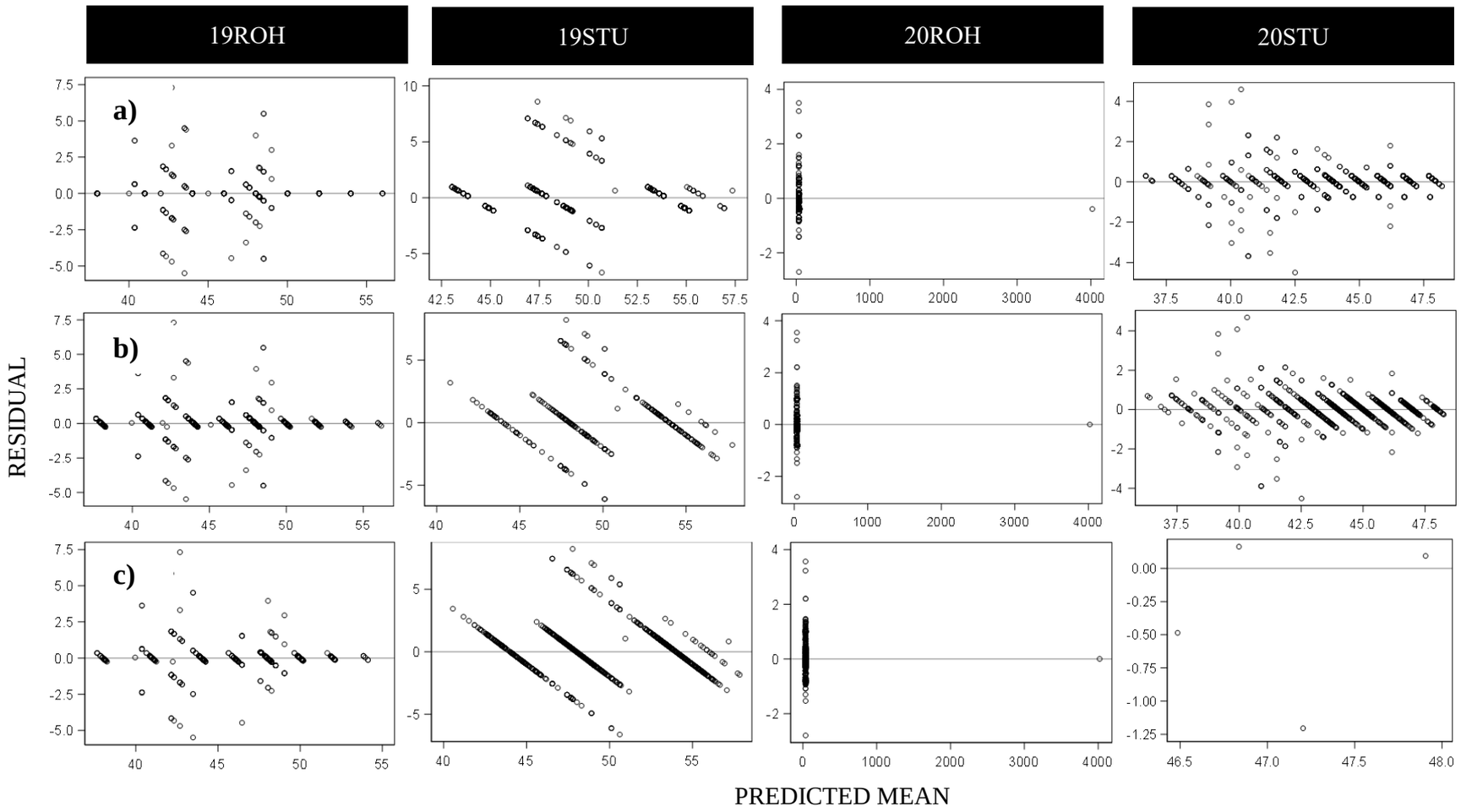

Figure 5.5. Student residuals for linear mixed model with blocking structure (a), exponential model (b), and gaussian model (c) in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean maturity.

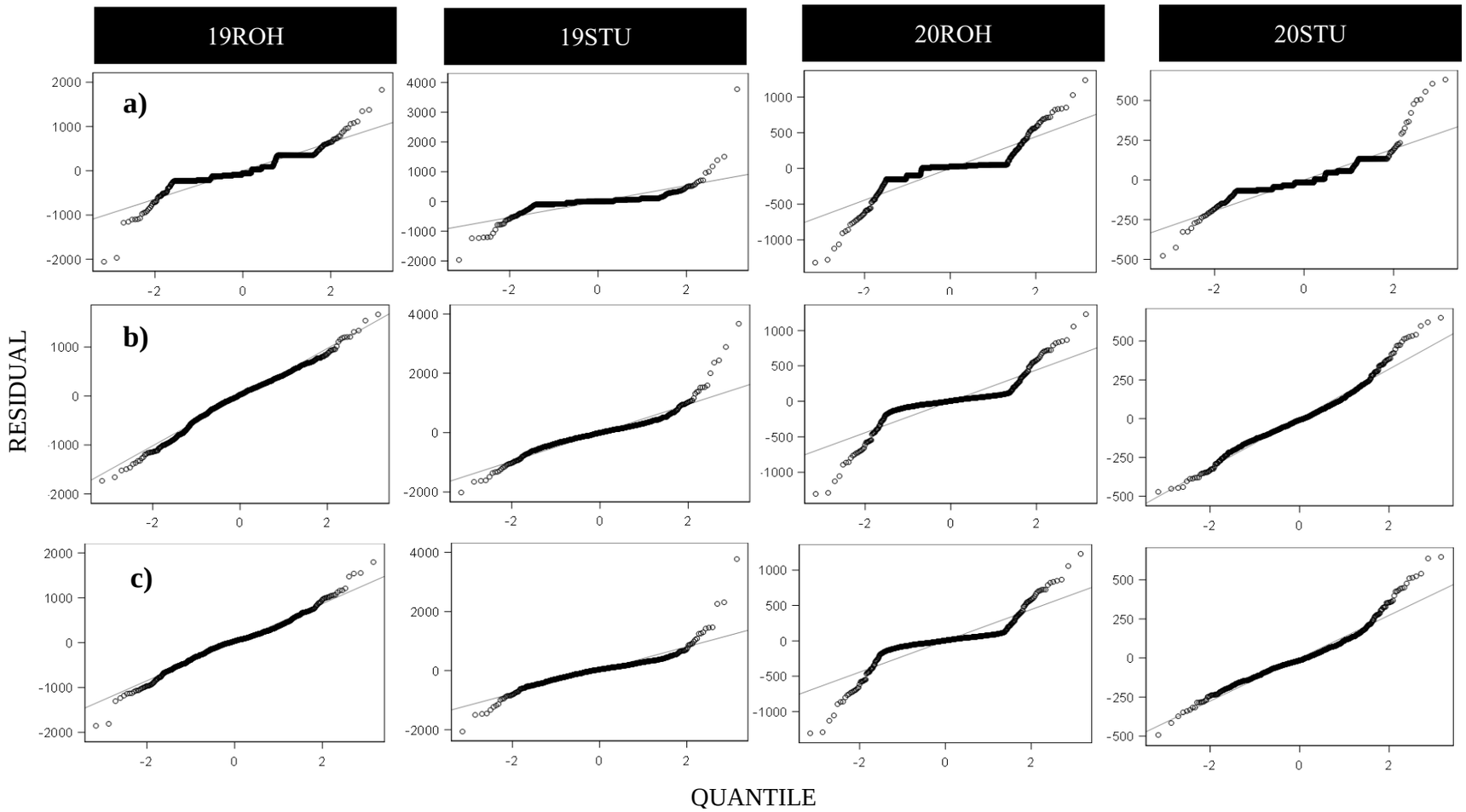

Figure 5.6. QQ plots linear mixed model with blocking structure (a), exponential model (b), and gaussian model (c) in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean seed yield.

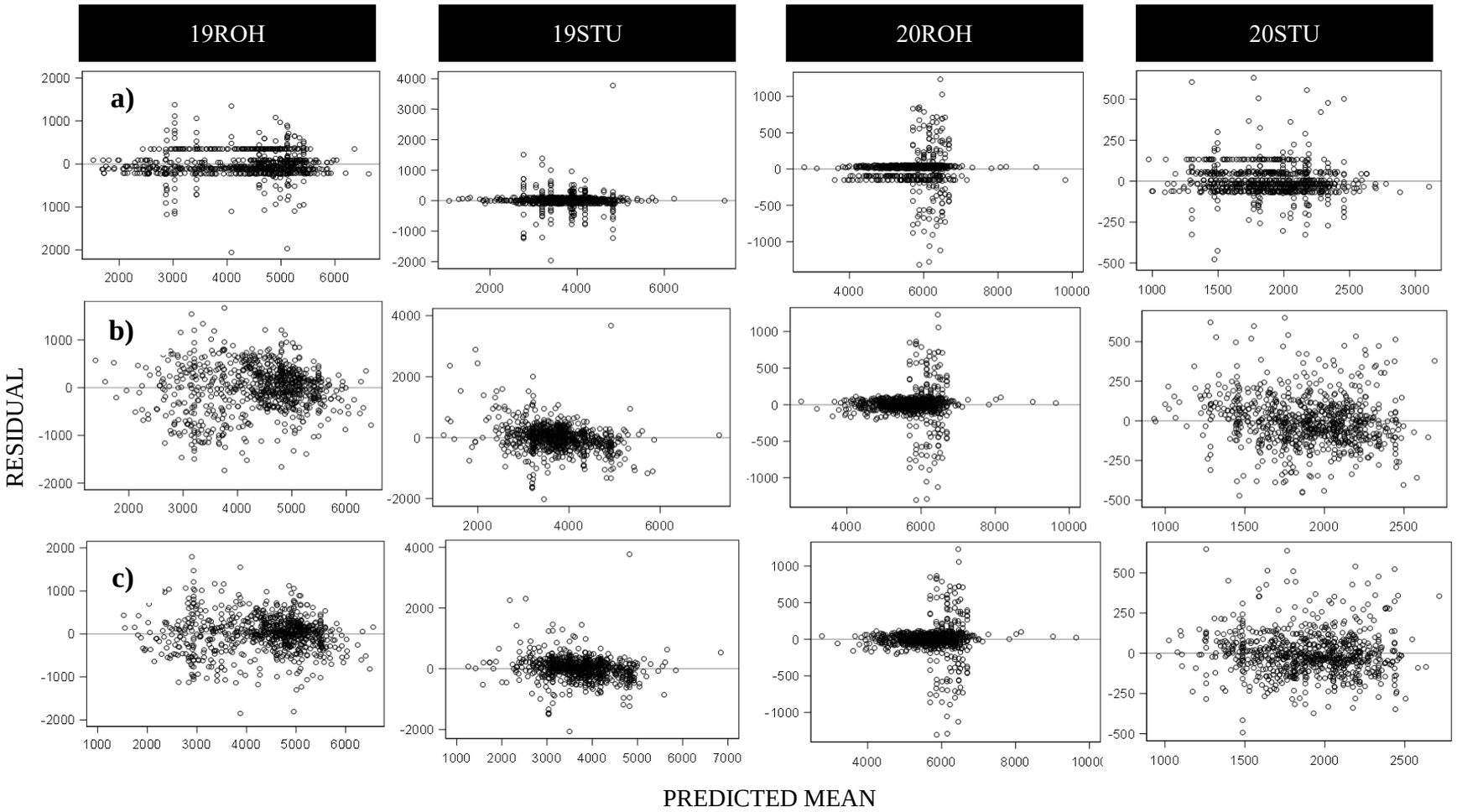

Figure 5.7. Student residuals linear mixed model with blocking structure (a), exponential model (b), and gaussian model (c) in each environment 19ROH, 19STU, 20ROH, and 20STU for data on soybean seed yield.

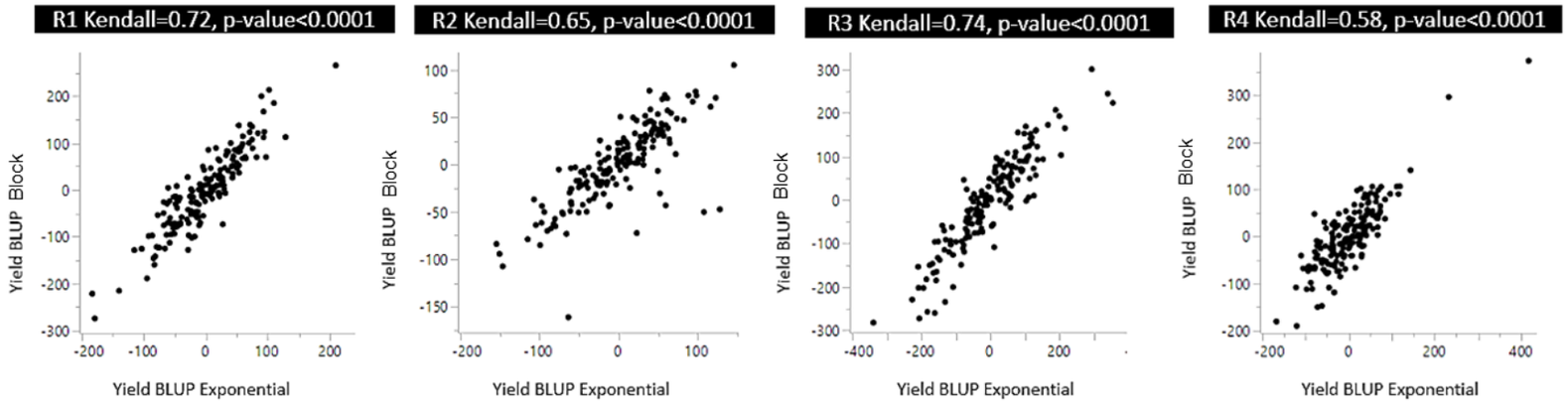

Figure 5.8. Kendall correlation for BLUP yield in each irrigation level at R1, R2, R3, and R4 growth stages between the block model and the exponential model, and p-value of the significance of the Kendall correlation

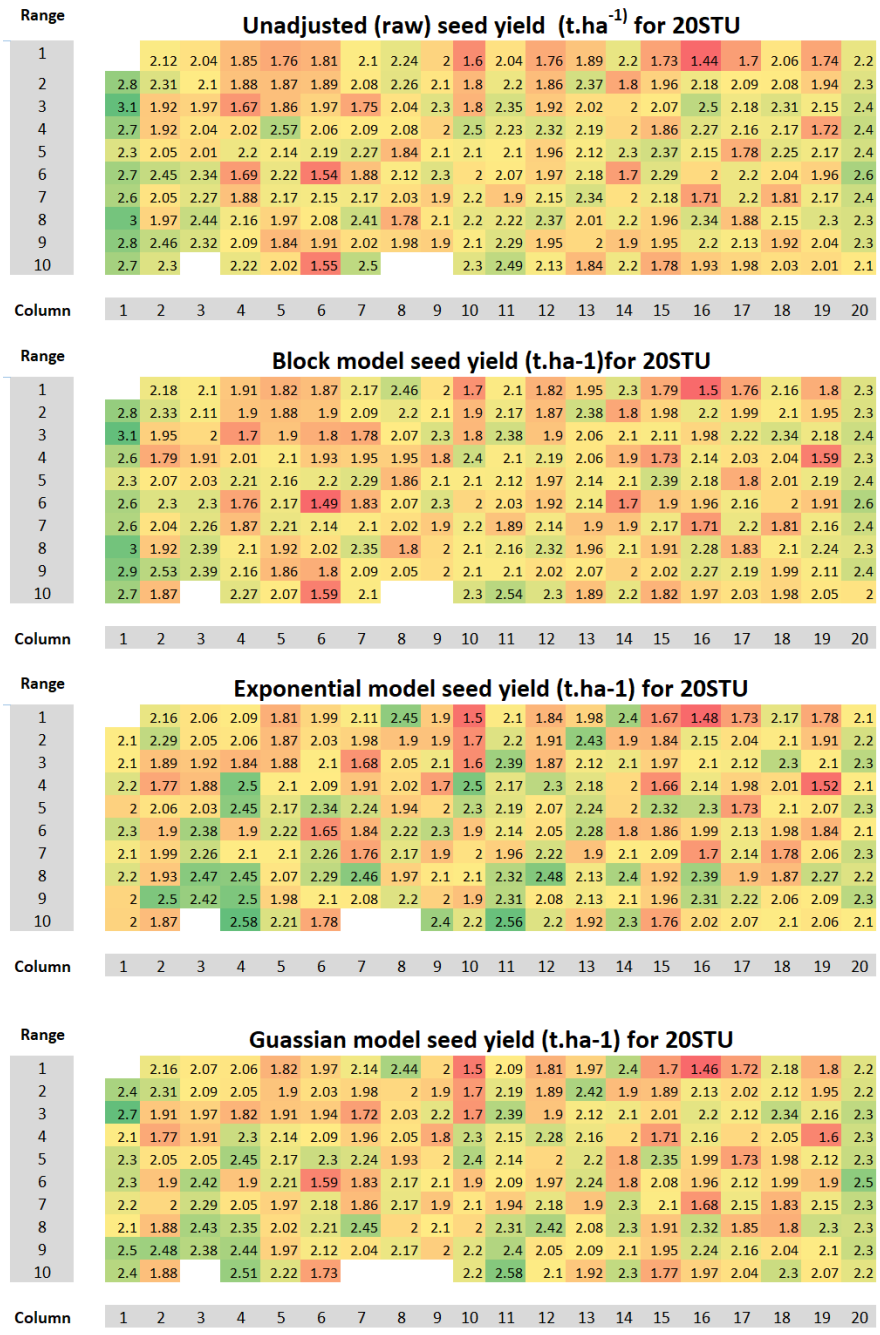

Figure 5.9: Range/column exemplary plot of spatial trends for unadjusted (raw) seed yield, and seed yield analyzed using block, exponential, and gaussian models for the R1-irrigation treatment in the 20STU environment

Table S 5.1. Yield BLUPs for each irrigation level and ranking of soybean genotype for the block model and the exponential model grown at four Arkansas environments.

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R1	AG55X7	41.4	40	50.0	43
R1	AG56X8	84.0	10	121.4	11
R1	P53AG7X	-5.2	93	-4.2	87
R1	P55A49X	53.0	30	81.7	29
R1	R11-1057	43.4	37	65.7	38
R1	R11-2933	10.6	72	6.0	76
R1	R18-292	32.7	45	74.3	30
R1	R18-293	-20.6	112	-14.6	103
R1	R18-294	-57.8	145	-59.7	133
R1	R18-295	-92.4	163	-98.5	153
R1	R18-296	-0.4	86	-8.9	96
R1	R18-297	-41.4	128	-73.4	143
R1	R18-298	-83.2	160	-159.4	165
R1	R18-299	8.5	75	23.3	61
R1	R18-300	-179.4	168	-274.1	169
R1	R18-301	-68.4	155	-49.3	125
R1	R18-302	-1.6	87	-23.6	111
R1	R18-303	20.0	65	-6.1	90
R1	R18-304	7.5	79	2.9	79
R1	R18-305	-5.7	95	-15.5	105
R1	R18-306	-23.3	116	-45.0	122
R1	R18-307	22.0	60	-13.2	100
R1	R18-308	20.2	64	65.7	37
R1	R18-309	6.1	81	-11.5	99
R1	R18-310	-103.6	165	-125.5	160
R1	R18-311	-62.6	150	-29.1	117
R1	R18-312	-7.0	96	0.6	84
R1	R18-313	0.7	84	-8.0	94
R1	R18-314	-45.9	131	-21.3	108
R1	R18-315	81.7	11	69.9	34
R1	R18-316	-44.2	130	-23.0	109
R1	R18-317	-8.6	98	5.2	77
R1	R18-318	-81.9	159	-141.5	163
R1	R18-319	-49.1	138	-29.6	118
R1	R18-322	-27.7	120	-96.9	151

Table S 5.2 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R1	R18-323	20.7	63	24.7	59
R1	R18-324	-1.7	88	10.3	70
R1	R18-325	53.4	27	89.2	22
R1	R18-326	-18.2	111	-99.5	154
R1	R18-327	-52.1	141	-14.7	104
R1	R18-328	-9.1	99	-69.6	135
R1	R18-329	-28.7	122	-127.9	161
R1	R18-330	-77.7	157	-53.6	128
R1	R18-331	-115.4	166	-128.4	162
R1	R18-332	-67.9	154	-75.4	144
R1	R18-333	-52.7	142	-81.9	147
R1	R18-334	-5.3	94	-8.2	95
R1	R18-335	4.0	83	12.7	67
R1	R18-336	15.2	68	6.6	75
R1	R18-337	-63.4	151	-95.9	150
R1	R18-338	-24.2	118	-51.9	126
R1	R18-339	27.3	50	-73.2	142
R1	R18-340	-79.5	158	-121.5	157
R1	R18-341	29.0	49	10.0	71
R1	R18-342	8.2	77	23.1	62
R1	R18-343	22.4	58	33.0	54
R1	R18-344	-39.1	126	-57.9	131
R1	R18-345	-61.5	149	-70.2	137
R1	R18-346	11.7	70	0.8	82
R1	R18-347	-14.5	106	-58.4	132
R1	R18-348	46.7	34	85.8	23
R1	R18-349	23.4	56	42.8	48
R1	R18-350	-58.4	146	-112.9	156
R1	R18-351	-4.5	91	-55.1	129
R1	R18-352	-5.0	92	49.8	44
R1	R18-353	-95.1	164	-188.6	166
R1	R18-354	-15.5	107	-46.0	123
R1	R18-355	-46.6	134	-78.6	146
R1	R18-356	-86.7	162	-97.4	152
R1	R18-357	-50.5	139	-71.0	139
R1	R18-358	29.6	48	1.6	81
R1	R18-359	-15.7	109	-25.9	115

Table S 5.3 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R1	R18-362	-46.5	133	-23.2	110
R1	R18-363	21.0	62	-10.8	98
R1	R18-364	16.3	67	89.5	20
R1	R18-365	-23.9	117	1.7	80
R1	R18-366	-56.4	144	-94.9	149
R1	R18-367	53.0	29	137.8	7
R1	R18-368	-28.7	123	-41.9	121
R1	R18-7389	74.4	14	135.2	8
R1	R18-7390	-10.9	101	-75.6	145
R1	R18-7391	71.2	15	139.2	6
R1	R18-7392	8.5	76	-4.7	89
R1	R18-7393	63.3	19	70.7	32
R1	R18-7394	50.1	32	46.3	47
R1	R18-7395	70.7	16	123.6	10
R1	R18-7396	-75.3	156	-123.3	158
R1	R18-7397	7.7	78	46.5	46
R1	R18-7398	-47.4	136	-73.2	141
R1	R18-7399	44.3	35	72.3	31
R1	R18-7402	-15.5	108	-28.1	116
R1	R18-7403	25.2	55	19.8	65
R1	R18-7404	-13.0	105	-46.2	124
R1	R18-7405	-7.2	97	-10.8	97
R1	R18-7406	14.3	69	31.6	55
R1	R18-7407	63.9	18	67.1	36
R1	R18-7408	22.3	59	10.5	69
R1	R18-7409	74.4	13	107.8	15
R1	R18-7410	59.2	21	98.6	18
R1	R18-7411	0.3	85	23.9	60
R1	R18-7412	51.8	31	82.4	27
R1	R18-7413	-67.9	153	-25.7	113
R1	R18-7414	75.2	12	89.2	21
R1	R18-7415	-11.4	103	-57.6	130
R1	R18-7416	57.3	23	92.3	19
R1	R18-7417	36.0	43	41.0	50
R1	R18-7418	22.4	57	62.6	39
R1	R18-7419	94.3	6	124.3	9
R1	R18-7420	93.1	7	167.2	5

Table S 5.4 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R1	R18-7421	42.9	39	26.1	58
R1	R18-7422	128.6	2	113.0	13
R1	R18-7423	-140.0	167	-215.1	167
R1	R18-7424	36.5	42	82.2	28
R1	R18-7425	55.1	24	82.9	26
R1	R18-7426	27.2	51	15.6	66
R1	R18-7427	109.9	3	185.4	4
R1	R18-7428	-30.5	125	-62.6	134
R1	R18-7429	-22.9	114	-102.0	155
R1	R18-7430	-47.0	135	8.3	73
R1	R18-7431	-46.3	132	-38.2	120
R1	R18-7432	-22.1	113	-2.3	85
R1	R18-7433	-55.5	143	-72.6	140
R1	R18-7434	17.3	66	9.8	72
R1	R18-7435	-42.0	129	-16.5	106
R1	R18-7436	-9.7	100	12.6	68
R1	R18-7437	-66.0	152	-124.7	159
R1	R18-7438	43.3	38	52.0	42
R1	R18-7439	9.5	74	-3.9	86
R1	R18-7442	-29.4	124	27.2	56
R1	R18-7443	31.4	46	-6.8	92
R1	R18-7444	4.0	82	85.8	24
R1	R18-7445	-3.8	90	-7.8	93
R1	R18-7446	-28.0	121	-70.8	138
R1	R18-7447	-60.9	148	4.8	78
R1	R18-7448	43.5	36	34.8	53
R1	R18-7449	26.8	52	20.1	64
R1	R18-7450	-50.8	140	-93.5	148
R1	R18-7451	40.2	41	100.3	17
R1	R18-7452	58.8	22	67.2	35
R1	R18-7453	53.6	25	35.8	52
R1	R18-7454	34.0	44	83.5	25
R1	R18-7455	-58.5	147	-23.7	112
R1	R18-7456	49.7	33	58.1	40
R1	R18-7457	-183.5	169	-221.4	168
R1	R18-7458	-23.1	115	-25.8	114
R1	R18-7459	-17.8	110	-53.1	127

Table S 5.5 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R1	R18-7460	59.7	20	114.8	12
R1	R18-7461	30.6	47	47.1	45
R1	R18-7462	-48.3	137	-14.1	102
R1	R18-7463	209.7	1	266.0	1
R1	R18-7464	-11.8	104	0.6	83
R1	R18-7465	21.8	61	-6.4	91
R1	R18-7466	7.4	80	-4.3	88
R1	R18-7467	-3.0	89	42.6	49
R1	R18-7468	53.4	28	52.8	41
R1	R18-7469	26.8	53	26.2	57
R1	R18-7470	89.7	9	200.3	3
R1	R18-7471	69.3	17	100.5	16
R1	R18-7472	-25.5	119	-13.6	101
R1	R18-7473	102.1	4	213.8	2
R1	R18-7475	25.5	54	20.7	63
R1	R18-7476	-11.3	102	-30.5	119
R1	R18-7477	93.0	8	112.8	14
R1	R18-7478	-84.6	161	-145.5	164
R1	R18-7479	97.0	5	70.2	33
R1	R18-7482	10.1	73	-18.1	107
R1	R18-7483	53.5	26	36.6	51
R1	R18-7484	-41.2	127	-70.2	136
R1	R18-7486	11.1	71	6.8	74
R2	AG55X7	41.7	40	31.8	37
R2	AG56X8	74.0	11	48.5	19
R2	P53AG7X	-1.2	93	1.8	87
R2	P55A49X	60.1	19	41.5	27
R2	R11-1057	34.0	47	30.4	41
R2	R11-2933	44.2	37	25.6	47
R2	R18-292	50.2	30	35.0	32
R2	R18-293	-61.7	149	-29.7	133
R2	R18-294	26.8	57	19.0	57
R2	R18-295	-104.3	164	-63.9	158
R2	R18-296	-40.2	130	-49.9	149
R2	R18-297	-18.6	110	-31.0	135
R2	R18-298	-47.4	136	-7.7	103
R2	R18-299	49.6	33	30.9	39

Table S 5.6 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R2	R18-300	-54.6	142	-3.0	92
R2	R18-301	-22.4	113	-18.1	114
R2	R18-302	-70.9	153	-50.7	152
R2	R18-303	-14.1	104	-9.8	105
R2	R18-304	54.2	28	38.4	29
R2	R18-305	-31.1	124	-35.6	137
R2	R18-306	-63.2	150	-161.3	169
R2	R18-307	52.5	29	-30.6	134
R2	R18-308	-99.0	163	-85.2	166
R2	R18-309	6.6	81	16.6	63
R2	R18-310	-59.8	145	-39.6	140
R2	R18-311	9.6	78	24.0	50
R2	R18-312	-96.7	161	-43.9	145
R2	R18-313	-38.6	128	-24.6	129
R2	R18-314	-59.9	146	-25.6	131
R2	R18-315	27.7	56	23.9	51
R2	R18-316	-22.6	114	2.0	84
R2	R18-317	21.0	64	1.8	85
R2	R18-318	-85.2	158	-67.7	160
R2	R18-319	-81.8	157	-62.6	157
R2	R18-322	23.3	62	-72.5	162
R2	R18-323	-50.3	140	-50.7	153
R2	R18-324	-93.7	160	-50.2	150
R2	R18-325	33.9	49	43.1	26
R2	R18-326	-97.3	162	-61.6	156
R2	R18-327	-50.2	139	-24.0	128
R2	R18-328	-150.4	168	-94.4	167
R2	R18-329	-0.8	91	22.9	52
R2	R18-330	-37.4	127	-3.2	93
R2	R18-331	-42.7	132	-20.5	121
R2	R18-332	-7.0	96	-14.4	110
R2	R18-333	-25.2	118	10.9	69
R2	R18-334	-59.3	144	-42.2	141
R2	R18-335	25.2	60	6.0	75
R2	R18-336	123.1	3	70.4	7
R2	R18-337	-154.2	169	-84.0	165
R2	R18-338	-47.3	135	-22.4	127

Table S 5.7 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R2	R18-339	44.5	36	38.3	31
R2	R18-340	-79.4	155	-65.3	159
R2	R18-341	0.0	89	12.1	66
R2	R18-342	2.5	86	50.3	17
R2	R18-343	46.9	34	38.4	30
R2	R18-344	-25.9	119	-21.4	125
R2	R18-345	-146.2	167	-107.5	168
R2	R18-346	-28.4	123	-31.5	136
R2	R18-347	7.7	79	0.0	88
R2	R18-348	-45.1	134	-42.6	142
R2	R18-349	18.7	66	49.4	18
R2	R18-350	-106.7	165	-36.9	138
R2	R18-351	23.1	63	46.5	21
R2	R18-352	42.5	39	34.9	33
R2	R18-353	-11.9	100	-44.0	146
R2	R18-354	-12.5	101	7.6	72
R2	R18-355	14.8	67	-24.8	130
R2	R18-356	-22.2	112	-19.5	120
R2	R18-357	-36.9	126	-38.0	139
R2	R18-358	-19.4	111	-18.5	118
R2	R18-359	-42.4	131	-11.3	108
R2	R18-362	33.1	50	24.3	49
R2	R18-363	54.3	26	43.2	25
R2	R18-364	6.0	82	7.2	73
R2	R18-365	-11.1	97	-6.8	102
R2	R18-366	-65.6	151	-73.2	163
R2	R18-367	57.5	22	30.1	42
R2	R18-368	-89.9	159	-70.2	161
R2	R18-7389	-23.2	117	25.5	48
R2	R18-7390	-26.5	120	-18.1	115
R2	R18-7391	-0.5	90	15.7	64
R2	R18-7392	54.2	27	32.9	35
R2	R18-7393	31.6	53	-0.3	89
R2	R18-7394	55.5	25	69.1	9
R2	R18-7395	40.4	42	58.1	12
R2	R18-7396	65.0	14	22.5	54
R2	R18-7397	-11.6	99	-4.4	95

Table S 5.8 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R2	R18-7398	59.8	20	-43.2	144
R2	R18-7399	-16.9	108	5.1	79
R2	R18-7402	-2.9	95	-20.9	122
R2	R18-7403	-16.8	106	-6.8	101
R2	R18-7404	9.8	77	-1.5	91
R2	R18-7405	-22.9	115	-22.2	126
R2	R18-7406	-59.9	147	-3.8	94
R2	R18-7407	49.8	31	-6.7	99
R2	R18-7408	-11.1	98	-42.8	143
R2	R18-7409	35.6	45	28.8	44
R2	R18-7410	2.2	87	-5.5	98
R2	R18-7411	39.3	43	77.9	2
R2	R18-7412	49.8	32	53.1	15
R2	R18-7413	5.1	83	3.1	82
R2	R18-7414	3.4	84	22.8	53
R2	R18-7415	56.5	23	30.7	40
R2	R18-7416	82.8	10	46.8	20
R2	R18-7417	-49.2	138	-16.9	112
R2	R18-7418	-48.8	137	-21.2	123
R2	R18-7419	21.0	65	32.8	36
R2	R18-7420	116.6	4	60.9	11
R2	R18-7421	-12.6	102	17.4	60
R2	R18-7422	146.0	1	105.1	1
R2	R18-7423	-60.7	148	-49.2	148
R2	R18-7424	6.6	80	27.6	45
R2	R18-7425	62.8	16	38.4	28
R2	R18-7426	14.6	69	10.6	70
R2	R18-7427	11.8	73	18.1	59
R2	R18-7428	14.6	68	2.2	83
R2	R18-7429	11.5	74	21.7	55
R2	R18-7430	37.8	44	45.3	23
R2	R18-7431	-28.2	122	-13.8	109
R2	R18-7432	93.9	8	66.4	10
R2	R18-7433	10.3	75	10.0	71
R2	R18-7434	98.6	6	72.8	5
R2	R18-7435	56.2	24	34.7	34
R2	R18-7436	25.3	59	21.0	56

Table S 5.9 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R2	R18-7437	-18.5	109	-9.8	106
R2	R18-7438	61.8	18	69.8	8
R2	R18-7439	43.3	38	18.4	58
R2	R18-7442	-15.0	105	-21.3	124
R2	R18-7443	-14.0	103	-18.2	117
R2	R18-7444	34.1	46	31.3	38
R2	R18-7445	-59.2	143	-25.7	132
R2	R18-7446	-75.3	154	-5.3	97
R2	R18-7447	33.9	48	51.3	16
R2	R18-7448	97.7	7	77.0	3
R2	R18-7449	-40.0	129	-18.1	116
R2	R18-7450	30.6	54	5.6	76
R2	R18-7451	-44.2	133	-6.7	100
R2	R18-7452	14.0	70	5.1	78
R2	R18-7453	108.2	5	-50.3	151
R2	R18-7454	88.6	9	72.7	6
R2	R18-7455	12.1	72	5.6	77
R2	R18-7456	-53.1	141	-17.2	113
R2	R18-7457	-23.0	116	-4.5	96
R2	R18-7458	-69.9	152	-52.3	154
R2	R18-7459	45.0	35	43.9	24
R2	R18-7460	66.8	13	54.2	14
R2	R18-7461	3.2	85	3.3	81
R2	R18-7462	-32.1	125	-11.2	107
R2	R18-7463	128.4	2	-47.4	147
R2	R18-7464	62.1	17	56.9	13
R2	R18-7465	12.4	71	4.2	80
R2	R18-7466	-1.0	92	-0.8	90
R2	R18-7467	1.1	88	14.5	65
R2	R18-7468	64.5	15	26.6	46
R2	R18-7469	32.2	51	16.9	62
R2	R18-7470	32.1	52	11.2	67
R2	R18-7471	-80.1	156	-57.4	155
R2	R18-7472	-2.5	94	6.8	74
R2	R18-7473	28.0	55	29.5	43
R2	R18-7475	26.6	58	17.2	61
R2	R18-7476	-114.5	166	-79.1	164

Table S 5.10 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R2	R18-7477	72.6	12	11.0	68
R2	R18-7478	24.7	61	1.8	86
R2	R18-7479	41.1	41	46.0	22
R2	R18-7482	58.6	21	73.4	4
R2	R18-7483	-27.2	121	-18.8	119
R2	R18-7484	-16.8	107	-9.2	104
R2	R18-7486	9.8	76	-15.2	111
R3	AG55X7	32.6	64	41.8	65
R3	AG56X8	32.4	65	52.9	55
R3	P53AG7X	-12.2	93	-2.0	94
R3	P55A49X	107.0	25	97.0	26
R3	R11-1057	61.4	45	76.2	39
R3	R11-2933	89.5	34	72.6	43
R3	R18-292	-67.4	129	-40.6	114
R3	R18-293	-11.0	92	12.4	79
R3	R18-294	-166.5	160	-146.2	153
R3	R18-295	-83.0	138	-94.9	139
R3	R18-296	-29.9	103	-51.7	119
R3	R18-297	57.6	49	50.1	57
R3	R18-298	-54.2	120	-20.5	103
R3	R18-299	-25.5	101	-26.5	107
R3	R18-300	-337.9	169	-282.2	169
R3	R18-301	-192.9	164	-202.2	163
R3	R18-302	14.2	78	63.2	48
R3	R18-303	-36.9	110	-34.2	110
R3	R18-304	-150.0	153	-95.3	140
R3	R18-305	-65.4	128	-24.0	104
R3	R18-306	-159.9	158	-260.0	167
R3	R18-307	117.4	19	142.4	15
R3	R18-308	-204.9	165	-272.2	168
R3	R18-309	-67.6	130	24.9	73
R3	R18-310	-107.2	142	-126.1	150
R3	R18-311	-34.9	108	-75.1	132
R3	R18-312	-4.3	89	31.2	70
R3	R18-313	14.1	79	68.5	45
R3	R18-314	-163.8	159	-167.6	158
R3	R18-315	126.6	15	10.4	81

Table S 5.11 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R3	R18-316	-61.7	124	-27.8	108
R3	R18-317	200.6	6	193.8	5
R3	R18-318	-76.2	136	-103.2	144
R3	R18-319	-38.5	112	-24.6	106
R3	R18-322	-32.4	107	-46.6	116
R3	R18-323	-53.9	119	4.1	87
R3	R18-324	-110.9	143	-62.4	126
R3	R18-325	28.4	68	82.3	36
R3	R18-326	-184.4	163	-182.3	159
R3	R18-327	-151.2	154	-133.0	151
R3	R18-328	-24.1	99	20.6	75
R3	R18-329	111.3	24	102.2	23
R3	R18-330	-174.8	161	-149.8	155
R3	R18-331	-129.8	148	-234.3	165
R3	R18-332	74.9	38	75.0	41
R3	R18-333	-32.0	106	-7.3	97
R3	R18-334	-78.5	137	-110.7	146
R3	R18-335	36.7	62	53.0	54
R3	R18-336	-57.6	121	-40.3	113
R3	R18-337	-208.3	167	-153.6	156
R3	R18-338	-30.4	104	-24.1	105
R3	R18-339	-83.8	139	-149.0	154
R3	R18-340	-154.5	155	-164.8	157
R3	R18-341	-61.2	123	-52.9	121
R3	R18-342	50.2	55	108.6	21
R3	R18-343	-0.2	87	2.6	88
R3	R18-344	-64.1	125	-51.8	120
R3	R18-345	-181.5	162	-257.2	166
R3	R18-346	137.4	10	93.3	29
R3	R18-347	-156.0	156	-184.7	160
R3	R18-348	-59.3	122	-84.4	135
R3	R18-349	190.0	7	207.8	4
R3	R18-350	-23.5	98	10.5	80
R3	R18-351	-24.8	100	-65.1	127
R3	R18-352	21.7	73	95.0	27
R3	R18-353	-99.1	140	-96.0	142
R3	R18-354	28.2	70	25.2	72

Table S 5.12 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R3	R18-355	-8.7	90	-11.5	99
R3	R18-356	-207.0	166	-201.9	162
R3	R18-357	-74.9	134	-119.2	148
R3	R18-358	16.9	76	63.9	47
R3	R18-359	-46.6	116	14.2	78
R3	R18-362	-76.0	135	46.8	62
R3	R18-363	-65.2	127	-71.8	130
R3	R18-364	63.0	43	33.5	69
R3	R18-365	-114.6	145	-93.4	138
R3	R18-366	-44.9	115	-75.7	133
R3	R18-367	-118.3	146	-119.6	149
R3	R18-368	-128.0	147	-91.2	137
R3	R18-7389	18.4	75	97.5	25
R3	R18-7390	52.6	52	1.4	90
R3	R18-7391	99.1	30	153.6	12
R3	R18-7392	-138.2	151	-58.5	124
R3	R18-7393	115.0	21	49.9	58
R3	R18-7394	39.9	60	55.3	53
R3	R18-7395	-38.5	113	-18.8	102
R3	R18-7396	60.3	47	40.4	66
R3	R18-7397	-69.7	131	-48.7	117
R3	R18-7398	340.9	2	245.5	2
R3	R18-7399	-225.5	168	-229.0	164
R3	R18-7402	136.3	11	123.8	18
R3	R18-7403	-159.8	157	-95.8	141
R3	R18-7404	-132.8	150	-69.9	129
R3	R18-7405	21.6	74	34.4	68
R3	R18-7406	117.0	20	152.9	13
R3	R18-7407	101.9	29	121.9	19
R3	R18-7408	90.8	33	66.4	46
R3	R18-7409	54.3	50	55.7	52
R3	R18-7410	-18.3	95	-12.2	100
R3	R18-7411	131.5	14	160.0	10
R3	R18-7412	74.1	39	110.9	20
R3	R18-7413	-23.0	97	-38.0	112
R3	R18-7414	50.6	54	82.5	35
R3	R18-7415	-1.6	88	49.9	59

Table S 5.13 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R3	R18-7416	93.8	32	69.5	44
R3	R18-7417	40.3	58	75.2	40
R3	R18-7418	30.1	66	89.6	34
R3	R18-7419	28.3	69	49.2	61
R3	R18-7420	216.8	4	165.8	8
R3	R18-7421	-37.1	111	4.4	86
R3	R18-7422	104.7	26	-1.6	93
R3	R18-7423	-114.0	144	-103.0	143
R3	R18-7424	97.1	31	90.0	32
R3	R18-7425	64.7	41	79.0	37
R3	R18-7426	24.0	71	49.8	60
R3	R18-7427	355.8	1	224.1	3
R3	R18-7428	-52.0	118	-54.4	122
R3	R18-7429	-30.8	105	-45.1	115
R3	R18-7430	168.0	8	173.4	6
R3	R18-7431	-65.1	126	-77.2	134
R3	R18-7432	1.2	86	6.1	84
R3	R18-7433	-145.8	152	-137.9	152
R3	R18-7434	40.3	59	26.5	71
R3	R18-7435	-48.0	117	-4.2	95
R3	R18-7436	-131.6	149	-114.5	147
R3	R18-7437	-11.0	91	-11.4	98
R3	R18-7438	120.8	18	144.8	14
R3	R18-7439	121.5	17	131.1	16
R3	R18-7442	4.8	83	-59.4	125
R3	R18-7443	9.3	82	-55.3	123
R3	R18-7444	71.6	40	100.0	24
R3	R18-7445	45.9	56	-7.1	96
R3	R18-7446	-70.8	132	-65.4	128
R3	R18-7447	-35.1	109	-51.2	118
R3	R18-7448	-70.9	133	-88.0	136
R3	R18-7449	24.0	72	23.5	74
R3	R18-7450	43.3	57	19.9	76
R3	R18-7451	80.8	36	156.0	11
R3	R18-7452	37.3	61	57.4	51
R3	R18-7453	126.0	16	73.0	42
R3	R18-7454	28.9	67	0.0	92

Table S 5.14 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R3	R18-7455	-27.4	102	-37.9	111
R3	R18-7456	80.0	37	46.8	63
R3	R18-7457	132.2	13	57.5	50
R3	R18-7458	133.5	12	161.3	9
R3	R18-7459	104.6	27	89.9	33
R3	R18-7460	59.4	48	-16.9	101
R3	R18-7461	1.5	85	0.7	91
R3	R18-7462	3.4	84	1.6	89
R3	R18-7463	111.7	23	6.5	83
R3	R18-7464	103.4	28	169.7	7
R3	R18-7465	11.4	81	43.5	64
R3	R18-7466	-12.4	94	4.8	85
R3	R18-7467	35.5	63	91.3	30
R3	R18-7468	294.6	3	301.9	1
R3	R18-7469	52.2	53	51.9	56
R3	R18-7470	12.0	80	-108.5	145
R3	R18-7471	-107.1	141	-200.1	161
R3	R18-7472	16.0	77	77.5	38
R3	R18-7473	-42.9	114	-32.7	109
R3	R18-7475	60.3	46	8.8	82
R3	R18-7476	-22.2	96	-73.1	131
R3	R18-7477	205.7	5	104.1	22
R3	R18-7478	64.1	42	36.0	67
R3	R18-7479	153.7	9	94.5	28
R3	R18-7482	113.0	22	124.7	17
R3	R18-7483	61.5	44	62.8	49
R3	R18-7484	85.7	35	90.1	31
R3	R18-7486	53.0	51	14.8	77
R4	AG55X7	8.1	77	-0.2	82
R4	AG56X8	5.5	79	26.9	50
R4	P53AG7X	-3.5	87	-4.1	87
R4	P55A49X	63.5	21	73.3	20
R4	R11-1057	42.6	37	55.1	27
R4	R11-2933	43.2	34	56.3	26
R4	R18-292	-8.8	94	19.2	63
R4	R18-293	16.2	68	-0.5	83
R4	R18-294	-16.0	100	-33.9	118

Table S 5.15 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R4	R18-295	23.1	65	4.2	77
R4	R18-296	-26.9	110	-41.0	125
R4	R18-297	30.5	53	26.4	51
R4	R18-298	-97.4	164	-112.5	164
R4	R18-299	-86.2	159	-67.9	147
R4	R18-300	72.4	13	48.7	34
R4	R18-301	-32.3	113	-63.8	142
R4	R18-302	-72.9	149	-39.3	122
R4	R18-303	-47.5	131	-76.9	157
R4	R18-304	12.8	73	12.4	71
R4	R18-305	-58.2	139	-70.3	152
R4	R18-306	-45.6	130	-109.1	162
R4	R18-307	43.2	35	51.8	31
R4	R18-308	-40.6	126	-60.5	140
R4	R18-309	30.9	51	-21.7	108
R4	R18-310	-78.8	154	-64.8	145
R4	R18-311	30.5	52	-38.9	121
R4	R18-312	-52.8	135	-52.4	134
R4	R18-313	-40.9	127	-41.3	127
R4	R18-314	-50.3	133	-79.0	158
R4	R18-315	-0.4	84	-26.3	114
R4	R18-316	-36.2	119	-30.6	116
R4	R18-317	15.4	69	-66.8	146
R4	R18-318	-91.3	161	-64.1	143
R4	R18-319	-13.1	98	-24.4	110
R4	R18-322	29.9	54	-43.2	129
R4	R18-323	55.1	27	105.3	6
R4	R18-324	-16.3	101	-59.0	139
R4	R18-325	0.8	83	37.8	44
R4	R18-326	-167.8	169	-181.0	168
R4	R18-327	-54.0	137	-75.4	156
R4	R18-328	29.6	55	65.9	24
R4	R18-329	-41.4	128	-27.1	115
R4	R18-330	-33.7	116	-119.3	165
R4	R18-331	7.9	78	-41.1	126
R4	R18-332	29.4	56	-19.1	102
R4	R18-333	-61.9	141	-57.1	137

Table S 5.16 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R4	R18-334	-79.0	155	-69.0	150
R4	R18-335	35.9	46	33.5	47
R4	R18-336	9.5	76	35.0	46
R4	R18-337	-35.6	118	-21.0	106
R4	R18-338	-79.4	156	47.2	37
R4	R18-339	-76.8	152	-46.2	131
R4	R18-340	-62.2	142	-148.0	166
R4	R18-341	-64.9	144	-25.4	113
R4	R18-342	55.9	25	44.9	38
R4	R18-343	-93.6	162	-64.2	144
R4	R18-344	14.2	70	20.0	60
R4	R18-345	-4.8	89	12.0	72
R4	R18-346	-4.3	88	24.0	53
R4	R18-348	-78.1	153	-32.7	117
R4	R18-349	40.8	39	20.7	58
R4	R18-350	-73.6	150	-20.0	104
R4	R18-351	-87.9	160	-98.8	160
R4	R18-352	-47.8	132	-58.9	138
R4	R18-353	34.9	48	81.4	18
R4	R18-354	-28.9	111	-53.4	135
R4	R18-355	43.1	36	47.2	36
R4	R18-356	-16.6	102	-38.1	120
R4	R18-357	26.5	60	-5.2	88
R4	R18-358	-105.9	165	-68.3	148
R4	R18-359	-9.7	95	-39.5	123
R4	R18-362	-72.3	147	-150.5	167
R4	R18-363	-72.6	148	-47.3	132
R4	R18-364	-21.4	104	-16.5	99
R4	R18-365	2.7	81	-20.7	105
R4	R18-366	-33.6	115	-36.8	119
R4	R18-367	-38.1	123	-14.4	98
R4	R18-368	-65.9	145	-12.6	93
R4	R18-7389	75.6	10	105.1	7
R4	R18-7390	13.1	72	67.0	23
R4	R18-7391	45.2	31	22.3	56
R4	R18-7392	-62.6	143	-55.2	136
R4	R18-7393	-120.8	167	-190.5	169

Table S 5.17 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R4	R18-7394	-52.3	134	17.2	66
R4	R18-7395	115.5	5	89.8	13
R4	R18-7396	-11.7	97	-44.6	130
R4	R18-7397	73.8	11	64.3	25
R4	R18-7398	-26.8	109	-74.1	155
R4	R18-7399	34.8	49	10.4	74
R4	R18-7402	10.0	75	-13.4	95
R4	R18-7403	55.7	26	53.5	29
R4	R18-7404	-66.2	146	-25.2	112
R4	R18-7405	38.1	42	8.6	75
R4	R18-7406	111.8	6	105.5	5
R4	R18-7407	67.8	15	38.1	43
R4	R18-7408	-7.4	91	-14.0	96
R4	R18-7409	64.7	19	39.1	42
R4	R18-7410	25.3	62	-16.7	100
R4	R18-7411	96.5	7	89.9	12
R4	R18-7412	44.4	32	70.3	22
R4	R18-7413	67.5	16	19.2	62
R4	R18-7414	28.0	58	101.1	8
R4	R18-7415	-33.4	114	30.0	48
R4	R18-7416	-22.9	106	1.4	81
R4	R18-7417	64.5	20	84.2	16
R4	R18-7418	-3.2	86	-3.9	86
R4	R18-7419	36.6	45	17.4	65
R4	R18-7420	37.8	44	89.7	14
R4	R18-7421	-23.2	107	47.7	35
R4	R18-7422	232.5	2	295.9	2
R4	R18-7423	-84.8	158	-111.7	163
R4	R18-7424	-84.3	157	-68.7	149
R4	R18-7425	67.2	17	2.9	79
R4	R18-7426	-8.2	92	-5.8	89
R4	R18-7427	-53.6	136	-21.0	107
R4	R18-7428	-41.9	129	-23.7	109
R4	R18-7429	51.7	28	92.1	11
R4	R18-7430	-22.3	105	-72.9	153
R4	R18-7431	56.9	23	15.4	69
R4	R18-7432	19.6	66	97.3	10

Table S 5.18 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R4	R18-7433	-34.8	117	-60.6	141
R4	R18-7434	-56.4	138	29.8	49
R4	R18-7435	11.5	74	-1.1	84
R4	R18-7436	73.1	12	54.1	28
R4	R18-7437	2.1	82	-10.1	92
R4	R18-7438	-24.7	108	22.7	55
R4	R18-7439	40.9	38	1.7	80
R4	R18-7442	-19.8	103	-85.6	159
R4	R18-7443	-11.5	96	15.6	68
R4	R18-7444	24.4	63	-12.7	94
R4	R18-7445	5.3	80	-41.3	128
R4	R18-7446	-38.5	125	6.7	76
R4	R18-7447	35.0	47	-19.8	103
R4	R18-7448	84.7	8	78.2	19
R4	R18-7449	-74.1	151	-9.6	91
R4	R18-7450	49.2	29	23.6	54
R4	R18-7451	62.2	22	89.5	15
R4	R18-7452	84.4	9	37.1	45
R4	R18-7453	47.1	30	72.9	21
R4	R18-7454	13.7	71	50.5	33
R4	R18-7455	417.7	1	372.7	1
R4	R18-7456	-36.4	121	-25.1	111
R4	R18-7457	-8.3	93	-69.6	151
R4	R18-7458	-6.7	90	53.2	30
R4	R18-7459	-14.1	99	-14.0	97
R4	R18-7460	37.8	43	10.7	73
R4	R18-7461	66.1	18	44.2	39
R4	R18-7462	-60.7	140	-49.8	133
R4	R18-7463	119.0	4	106.1	4
R4	R18-7464	27.1	59	19.8	61
R4	R18-7465	19.1	67	21.7	57
R4	R18-7466	68.9	14	98.2	9
R4	R18-7467	38.4	41	14.7	70
R4	R18-7468	-38.3	124	25.1	52
R4	R18-7469	25.6	61	51.7	32
R4	R18-7470	24.0	64	-17.4	101
R4	R18-7471	39.4	40	44.0	40

Table S 5.19 (Cont.)

Irrigation Level	Genotype	BLUP Yield Block Model	Ranking Block Model	BLUP Yield Exponential Model	Ranking Exponential Model
R4	R18-7472	29.3	57	39.5	41
R4	R18-7473	-31.7	112	17.4	64
R4	R18-7475	-110.6	166	-41.0	124
R4	R18-7476	-122.5	168	-108.7	161
R4	R18-7477	143.5	3	140.6	3
R4	R18-7478	56.2	24	15.8	67
R4	R18-7479	-2.1	85	20.5	59
R4	R18-7482	44.3	33	83.4	17
R4	R18-7483	34.1	50	4.0	78
R4	R18-7484	-37.0	122	-5.9	90
R4	R18-7486	-36.4	120	-2.6	85

Table S 5.20. Soybean Yield BLUP for the 2019 Final Yield Trials Maturity group IV (19 AF4E) grown in four Arkansas environment under a randomized-complete block design (RCBD) with two replications and ranking of the genotypes after analysis using an exponential model and RCBD models.

Genotype	BLUP Yield Exponential	Ranking Exponential	BLUP Yield RCBD	Ranking RCBD
P48A60X	634.05255	1	584.45757	1
R16-2711	604.47876	2	484.94721	2
AG46X6	480.35846	3	450.05111	3
R17C-257	467.02821	4	415.82963	4
R17-2115	410.39857	5	388.4304	5
R16-247	366.00511	6	375.56648	6
R16-2463	351.26132	7	265.62862	7
R13-1463	272.253	8	239.25551	10
R17-2000	258.40257	9	222.0892	12
R17-2069	256.85386	10	232.96907	11
R17C-130	173.29599	11	173.13314	13
R17C-135	161.32367	12	246.16394	8
R17-56	153.12137	13	245.90949	9
R17C-118	70.256797	14	86.313741	14
R17C-126	27.289597	15	7.0463897	19
R17-2040	12.825484	16	9.3842326	18
R17C-133	-6.080089	17	59.613002	15
R17C-334	-6.802705	18	-3.579024	20
R16-1807	-16.0208	19	11.774524	17
R17C-129	-19.47522	20	-115.4194	25
R17C-105	-24.91122	21	19.349851	16
R17C-128	-52.72936	22	-59.7778	21
R17C-138	-62.50106	23	-82.17117	22
R17C-106	-81.41497	24	-85.22125	23
AG43X8	-119.2773	25	-89.13781	24
R17C-110	-145.5191	26	-184.2027	27
R17C-127	-158.2158	27	-226.7044	28
R17-2170	-246.406	28	-183.0835	26
R17C-132	-271.5854	29	-249.0712	29
R17C-410	-521.6261	30	-458.5194	30
AG39X7	-529.7405	31	-464.9272	31
R17-1945	-606.6862	32	-589.6093	33
R17C-142	-659.3112	33	-584.2931	32
R17C-587	-1170.902	34	-1142.196	34

Table S 5.21. Soybean Yield BLUP for the 2020 Final Yield Trials Maturity group IV (20AF4E) grown in two or four Arkansas environments as a randomized-complete block design (RCBD) with two replications and ranking of the genotypes after analysis using an exponential model and a block model.

Genotype	Yield BLUP Block	Ranking Block	Yield BLUP Exponential	Ranking Exponential
AG46X6	1133.8061	1	993.30648	1
AG51X8	1060.0022	2	903.55705	2
AG39X7	672.30601	3	546.04518	3
AG43X8	546.08446	4	333.08319	5
R18-1417	419.59607	5	353.94512	4
R18-1427	299.00858	6	121.57391	7
R18C-175	184.00012	7	84.869285	8
AG48X9	147.72264	8	23.415576	9
R18-1419	76.433135	9	127.54173	6
R18C-117	-8.099703	10	-42.72571	10
R18-1421	-71.08118	11	-54.03118	11
R18-1420	-94.9929	12	-70.70525	12
R18C-197	-193.1597	13	-287.9568	15
R18C-131	-340.2449	14	-248.4175	14
R18C-137	-535.0847	15	-130.6421	13
R18-1479	-565.5356	16	-503.4771	16
R18-1440	-750.5323	17	-689.8923	17
R18C-120	-863.5431	18	-729.9606	19
R18C-116	-1116.685	19	-729.5289	18

CHAPTER VI

Overall Conclusions

The research studies presented in this dissertation provide in first the part a novel insight of the localization of the high protein gene inherited from 'BARC-7. We have investigated the genetic architecture high protein 'BARC-7' gene using a F_2- derived populations. In fact, one QTL further downstream in Chr. 20 (only detected on population two), explaining 18% of protein variation. The results of our study suggest that 'BARC-7' may carry alleles different from Danbaekkong; this could be useful for breeders to diversify sources of higher protein. However, in order to efficiently implement that in a breeding program, an ongoing fine-mapping using an advanced inbred line mapping approach will help confirm and fine-map the regions associated with high protein and oil in BARC-7 genetic background. Afterwards, we could validate identified SNPs using a Kompetitive allele specific -KASP analysis.

The work presented throughout this dissertation should allow to assess the impact of delaying irrigation on wilting, seed yield, and other agronomic traits of determinate MG 5 soybean. The study suggests that even if high water deficits are experienced at early stages from delayed or inadequate irrigation that yields will likely not be significantly reduced in a furrow irrigation production system for soybean in silt loam soil. A deficit irrigation which is a water-saving irrigation strategy without compromising seed yield, could be implemented for farmers in the Mid -South as result of a groundwater shortage. There are several opportunities for future work stemming from this dissertation. The irrigation was triggered using an atmometer and the reproductive stages due to the variability of the field. A deficit irrigation study would also be meaningful for famers as they are facing groundwater shortage. In fact, a good question that could be interesting to answer is also what happen if we just have specific amount of water to use? How that impact irrigation management on soybean farmers? How could we implement

TDR vs watermark sensors in soybean farm? Additionally, how those results could impact soybean selection on fast- and slow-witling genotype under deficit irrigation? On the other hand, as the canopy wilting is a visual rating that prone to be subjective, using high throughput technology such a drone with thermal sensor would help to discriminate fast- and slow-wilting and might be leading a selection a drought tolerant soybean genotype.

The results of the dissertation pinpointed that canopy wilting and seed yield were quantitative traits. Also, our current showed that genomic selection was efficient to select superior individuals. This investigation will contribute to a better understanding of genetic architecture of soybean lines under reduced irrigation. Also, that breeders should perform independent selection experiments for soybean under full irrigation as opposed to those targeted to withstand any level of water restriction. In fact, implementing genomic selection in breeding could earlier in a breeding pipeline would enhance genetic gain.

The last objective of the dissertation was to evaluate spatial models for seed yield, wilting, and maturity in furrow-irrigated soybean plots. Results revealed that in variety trials with large numbers of genotype, spatial analysis allowed better discrimination among genotypes and increased heritabilities. The spatial analysis led to a different ranking of the genetic materials in comparison with the non-spatial analysis, and selections could have been less influenced by local variation. Such differences in selections may have significant consequences for the outcome of plant breeding programs. Sound recommendations might be applied on the preliminary trials where large genotypes are evaluated in the Soybean Breeding program in Arkansas.