ADAPTATION OF SOYBEAN TO TROPICAL ENVIRONMENTS FOR SMALLHOLDER FARMERS

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

CARRIE MIRANDA

Dr. Kristin Bilyeu, Dissertation Supervisor

May 2018

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation titled:

ADAPTATION OF SOYBEAN TO TROPICAL ENVIRONMENTS FOR SMALL HOLDER FARMERS

presented by Carrie Miranda

a candidate for the degree of Doctor of Philosophy

and hereby certify it is worthy of acceptance

____

Dr. Kristin Bilyeu

____

Dr. Felix Fritschi

____

Dr. Andrew Scaboo

____

Dr. Elroy Cober

____

Dr. David Braun

DEDICATION

To my dog. He’s my Juan and only.

ACKNOWLEDGMENTS

Most importantly, I must thank Kristin Bilyeu for her support even before my dissertation research began. I first approached her during my masters with the uncompromisable research idea of international plant breeding. When every other plant breeding advisor I asked said, “No”, she was the only one who said, “Calm down. We’ll figure something out”. After a fateful trip to the winter nursery in Costa Rica, we came up with a research project to be tested in Ghana, and five days later a grant was written. She supported me as an advisor and as a friend during the difficult times in Ghana. She even showed unparalleled patience and tenacity as I wrote my dissertation in 6 weeks. I am truly grateful to have had her as my advisor.

There are so many people who assisted me for this project. Carolyn Culp is an undergraduate in the Bilyeu Lab and assisted me for so much of this research. I did not have lab access while I was in Ghana and I was dependent on her for most of my genotyping results. She was also a major contributor for Chapter 2. My graduation would have been delayed without her help.

I also want to thank Dr. Kerry Clark. She had experience working Ghana before I arrived, and she connected me to most of the contacts I had there. She also knew what supplies I need and prepared me for the unique work conditions I would experience. My research quality would have suffered dramatically if it weren’t for her advice. She even

watched my dog for me when I was abroad and helped me find my new job! I am extremely grateful to have her as a colleague and a friend.

I need to thank Nicholas Denwar, Stephen Yenori, and Hakeem Apullah. Nicholas Denwar was my in-country mentor in Ghana. He supported my research at the Savannah Agricultural Research Institute by providing me room in his office, a vehicle for work, and most importantly, someone with whom to enthusiastically share my discoveries. Stephen Yenori was my driver, translator, and after thousands of kilometers together, my friend. Hakeem Apullah was my technician in 2017 and is one of the most diligent workers I have ever seen. His passion for this project was as strong as my own.

In addition, I want to thank all members of the Soybean Innovation Lab especially Peter Goldsmith, Courtney Tamimie, Brian Diers, Andrew Scaboo, and Glen Hartman. Every one of them observed my research in Ghana and gave me thoughtful advice about my project. My research would have suffered without their support.

Finally, I want to thank my funding agencies. USAID funded this project through the Soybean Innovation Lab and my Borlaug Graduate Research Fellowship Grant. I was awarded a research fellowship through the University of Missouri Brown Fellowship and the research assistant fellowship through the Gus T. Ridgel Fellowship. This research would not have been possible without their support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS…………………………………………………………...ii

LIST OF FIGURES AND TABLES……………………………………………………viii

ACADEMIC ABSTRACT……………………………………………………………….1

CHAPTER

1. Literature Review…………………………………..…………………………….………..2

    E genes
    Long juvenile trait
    Genetic control of the long juvenile trait
    Current practices with the long juvenile trait
    Stem termination
    Pod shatter
    Current breeding practices in Africa
    Conclusion

    References…………………..……………………….…………………..…28

2. Molecular Tools for Detecting Pdh1 Can Improve Soybean Breeding Efficiency by Reducing Yield Losses Due to Pod Shatter……………………………….…………………………………….….…..36

    Summary……………………………………………………………………37

    Introduction………………………………………………………………....38

    Methods and Materials ………………………………………………………..40

    Pdh1 simpleprobe assay
    Genome wide association study with Pdh1 allele as phenotype
    Associated marker validation
    DNA preparation for IITA lines

    Results ………………………………………………………………..……….43

    Prevalence of Pdh1 in African soybean germplasm
    Discovery of a Pdh1-associated marker
    Prevalence of the shatter allele in the Soybean GRIN collection

Correlation with predicted shatter allele to GRIN shatter score data
Pdh1 in the GRIN data explorer

Discussion ……………………………………………………………….…………47

Figures…..……………………………………………………………….…….52

Tables ……………………………………………..…………………………55

References ………………………………………………………………………58

3. Adaptation of Soybean to Tropical Environments for Smallholder Farmers…………………………………………………………………..…....61

Summary…………………………………………………………...……..62

Introduction……………………………...…………………………………63

Methods and Materials ………………………………………………………..67

Plant materials
RIL populations and field experimental design
Genotyping
Statistical Analysis

Results ………………………………………………………………………...76

Analysis of variance of days to flower, days to maturity, and yield for six RIL populations
Frequencies for days to flower and days to maturity of six RIL populations
In a Jake background, 2 variants of the long juvenile trait exhibit differing means for dtf and dtm
The missense allele of the major maturity gene E1, e1-as, influences days to flower but does not affect days to maturity
Effects of the j-1 and j-x alleles of the long juvenile trait in different genetic backgrounds
E2 affects days to flower and days to maturity in the 534-Can population, but does not have an effect in the X97-Can population
E2 and E3 have an additive effect to delay flowering and maturity in a E1 background in 534-Can but not in X97-Can
Genotype data and means to flower and maturity for other inbred lines of interest

Discussion ………………………………………………………………………86

Figures…..……………………………………………………………………90

Tables ……………………………………………………………………………99

References ……………………………………………………………………...103

4. The Effects of Dt1 on Days to Flower, Days to Maturity, and Height in a Tropical Environment ……………………………………………………………….107

Summary……………………………………………………………………..108

Introduction…………………………………………………………………..109

Methods and Materials ………………………………………………………111

Results ……………………………………………………………………….111

Determinate varieties are the most accessible varieties in Ghana and Mozambique
Population and location effects and their interaction influence days to flower; genotype, population, location and population x location affect days to maturity
Dt1 does not affect days to flower or days to maturity in a tropical environment
E1 and J affect dtf but their interaction with Dt1 does not; Dt1 affects dtm in a conventional background
Dt1, environment, and genetic background influence height
Determinate alleles R166W and P113L height means are not significantly different
Dt1 and environment influence height

Discussion ……………………………………………………………...……115

Figures…..…………………………………………………………….……..117

Tables ………………………………………………………………………..125

References ……………………………………………………………………...135

APPENDIX

1. Supplementary data for Chapter Two ……………………..……………….…..137

Figures…..…………………………………………………………………….…...138

Tables …………………………………………………………………..…….139

2. Supplementary data for Chapter Three…………………….……………………...160

Figures…..……………………...…………………………………………….161

Tables …………………………………..………………………………162

VITA…………………………………………..………………………………..185

LIST OF FIGURES AND TABLES

Figure Page

1. Manhattan plot of Genome Wide Association Analysis Study (GWAS) ……….54
2. The frequency of shatter scores for each allele of *Pdh1*………………..……….55
3. Frequencies of agronomic traits from six RIL populations……………………..92
4. Days to flower and days to maturity in Jake x long juvenile RIL populations…..93
5. Days to flower and days to maturity in a RIL population that was segregating for *e1-as, E1, LJ, and j-1*……………………………………………………..…….94
6. Days to flower and days to maturity for all RILs and parents with a fixed *E1* background…………………………………………………………………….95
7. Days to flower and days to maturity for RILs from the 534-Can population……96
8. Days to flower and days to maturity for RILs from the X97-Can population…...97
9. Days to flower and days to maturity for RILs from the 534-Can population (*E1, E2, E3*, and *j*)…………………………………………………………………..98
10. Days to flower and days to maturity for RILs from the X97-Can population (*E1, E2, E3*, and *j*)…………………………………………………………………..99
11. Days to flower and days to maturity of *Dt1* and *dt1* alleles in all populations....119
12. Days to flower and days to maturity in two populations that were segregating for *Dt1/dt1*, X97-15 and X97-Jen………………………………….120
13. Days to maturity in the X97-15 population that was segregating for *Dt1/dt1*….121
14. Days to flower and days to maturity X97-15 that was segregating for *E1/e1-as, J/j-1,* and *Dt1/dt1*…………………………………………………….122
15. Height for alleles *Dt1* and *dt1* in all populations…………………...………..123
16. Height for alleles *Dt1* and *dt1* (R166W) and *dt1* (P133L) in all populations…..124
17. Height for alleles *Dt1* and *dt1* (P133L) in the X97-Jen population………….....125

Table

1. Prevalence of *Pdh1* shatter allele in 7 released Ghanaian soybean varieties…….57

2. F-test results for each allele of the *Pdh1* associated marker candidate ss715624199. For each allele of ss715624199 it was highly accurate for predicting the *Pdh1*…………………………………………………………………58

3. Genotype data and days to flower and maturity for RIL parents……………..…101

4. RIL population names, parents, and segregating genes of interest……………..102

5. Genotype data and days to flower and maturity for other inbred lines of interest………………………………………………………………..……103

6. *Dt1* alleles of released African varieties and the country in which they were released…………………………………………………………..……….127

7. ANOVA results for *Dt1* alleles on two traits in all six RIL populations: days to flower and days to maturity……………………………………………128

8. ANOVA results for days to flower and days to maturity in two populations that were segregating for *Dt1/dt1*, X97-15 and X97-Jen………………………129

9. ANOVA results for days to flower and days to maturity in the X97-15 population that was segregating for *Dt1/dt1*……………………...………….130

10. ANOVA results for days to flower and days to maturity in the X97-Jen population that was segregating for *Dt1/dt1*……………………………...….131

11. ANOVA results for days to flower and days to maturity for X97-15 on allelic combinations of *E1, J,* and *Dt1*………………..…………………………..132

12. ANOVA results for height in all six RIL populations with the *Dt1* or *dt1* allele……………………………………………………………………………….133

13. ANOVA results for height in all populations with the *Dt1* or *dt1* (R166W) or dt1 (P113L) allele: X97-15 and X97-Jen……………………...………….134

14. ANOVA results for height in the X97-Jen population with the *Dt1* and *dt1* (P113L) alleles……………………………………………………………….135

ACADEMIC ABSTRACT

There are many traits that influence crop adaption to a new environment so that it can perform as a farmer would prefer. In these chapters, I have shown the influence of the genetic mechanisms behind pod shatter, days to flower, days to maturity, and height for soybean in a tropical environment. We developed two molecular tools to identify the allele status of *Pdh1*, a gene that influences shatter. Using those tools, we discovered that this genetic source of shatter susceptibility is still prevalent in African breeding materials and released varieties. I also contributed knowledge of the effects of the maturity genes *E1, E2, E3* and *ELF3* on days to flower and days to maturity. It was discovered that season length can be controlled by choice of the long juvenile trait *ELF3* allele. Days to flower is influenced by *E1* alleles in a *j-1* background, and is influenced by *E1, E2,* and *E3* in a *j-x* background in some cases. I also discovered that the determinate and indeterminate phenotypes do have different influences on height in this environment, but the gene *Dt1* does not affect maturity. The next step is to conduct yield testing to understand how these traits influence yield. In addition, other alleles of *ELF3* should be bred into different backgrounds to see if they influence different season lengths as well. Finally, the genetic source of the long juvenile trait in the current African released varieties need to be discovered. Taken together the future data combined with the data presented here can assist a local breeder in Africa to choose the germplasm they want to control their season length or protect yields from pod shatter and ultimately create a new, elite African variety.

CHAPTER ONE

Literature Review

Soybean is an economically important legume crop that has multiple uses including oil production, livestock feed, and high protein content for human consumption (Masuda and Goldsmith 2009). Due to the potential uses and profits, many smallholder farmers in tropical developing countries grow soybean; however, yields are lower than their potential (Abate and Orr 1981, Mbanya 2011, Abate et al. 2012, Alene et al. 2012, IITA 2014). Low soybean yields are a multifaceted problem for the smallholder farmer in tropical Africa, due to many factors such as adaptation, rainfall, soil health, mineral utilization, access to supplies and disease pressure. Developing a core of knowledge that defines what constitutes maximum adaptation to the tropical African target soybean production regions is a key step to addressing poor yields in this environment. Important traits for adaptation are days to flowering, days to maturity, and plant height at specific latitudes based on the genes underlying the photoperiod and plant architecture responses that will all influence yield.

Soybean is a photoperiod sensitive, short day plant where flowering is induced when the daylength is shorter than a maximum critical value (Garner and Allard 1920, Whigham and Minor 1978, Destro et al. 2001, Watanabe et al. 2012). This attribute has limited the expansion of soybean cultivation in latitudes less than 20° where daylengths never deviate from around 12-13.5 hours. Soybean was domesticated at high latitudes north of 30°N and was predominantly cultivated in the Northern Hemisphere. In temperate production zones, soybean varieties have been developed that are adapted to fairly narrow bands of latitude, and a system of classification has been developed to assign "maturity groups" of 000 to VIII for production in North America (Zhang et al. 2007).

When temperate soybean cultivars are grown in low latitudes, plants of most maturity groups begin the reproductive flowering stage somewhat synchronously approximately four weeks after emergence (Destro et al. 2001). Very short days (~12 hours) can cause a reduction of the soybean vegetative stage of growth. When the vegetative stage of development is stunted, the plant has a short stature and a low leaf area index (Sinclair and Hinson 1992) which leads to reduced yields.

Brazil and Argentina were the first countries to introduce commercial soybean production to the Southern Hemisphere, but growth was initially restricted to high latitudes south of 22° (Carpentieri-Pípolo et al. 2000, Carpentieri-Pipolo et al. 2002). In the late 1970s, Brazil was the forerunner of soybean cultivation in latitudes closer to the equator, and they were able to expand soybean growth to low latitudes at less than 15°S (Spehar 1995, Carpentieri-Pipolo et al. 2002), however the genetic mechanisms responsible were not fully known until recently. Understanding soybean adaptation to low latitudes is a key step in order to expand and increase soybean yields in tropical climates as photoperiod response is the most important trait for adaptation (Bandillo et al. 2017) Four traits are critical components to build a foundation of a successful tropical soybean variety: general maturity group control through the *E* genes, eliminating pod shatter by selecting against the *Pdh1* gene (Funatsuki et al. 2014), control of stem termination through the determinate/indeterminate and semi-determinate genes, and utilization of the long juvenile trait which allows for prolonged vegetative growth in short daylengths.

E genes

Maturity group classification in North America is well characterized and can now be determined by the assortment of alleles at the *E* loci (Bernard 1971, Buzzell 1971, Buzzell and Voldeng 1980, McBlain and Bernard 1987, Bonato and Vello 1999, Cober et al. 2010, Langewisch et al. 2017). Genetic control of soybean maturity, the culmination of photoperiod responses of plant development, has been researched starting in the 1920s (Woodworth 1923), and as of 2018, eight unique loci have been identified that have strong control on local maturity adaptation, named *E1-E10*.

The science behind germplasm adaptation for tropical climates is still in the discovery phase, and success begins by understanding genetic control of maturity through the *E* genes.

E1 and *E2*: In early 1971, Richard Bernard published the first paper using the nomenclature *E1/e1* and *E2/e2* to describe the two major genes that affect time to flower and maturity in soybean (Bernard 1971). In order to understand the difference between late and early maturity, he backcrossed two lines: T175 and T245 respectively, using Clark as the recurrent parent for both. Both series of backcrosses provided sufficient evidence to suggest a gene that controls lateness of flowering (*E1/E1*) and a gene that controls early flowering (*e2/e2*) when contrasted to the Clark control. To determine linkage of the two genes, near isogenic lines of both lineages exhibiting the extreme of either the early or late phenotype were crossed (named Clark_{E1} and Clark_{e2}). To identify which lines had the *E1* allele, the linkage of *E1* and pubescence color was utilized.

Hanson (Hanson 1961) and Weiss (Weiss 1970) and later Cober (Cober and Voldeng 2001) discovered that early maturing varieties have a tawny pubescence color and late maturing have a gray color. By analyzing the segregation ratios of pubescence color and contrasting it to flowering dates, it was determined that there was a linkage relationship to pubescence color and *E1* i.e. *E1/t* and *e1/T*. These data were used to classify *E1, e1, E2, or e2*. Bernard was able to create three new maturity varieties of Clark which all flowered and matured at different dates (Bernard 1971). He determined that the Clark variety is *e1/E2*, where $\text{Clark}_{\text{E1E2}}$ flowers and matures almost 30 days later than the original variety. Clark_{e1e2} flowered and matured earlier than the original variety; interestingly, Clark_{E1e2} flowered later than the original variety but earlier than *E1/E2*. These data suggest that both dominant alleles of the *E1* and *E2* genes play a role in delaying flowering and maturity, however *E1* has a stronger effect.

The genetic location and molecular identity of the *E1* gene was elusive until recently. QTL mapping, conducted by Tasma et al in 2001(Tasma et al. 2001), located an area that regulated about 47% of phenotypic flowering variance and was tightly linked to pubescent color locus *T*, which highly suggested *E1* and was later confirmed in 2005 (Yamanaka et al. 2005). Due to the pericentromeric location of *E1* (Schmutz et al. 2010), it was difficult to identify and characterize. In 2012, Xia et al were able to definitively locate *E1* and its molecular identity through positional cloning (*Glyma.06g207800*) (Xia et al. 2012). They determined that the *E1* gene encodes a B3 domain which includes it in a superfamily that plays many roles in plant growth and development (Swaminathan et al. 2008); however, amino acid identity was too low with known B3 proteins to predict

function or homology. An arginine residue in the first domain suggests that the protein has a nuclear localization, and with the DNA binding capabilities from the B3 domain, it is possible that *E1* is a transcription factor. Interestingly, through discovery of the importance of the arginine residue for nuclear transport, the main recessive allele of *E1, e1-as*, was named that has a R15T missense mutation at that position. Varieties with that genotype were previously classified as *e1*, due to the early flowering phenotype observed. Xia's group showed that there are not differences in rates of transcription of *E1* versus *e1-as*, but a nuclear transport issue, which suggests that the *e1-as* allele may have a weaker effect on phenotype than *e1* (now called *e1-null* and presumably the *e7* variants (Cober and Voldeng 2001)). Thus, an allelic series exists at the *E1* locus: the functional *E1* allele, the early flowering missense *e1-as* allele, and the very early flowering null *e1* alleles (Xia et al. 2012). Using transgenic crops that overexpress *E1* they discovered a down regulation of *GmFT2a* (also known as *E9*) and *GmFT5a*, orthologs of *Arabidopsis FLOWERING LOCUS T*, a well characterized locus which produces a florigen signaling molecule that leads to early flowering (Samach et al. 2000, Robson et al. 2001, Kong et al. 2010).

Besides the genetic work by Bernard with *E2*, an independent group in Japan characterized the FT2 locus for early flowering. It became obvious only later that *FT2* and *E2* were the same gene (Yamanaka et al. 2001). The molecular identity of *E2* was solved by Watanabe et al in 2011 (Watanabe et al. 2011) through map-based cloning techniques. *E2* (*Glyma.10g221500*) was identified to be an ortholog of *GIGANTEA (GI),* a protein that plays a role in regulation of the circadian rhythm in barley and *Arabidopsis*

(Fowler et al. 1999, Dunford et al. 2005). Mutants of *GIGANTEA* in soybean have elevated levels of *GmFT2a* and early flowering, which genetics and phenotyping confirm (Watanabe et al. 2009). Functional *E2* and its nonfunctional, early flowering allele variant *e2* have been described and are widely dispersed in soybean.

E3: In the fall of 1971, Buzzell (Buzzell 1971) conducted an experiment contrasting the days to flower in the soybean varieties Harosoy 63 and Blackhawk in 20 hour incandescent lighting, continuing research initiated by Fisher (Fisher 1963). Harosoy 63 displayed delayed flowering by 28 days compared to Blackhawk in greenhouse experiments with regulated 20-hour white light exposure. This experiment classifies Harosoy 63 as fluorescent sensitive (flowering can be delayed by fluorescent lights) and Blackhawk as fluorescent insensitive. The sensitivity translated to only eight days difference in delayed flowering of Harosoy 63 in field conditions. Segregation ratios of 3:1 for fluorescent sensitivity were observed, leading to the conclusion that this trait is under the control of one gene. Since the previous Bernard *E1/E2* studies were done in the Clark background (fluorescent sensitive) and the trait observed here cannot be controlled by *E1* or *E2*, the term *E3* was created where *e3* conditions earliness. Independent work in Japan also genetically identified the *E3* gene (Watanabe et al. 2004). The molecular identity of *E3* (*Glyma.19g224200*) was discovered in 2009 by Watanabe et al, the same group that mapped *E2* (Watanabe et al. 2009). Using the same techniques that led to the cloning of *E2*, they determined that *E3* is the phytochrome A protein, *GmPhyA3*. Two functional *E3* alleles differing in the size of the last intron and several

nonfunctional *e3* alleles have been described. The soybean genome contains potentially functionally redundant copies of the phytochrome A genes.

E4: In 1980, Buzzell and Voldeng (Buzzell and Voldeng 1980), using a similar experimental design as the *E3* study, explored more varieties of soybean under 20 hour fluorescent exposure, and discovered another locus, *E4*, where *E4* is observed in fluorescent and photoperiod insensitive varieties, and *e3e4* is found in incandescent insensitive varieties (Voldeng and Saindon 1991).

The molecular identity of *E4* was discovered in 2008 by Liu et al, and it is also a phytochrome A ortholog, named *GmPhyA2* (*Glyma.20g090000)* (Liu et al. 2008). *E3* and *E4* (*GmPhyA3* and *GmPhyA2*) both play a role in regulating *E1*, however that exact model is not known (Kong et al. 2010). Although variants of *E4* are relatively rare, the most common dysfunctional allele was *e4-SORE1*, followed by *e4-kes, e4-kam*, and *e4-oto* (Xu et al. 2013)

E5: In 1987, McBlain and Bernard (McBlain and Bernard 1987) described the identification of the trait *E5*. Their research sought to answer how genetic control of maturity groups greater than V is achieved. Their previous research of the *E1-E4* loci allowed for the reproduction of MG I-V varieties through manipulation of the *E* genes, however they were not able to account for lines in the later maturity groups. It was also unclear if knowledge of all loci that control MG I-V were known. Using Harosoy (*e1, e2, E3*; MGII) and NILs with single allelic substitutions (*E1, E2*, and *e3*) crosses were made with the late experimental line L64-4830 (BC_5F_5 Harosoy x PI 80.837). L64-4830

was chosen for further testing as it exhibited lateness from a single unknown recessive gene that was inherited from the PI line. Crosses with L64-4830 with Harosoy and the Harosoy NILs clearly demonstrated that the lateness observed is not attributed to any of the *E* loci that has been previously described. *E5* is used to describe lateness (inherited from PI 80.837) and *e5* is observed in Harosoy. In 2016, Dissananyaka et al. performed a QTL mapping experiment to discover the loci with the *E5* gene. They generated mapping populations using Bernard's original germplasm used to discover *E5* but were unable to discover any new QTL peaks that did not associate with an already known *E* gene. They proposed that the lateness observed in *E5* was due to an interaction with different alleles of *E2*: *E2-in* and *E2-dl* and was not a new gene (Dissanayaka et al. 2016).

E6: When soybean commercial production was expanded to southern Brazil, the MG VI Paraná variety was released from North Carolina in 1977 and described by Laster et al in 1979. From this original variety, two naturally occurring variants were observed: Paranagoiana and SS-1 (Bonato and Vello 1999). These two variants exhibited delayed flowering compared to Paraná in daylengths between 13.5 and 14.5 hours. Crosses between Paraná and the two variants and the two variants themselves were made and contrasted to the parents. In both Paraná crosses with the variants, segregation ratios were 3:1, suggesting that both varieties were derived from a single recessive mutation. In the Paranagoiana x SS-1 cross, phenotypes were difficult to discern from each parent as all displayed similar, delayed flowering and maturity, and the result suggested high genetic similarity. Based on these data, the new gene locus was named *E6*, where *e6*

delays flowering in short days. This locus will be elaborated on in more detail in the long juvenile section.

E7: In 2001, the locus *E7* was described by Cober and Voldeng. The new photoperiod sensitivity locus was found to be linked to *E1* and the tawny gene (Cober and Voldeng 2001). With the molecular genetic characterization of *E1* in 2012, it became apparent that *E7* is actually a null allele of *E1* that is distinct from the R15T missense allele now named *e1-as* (Xia et al. 2012).

E8: In 2010, Cober et al. (Cober et al. 2010) described the discovery of the maturity gene *E8*. Working in Harosoy, NILs of different *E* gene alleles, and Maple Presto, a late maturing line, were crossed with early maturing PI lines. A 3:1 segregation of late maturing phenotypes was observed, leading to the conclusion that early maturity was under the control of a single recessive allele, which was named *e8*. One of the final conclusions of this publication states that recessive alleles at all known *E* loci results in the maturity group 000.

E9: In 2014, Kong et al described a new dominant gene for early flowering and maturity that was identified in genetics and mapping experiments with early maturing soybean lines. The new gene, named *E9*, was mapped to a small interval on chromosome 16 that includes two orthologues of the *Flowering Locus T* genes (Kong et al. 2014). Fine mapping was conducted on *E9* and was revealed to be the *Arabidopsis* florigen ortholog

FT2a (Glyma.16g150700) (Zhao et al. 2016), where the recessive allele caused by the same *SORE-1* insertion as also found in *E4*, delays flowering.

E10: In 2017, Samanfar et al. discovered that isolines of varieties Maple Presto and Harosoy showed a 6 days difference in days to maturity even though they had identical recessive *e1-e8* and functional *E9* genes in both backgrounds. SSR marker analysis of populations developed through crosses of the isolines showed a correlated region on chromosome 8 named *E10*, where the recessive allele *e10* conditions earlier maturity than *E10.* A functional genomics approach was used to determine which of the 75 genes in the genomic region could be candidate genes for the molecular identity of *E10* and discovered that *FT4*, an *Arabidopsis* flowering ortholog, was predicted to be the mechanism behind *E10* (Samanfar et al. 2017).

It is believed that manipulation of the *E1-E4* genes plays the biggest role in determining American soybean maturity groups in temperate latitudes (Langewisch et al. 2017). These soybean varieties are generally considered photoperiod sensitive and that allows them to respond to relatively small differences in daylength to optimize growth in appropriate latitudes with long days (the opposite of the tropics). A study of North American cultivars conducted by Langewisch et al. showed that 70% utilize the *e1-as* allele, showing strong artificial selection for that *E1* allele. They also discovered that 71% of the North American soybean ancestors were *E2,* even though the recessive *e2* allele is preferred in landraces and Chinese cultivars, but there was an even distribution of both alleles in the American cultivars. In all they discovered that 28% of all US

cultivars have the genotype *e1-as, E2, E3*, 22% have the genotype *e1-as, e2, E3,* and 17% have the genotype *E1, E2, E3.* These genotypes are predominant in the Midwest and Southern United States (Langewisch et al. 2017).

In the early maturing, northern limits of soybean adaptation, Xu et al tested 53 photoperiod insensitive varieties for specific/novel alleles and allelic combinations at the four loci (Xu et al. 2013). They determined that all accessions were recessive at *e2*, which must be essential for adaptation at higher latitudes. They also found ~70% of the varieties tested were *e3/e4* which is also crucial for photoperiod insensitivity. Other studies have shown that dysfunctional *E1* leads to the earliest flowering, which lends more support about the strong influence of *E1* (Zhai et al. 2014). When dominant alleles were present at all four loci, the greatest affect in delayed flowering was observed (Zhai et al. 2014); however, the effect of dominant alleles at all four *E* loci is not strong enough to delay flowering to allow for optimum vegetative growth in tropical climates. The other *E* gene loci and their most advantageous allelic combinations, or perhaps other genes such as the long juvenile trait, are necessary for adaptation of soybean in low latitudes.

Long juvenile trait

In 1979, Hartwig and Kiihl identified a plant introduction line, PI 159925, that had delayed flowering under short day conditions and since then has been described as the long juvenile trait (Hartwig and Kiihl 1979, Cregan and Hartwig 1984, Sinclair et al. 1991, Sinclair and Hinson 1992, Collinson et al. 1993, Ray et al. 1995, Cober et al. 1996, Cober 2011).

In 1977, Paraná was cultivated throughout southern Brazil, (LIMA et al. 2000) and two natural variants were observed, SS-1 and Paranagoiana, which exhibited delayed flowering and time to maturity compared to its progenitor (de Pesquisa 1986, Destro et al. 2001). It was determined that this phenotype was attributed to genetic control other than that of maturity; the long juvenile trait was again observed through a separate event (Bonato and Vello 1999).

It is important to discern the difference between late maturity (MG V+) and the long juvenile trait. A soybean plant that exhibits the long juvenile trait will have a longer period of vegetative growth during which flowering cannot be induced even by a reduced photoperiod of 12 hours (Bäurle and Dean 2006). A late maturing variety (V+) with a conventional juvenile characteristic is induced to have a shortened vegetative stage when photoperiod is manipulated to a critical point (12-hour daylength) during early growth. Elroy Cober conducted an experiment to prove this point in 2011 (Cober 2011). First, he grew the conventional juvenile (CJ) line Paraná and the long juvenile (LJ) line Paranagoiana in growth chambers with controlled lighting ranging from 4 to 16 hours of daylength in 2 hour increments and a constant temperature of 25°C. It is important to note that the two lines come from near identical genetic background, as Paranagoiana is a natural variant of Paraná. Photoperiods and days to flowering were compared for the two lines. Interestingly, it is observed that Paraná (CJ) exhibits days to flower similar to Paranagoiana at certain photoperiods including 4 hour and 16 hour photoperiods. In all other photoperiod lengths, Paranagoiana (LJ) flowered later than Paraná, and the largest

difference was observed at the 12 hour daylength. Days to flowering in the field between these two lines is similar to these experimental results (Cober 2011).

In a following experiment, Cober observed days to flowering of Paraná (CJ), Paranagoiana (LJ), PI 159925 (LJ) and X5063-39 (a line developed by backcrossing the LJ trait from Paranagoiana into OT94-47: MG 0, *e1-null*) in growth photoperiods from 3-12 hours; the 12 hour daylength showed the greatest contrast in flowering. PI 159925 and Paranagoiana show similarity in days to flower contrasted to the photoperiods, and exhibit delayed flowering compared to Paraná. The X5063-39 line exhibited delayed flowering similar to the other LJ lines at 3 and 12 hour photoperiods, but interestingly showed days to flowering intermediate to that of Paraná and the other LJ lines at all other photoperiods.

These results provide interesting insight into the classification of the long juvenile phenotype. First, the distinction between CJ and LJ lines is photoperiod dependent; the days to flowering phenotypes are identical at both high and low extremes of day length. Second, naturally occurring LJ lines (Paranagoiana and PI 159925) share nearly identical phenotypes across all photoperiods. Lastly, when the trait is bred into another background (here, OT94-47, a Harosoy isoline) the dramatic difference between CJ and LJ days to flower in intermediate photoperiods (6 and 8 hours) is lessened. When the LJ trait is used in a breeding program, a sharp contrast may not be observed in flowering depending on the photoperiod, influence from *E* gene allelic combination, and presence of the LJ trait, leading to ambiguous distinction between CJ and LJ lines based on

phenotype alone. It is necessary to understand genetic control to properly discern when the long juvenile trait is present.

Genetic control of the long juvenile trait

Opinions of genetic control have been discussed since the 1970s (Hartwig and Kiihl 1979), however there was not enough data to conclude the source of inheritance. In 1995, Ray et al. conducted field experiments in Florida over several years to determine the number of genes that control the long juvenile trait based on segregation ratios (Ray et al. 1995). They first created crosses between four conventional juvenile (CJ) varieties (Bedford [MG V], Will, F85-431, F85-459 [MG VII]) and a long juvenile (LJ) donor (PI 159925). Progeny were described as either CJ or LJ by comparing the flowering date distribution to the respective parent flowering date distribution on two different growing dates at either 13 or 14 hour daylengths in Florida latitudes (~29°N). The F_2 generation of Bedford x 159925 and F85-431 x 159925 showed a 3:1 CJ/LJ segregation on both planting dates, as did F85-459 x 159925 on one date. Will x 159925 was only evaluated on one date; however, it showed continuous segregation for days to flowering. These results suggest control of LJ by one recessive gene; however, since the segregation ratio was not consistent with all crosses at all planting dates, the authors believed that genetic background also affects flowering.

From the F_2 progeny, early flowering and late flowering individuals were chosen from the 4 crosses and used to create near isogenic lines (NIL) pairs, where one individual flowers early and the other late but in the same background (Bedford, Will, F85-431, or F85-459). Individuals from each NIL pair were crossed and F_2 generations were

analyzed for segregation ratios of the LJ trait as previously described. These four crossed pairs were sown in three separate years (1987, 1990, and 1991), and with the exception of an F_2 cross in 1990 (which was not significant) all F_2 generations from all four NIL pairs showed a 3:1 CJ/LJ segregation. This adds support to the LJ trait being under the control of one recessive gene but also that genetic background does influence days to flowering, which was not elaborated on. This paper designates the nomenclature of the *J locus* for the unknown gene controlling the long juvenile trait in PI 159925, where *JJ* is found in CJ lines and *jj* is found in LJ lines (Ray et al. 1995).

In 1999, Bonato and Vello published an independent set of experiments to determine genetic control of LJ in Brazilian soybean varieties. In their experiments they used the CJ variety Paraná (MG VI), and natural LJ variants of Paraná, Paranagoiana and SS-1 (Bonato and Vello 1999). They made crosses, Paraná (CJ) x Paranagoiana (LJ), Paraná (CJ) x SS-1 (LJ), and Paranagoiana (LJ) x SS-1 (LJ) and compared the F_2 generation for segregation of the LJ trait. In both CJ x LJ populations, they also observed a 3:1 CJ/LJ segregation, but segregation could not be determined in the LJ x LJ due to phenotype similarity. The F_2 and the F_3 progeny of the Paranagoiana x SS-1 cross did not segregate into classifiable groups. They stated this could be attributed to "alleles of the cultivars SS-1 and Paranagoiana are different alleles at the same locus". The nomenclature *E6* locus was assigned to denote early flowering and maturity, where *E6* is present in CJ lines and *e6* is present in LJ lines.

At this point in the literature, there was no distinction between the *J* locus and the *E6* locus, except that *j* was used to refer to LJ varieties that used 159925 as a donor and *e6* was used in Brazilian LJ research. In 2000 and 2002, two papers were published by Carpentieri-Pípolo et al., suggesting that the LJ trait was under the control of two recessive genes. In 2011, Elroy Cober published a paper to address all of these dissimilarities (Cober 2011). Cober noted that all previous experiments conducted with LJ lines had been done in late maturing varieties (MG VI-VIII), so he chose OT94-47, an early maturing Harosoy line (MG 0) for LJ crosses to determine long juvenile effect on flowering. The F_2 results of OT 94-47 x Paranagoiana show a segregation ratio of 15:1 CJ/LJ. The same ratio is seen in 3 consecutive backcrosses, using OT 94-47 as the recurrent parent. This same ratio is seen again in the F_2 generation of OT 94-47 x PI 159925. These results highly suggest that in an MG 0 background, observance of the LJ trait is under the control of two recessive genes, which may include influence from one of the recessive *E* genes that are necessary to achieve MG 0. When the exact sequence of the gene controlling the LJ trait was not known, it was difficult to discern the number of genes that influence the phenotype, especially when the maturity group also played a strong role. It is important to note, though, that MG V+ may be necessary for cultivating soybean in the tropics, and previous studies suggest that in this maturity background, the LJ segregates under Mendel's law of a single gene. In maturity groups later than MG V, it has been shown that the *E* genes have dominant alleles at *E1-E4* (Langewisch et al. 2017). If the LJ trait has an epistatic effect with *E1, E2, E3,* or *E4,* the 3:1 ratio may be present when the *E* gene is functional, and 15:1 when recessive at one of the *E* loci.

Understanding the relationship of maturity grouping with the long juvenile phenotype is essential to adapting temperate soybean to the tropics.

Recently, the genetic mechanism controlling the *J* allele has been discovered. First, it was mapped to a QTL on chromosome 4 (Cairo et al. 2002) and the causative gene was cloned by Yue and Lu (Lu et al. 2017, Yue et al. 2017). The gene controlling the *J* long juvenile trait was identified as the *Arabidopsis* flowering ortholog *ELF3* (Zagotta et al. 1996). *ELF3 (Glyma.04G050200.1)* has 4 exons/3 introns and is a highly conserved protein that controls flowering time in multiple species (Lu et al. 2017). Yue cloned *ELF3* from the Chinese variety Huaxia 3 and discovered a thymine deletion resulting a frameshift mutation creating a truncated, nonfunctional protein. They also sequenced *ELF3* in 170 other soybean varieties and discovered 8 other polymorphisms, some that were synonymous mutations and may not affect the phenotype (Yue et al. 2017). Lu also identified *ELF3* as the genetic mechanism behind the long juvenile trait close to the time of Yue (Lu et al. 2017). Their work went into more detail confirming *ELF3* as the genetic mechanism of the *J* long juvenile trait by doing positional cloning of a Brazilian long juvenile trait variety, BR121. They identified a 10bp indel in exon 2 of *ELF3*, a new polymorphism compared to the Yue SNPs. They then sequenced *ELF3* in PI 159925, the plant introduction line where the *J* allele was discovered (Hartwig and Edwards 1970, Hartwig and Kiihl 1979, Ray et al. 1995) and discovered 4 polymorphisms contrasted to CJ Harosoy, 3 SNPS and a cytosine deletion that causes a frameshift. This C deletion was also a new polymorphism contrasted to Yue's work. To discover other polymorphisms, they examined *ELF3* in the 302 sequenced Zhou soybean lines (Zhou et

al. 2015) and an additional 125 lines from low latitude areas. They discovered 34 SNPs and six indels in *ELF3* and named 34 haplotypes. The eight significant SNPs and indels that are frameshift mutations predicted to cause the delayed flowering phenotype are named *j-#* where *j-1* denotes the C deletion of PI 159925, *j-2* names the 10bp deletion in BR121, etc.

Lu et al. also confirmed the role of *ELF3* in the soybean flowering pathway. They discovered that *J* binds with *ELF4/LUX* proteins and acts upstream of *E1* to suppress *E1* expression when *J* is functional, which was consistent with previous findings (Xia et al. 2012). When *j* is nonfunctional, the repression is not observed in a 12 hour day length, showing a delay in flowering time (Lu et al. 2017), suggesting that *ELF3* is a transcriptional repressor of *E1*. They propose a simple flowering model, where *E3* and *E4* partially suppress *J*, and *J* suppresses *E1* which allows for the expression of *FT2a* and *FT5a* and early flowering in short day conditions. Conversely, when *j* is inactive, *E1* is expressed normally and is able to repress *FT2a* and *FT5a,* allowing for delayed flowering in short days. When they sequenced *ELF3* in low latitude lines, not all lines showed polymorphisms in the gene, suggesting other genes may be responsible in a quantitative manner (Cober et al. 1996, Carpentieri-Pipolo et al. 2002, Lu et al. 2017). These data also provide evidence that *ELF4, LUX, FT2a* or *FT5a* recessive alleles could also be candidates for delayed flowering in short days (Lu et al. 2015).

At this point in the literature, the difference between *J* and *E6* has not been discussed molecularly. Specifically, the genetic mechanism behind the long juvenile trait in the

Brazilian line, Paranagoiana where *E6* was discovered (Bonato and Vello 1999), was unknown. Li et al (Li et al. 2017) conducted QTL mapping by crossing Paranagoiana (*E1,* LJ) by Harosoy (*e1-as,* CJ), OT94-47 (*e1-null* [also known as *e7* (Cober and Voldeng 2001)] CJ), and PI 159925 (*E1,* LJ). When Paranagoiana was crossed by the two conventional juvenile varieties, a QTL on chromosome 4 was consistently observed as well as a peak corresponding with *E1* (Xia et al. 2012) since both populations were segregating for *E1* with either *e1-as* or *e1-null*. The same QTL on chromosome 4 was seen in the Paranagoiana x PI 159925 mapping population and was at the location that *ELF3* was mapped to suggesting that *J* and *E6* are the same gene. However, when sequencing *ELF3* in Paranagoiana, no polymorphisms were detected (Li et al. 2017).

When analyzing days to flower, it appears that Paranagoiana had more delayed flowering when compared to PI 159925, showing that they do not have identical phenotypes. These results suggest that the gene controlling the long juvenile trait in Paranagoiana is closely linked to *ELF3* or there is a complicated mutation in *ELF3* that was not detected.
After these discoveries the semantics to describe the long juvenile trait herein will be *J* for conventional lines and *j* for lines that exhibit delayed flowering in short days. The name of the causative SNP, for example *j-1* to describe the C deletion in PI 159925, is the most accurate.

Current practices with the long juvenile trait

As already stated, Brazil attributes the long juvenile trait to expansion of soybean farming to the northern, tropical part of the country. Unfortunately, current Brazilian varieties

cannot be released in Africa as the majority of Brazilian commercial soybean lines have transgenic traits subject to regulation in many African countries. Many food insecure countries in Sub-Saharan Africa do not allow the cultivation of GMOs (Protocols 2000). In addition to that major hurdle, the environment of northern Brazil is vastly different than that of most tropical Africa in terms of rainfall and soil (EMBRAPA 2014), and the Brazilian alleles of the long juvenile trait may not be suited to the short season varieties desired in Africa.

Interestingly, in northern Australia (22-28°S), breeding research is being conducted using the long juvenile trait as a way to delay flowering due to temperature cues. One paper comments on the difficulty of trying to grow southern US varieties in eastern Australia that were "confounded by large G x E effects on yield" (Lawn and James 2011) adding support that maturity alone is not sufficient to introduce soybean into tropical climates. In a breeding experiment to create a subtropical soybean variety, James and Lawn backcrossed a *j* line developed in Florida by Dr. Kuell Hinson with a temperate semi-dwarf variety that demonstrated high yield and lodging resistance as the recurrent parent (James and Lawn 2011). Compared to controls, they were able to maintain high yield, improve lodging resistance, and also delay flowering, demonstrating the capability of introducing temperate varieties into subtropical environments by backcrossing in the long juvenile trait.

While several of the most intensively studied sources of the long juvenile trait appear to have very similar and drastic delayed flowering phenotypes (PI 159925, Paranagoiana,

and SS-1), the possibility exists that an allelic series at the same *J* locus could present an opportunity to fine tune the vegetative and reproductive periods in tropical environments. Alternative allele sources of the long juvenile trait may offer opportunities to create shorter season, but in high yielding soybean cultivars adapted to the tropics.

Stem termination

The genetic and molecular control of soybean plant architecture traits are known. In 1972, Bernard discovered two genes, *Dt1* and *Dt2*, which through allelic and epistatic interactions produce the three soybean stem architecture phenotypes, indeterminate, determinate, and semi-determinate. In *Dt1/Dt1* backgrounds, *Dt2/Dt2* produce semi-determinate varieties and *dt2/dt2* have an indeterminate phenotype. The determinate varieties have *dt1/dt1* regardless of the alleles at *Dt2* due to an epistatic effect (Bernard 1972, Tian et al. 2010). Thompson et al identified a third *Dt1* allele, *dt1-t*, that has an effect in-between determinate and semi-determinate termed tall determinate (Thompson et al. 1997).

Of the three known variations of stem termination, the most common in North America and ancestrally are the indeterminates, (*Dt1*), where vegetative growth continues during the reproductive stage (Bernard 1972). Determinates, (*dt1*), halt vegetative growth of the main stem immediately at the start of flowering; these varieties are most common in the southern USA (Tian et al. 2010). Third are the semi-determinates (*Dt2*) which express, in the indeterminate genetic background only, a phenotype of intermediate stature with a terminal raceme. The indeterminate gene *Dt1* was shown to encode a homologue of the

Arabidopsis regulatory protein *Terminal Flower 1* (Liu et al. 2010, Tian et al. 2010). Four missense mutations were identified that defined the known *dt1* alleles conditioning the determinate plant architecture trait (Tian et al. 2010). The *Dt2* gene was recently shown to be a gain of function MADS-Domain Factor gene that was thought to regulate the *Dt1* gene (Ping et al. 2014).

Plant architecture traits will play a role in the success of soybean cultivation in the tropics as vegetative growth/height/number of nodes is currently limiting yield potential. There are pros and cons to all three plant architecture phenotypes. It has been demonstrated that indeterminate growth is advantageous for yield in the southern US states; however, yield advantages can disappear when the environment favors lodging of plants too large to support themselves. Determinate types are most likely not the ideal trait necessary for soybean success in the tropics, especially if the first reproductive nodes are too close to the ground, which would lead to those pods rotting. Semi-determinate types hold promise to be successful; however, due to different allelic combinations, there are still several phenotypes that need to be assayed for optimum adaptation.

Pod Shatter

Pod shatter is an ancestral characteristic from *Glycine soja* that facilitates seed dispersal. Due to the heavy yield losses shatter causes, it was one of the first traits selected against during cultivation of soybean (Hymowitz 1970, Fuller et al. 2014). Several QTLs were discovered that influenced the shatter phenotype (Saxe et al. 1996, Bailey et al. 1997, Liu et al. 2007, Suzuki et al. 2009). A gene *SHAT-1* was also cloned that plays a role in

shatter resistance, however its use in current breeding programs has not been demonstrated (Dong et al. 2014).

In 2014, Funatsuki et al. cloned a gene, *Pdh1*, that is responsible for ~45% of the pod shattering phenotype (Bailey et al. 1997, Funatsuki et al. 2014). When the gene is nonfunctional, *pdh1*, the pod stays intact after maturity. When it is functional *Pdh1,* the pod walls undergo a strong torsion force after dehiscence, twisting the pod walls open and causing the shatter phenotype (Funatsuki et al. 2014). Selection against the functional allele of *Pdh1* is necessary in all breeding programs to eliminate potential, unnecessary yield losses due to pod shatter.

Current Breeding Practices in Africa

The International Institute for Tropical Agriculture (IITA) is one of leading drivers of soybean development in Africa. Research on soybean started in the 1970s to overcome production problems such as low yield, low seed viability, pod shattering, and disease (Tefera et al. 2010). IITA has released numerous varieties of early, middle, or late maturities to Nigeria, Ghana, Benin, Togo, and the Democratic Republic of the Congo (Tefera et al. 2010). One constraint of soybean production in several countries in Africa is the lack of genetic diversity or germplasm available. Only a handful of varieties have been released in Ghana since 1990: Salintuya-1, Quarshie, Jenguma, Salintuya-2 to name a few (Appiah-Kubi et al. 2014). The source of genetic materials IITA utilized to build its germplasm collection is unknown, however genotype by sequencing analysis of 298 IITA breeding lines indicate that there exists as much diversity in those lines as varieties from the USA, Canada, and Brazil (PESSOA FILHO et al. 2016). Regardless of the

limited number of variety releases, newly released varieties continue to see a yield increase (Appiah-Kubi et al. 2014).

New knowledge surrounding the genetic mechanisms behind adaptation of soybean to tropical environments will help strengthen local breeders' efforts to improve their programs by providing them insights on their germplasm currently available. It will also facilitate the introduction of traits from unadapted lines by selecting for genes that are necessary for their environment.

Conclusion

Improving soybean yields is potentially one way to help lift African smallholder farmers out of poverty and food insecurity. Improving maturity adaptation, photoperiod sensitivity, pod shatter, and stem development is only a small piece of a large picture to create elite cultivars for tropical climates. It will, however, provide a strong background to allow local breeders to add necessary advantageous traits, such as disease resistance, mineral utilization, etc. This knowledge may also help breeders determine the an optimal season length to be most beneficial for their local weather pattern and to reduce risk. The culmination of experiments reviewed here show that it is indeed possible to expand elite temperate soybean cultivars to tropical areas. However, there is much ambiguity surrounding genetic control of the desirable phenotype. Currently in tropical Africa, the American classification system of maturity groups cannot be applied. To optimize ideal environmental adaptation, the long juvenile trait needs to be utilized to allow for delayed flowering in short daylengths. When implemented into breeding programs, it has been shown that this trait, although it does delay flowering compared to the conventional

parent, may have background effects that cause earlier flowering than the long juvenile donor parent. In maturity group varieties that are greater than III, the long juvenile trait has been demonstrated in a short daylength to be controlled by a single Mendelian recessive gene. Since there are multiple sources of the long juvenile trait that may have different effects on delayed flowering in short days, this alleles need to be tested in different low latitude environments to understand their effect on flowering time and season length. Once the optimal allele of the long juvenile trait is determined for a specific environment, breeders can select for the correct tropical variety for their specific latitude by determining the correct allele composition of *E1, E2, E3,* and *J* depending on the desired season length.

REFERENCES

(2000). "Restrictions on Genetically Modified Organisms: International Protocols." from http://www.loc.gov/law/help/restrictions-on-gmos/international-protocols.php.

Abate, T., A. D. Alene, D. Bergvinson, B. Shiferaw, S. Silim, A. Orr and S. Asfaw (2012). Tropical grain legumes in Africa and south Asia: knowledge and opportunities, International Crops Research Institute for the Semi-Arid Tropics.

Abate, T. and A. Orr (1981). "Research and development for tropical legumes: Towards a knowledge-based strategy." Journal of SAT Agricultural Research **10**: 1-12.

Alene, A., O. Coulibaly and T. Abdoulaye (2012). "The world cowpea and soybean economies: Facts, trends, and outlook." Lilongwe, Malawi: International Institute of Tropical Agriculture.

Appiah-Kubi, D., J. Asibuo, M. Quain, A. Oppong and R. Akromah (2014). "Diversity studies on soybean accessions from three countries." Biocatalysis and Agricultural Biotechnology **3**(2): 198-206.

Bailey, M., M. Mian, T. Carter, D. Ashley and H. Boerma (1997). "Pod dehiscence of soybean: identification of quantitative trait loci." Journal of Heredity **88**(2): 152-154.

Bandillo, N. B., J. E. Anderson, M. B. Kantar, R. M. Stupar, J. E. Specht, G. L. Graef and A. J. Lorenz (2017). "Dissecting the Genetic Basis of Local Adaptation in Soybean." Scientific reports **7**(1): 17195.

Bäurle, I. and C. Dean (2006). "The timing of developmental transitions in plants." Cell **125**(4): 655-664.

Bernard, R. (1972). "Two genes affecting stem termination in soybeans." crop Science **12**(2): 235-239.

Bernard, R. L. (1971). "Two major genes for time of flowering and maturity in soybeans." Crop Sci **11**: 242-244.

Bonato, E. R. and N. A. Vello (1999). "E6, a dominant gene conditioning early flowering and maturity in soybeans." Genetics and Molecular Biology **22**(2): 229-232.

Buzzell, R. (1971). "Inheritance of a soybean flowering response to fluorescent-daylength conditions." Canadian Journal of Genetics and Cytology **13**(4): 703-707.

Buzzell, R. and H. Voldeng (1980). "Inheritance of insensitivity to long daylength." Soybean Genet. Newsl **7**: 26-29.

Cairo, C. A., J. Stein, L. Delgado, S. Bortolotti, S. A. Guelman, J. P. A. Ortiz and E. N. Morandi (2002). "Tagging the juvenile locus in soybean [Glycine max (L.) Merr.] with molecular markers." Euphytica **124**(3): 387-395.

Carpentieri-Pipolo, V., L. Alves de Almeida and R. Afonso de S. Kiihl (2002). "Inheritance of a long juvenile period under short-day conditions in soybean." Genet Mol Biol **25**(4): 463-469.

Carpentieri-Pípolo, V., L. A. de Almeida, R. A. de Souza Kiihl and C. A. Rosolem (2000). "Inheritance of long juvenile period under short day conditions for the BR80-6778 soybean (Glycine max (L.) Merrill) line." Euphytica **112**(2): 203-209.

Cober, E., J. Tanner and H. Voldeng (1996). "Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines." Crop science **36**(3): 601-605.

Cober, E. R. (2011). "Long juvenile soybean flowering responses under very short photoperiods." Crop science **51**(1): 140-145.

Cober, E. R., S. J. Molnar, M. Charette and H. D. Voldeng (2010). "A new locus for early maturity in soybean." Crop science **50**(2): 524-527.

Cober, E. R. and H. D. Voldeng (2001). "A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*." Crop Sci **41**: 698-701.

Collinson, S., R. Summerfield, R. Ellis and E. Roberts (1993). "Durations of the photoperiod-sensitive and photoperiod-insensitive phases of development to flowering in four cultivars of soyabean [Glycine max (L.) Merrill]." Annals of Botany **71**(5): 389-394.

Cregan, P. and E. Hartwig (1984). "Characterization of flowering response to photoperiod in diverse soybean genotypes." Crop Science **24**(4): 659-662.

de Pesquisa, O. P. (1986). "Recomendacoes tecnicas para a cultura da soja no Parana 1986/87." OCEPAR. Boletim Tecnico, 19; EMBRAPA-CNPSo. Documentos.

Destro, D., V. Carpentieri-Pipolo, R. Afonso de S. Kiihl and L. Alves de Almeida (2001). "Photoperiodism and genetic control of the long juvenile period in soybean: a review." Crop Breed Appl Biot **1**(1): 72-92.

Dissanayaka, A., T. O. Rodriguez, S. Di, F. Yan, S. M. Githiri, F. R. Rodas, J. Abe and R. Takahashi (2016). "Quantitative trait locus mapping of soybean maturity gene <i>E5</i>." Breeding Science **66**(3): 407-415.

Dong, Y., X. Yang, J. Liu, B.-H. Wang, B.-L. Liu and Y.-Z. Wang (2014). "Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean." Nature communications **5**.

Dunford, R., S. Griffiths, V. Christodoulou and D. Laurie (2005). "Characterisation of a barley (Hordeum vulgare L.) homologue of the Arabidopsis flowering time regulator GIGANTEA." Theoretical and Applied Genetics **110**(5): 925-931.

EMBRAPA. (2014). "Public Consultation on North Brazilian Soils." from http://www.sisolos.cnptia.embrapa.br/.

Fisher, J. (1963). "The effects of short days on fruitset as distinct from flower formation in soybeans." Canadian Journal of Botany **41**(6): 871-873.

Fowler, S., K. Lee, H. Onouchi, A. Samach, K. Richardson, B. Morris, G. Coupland and J. Putterill (1999). "GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in Arabidopsis and encodes a protein with several possible membrane-spanning domains." The EMBO journal **18**(17): 4679-4688.

Fuller, D. Q., T. Denham, M. Arroyo-Kalin, L. Lucas, C. J. Stevens, L. Qin, R. G. Allaby and M. D. Purugganan (2014). "Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record." Proceedings of the National Academy of Sciences **111**(17): 6147-6152.

Funatsuki, H., M. Suzuki, A. Hirose, H. Inaba, T. Yamada, M. Hajika, K. Komatsu, T. Katayama, T. Sayama and M. Ishimoto (2014). "Molecular basis of a shattering resistance boosting global dissemination of soybean." Proceedings of the National Academy of Sciences **111**(50): 17797-17802.

Garner, W. W. and H. A. Allard (1920). "Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants." Monthly Weather Review **48**(7): 415-415.

Hanson, W. (1961). "Effect of calcium and phosphorus nutrition on genetic recombination in the soybean." Crop Science **1**(5): 384-384.

Hartwig, E. E. and C. J. Edwards (1970). "Effects of morphological characteristics upon seed yield in soybeans." Agronomy Journal **62**(1): 64-65.

Hartwig, E. E. and R. A. Kiihl (1979). "Identification and utilization of a delayed flowering character in soybeans for short-day conditions." Field Crops Research **2**: 145-151.

Hymowitz, T. (1970). "On the domestication of the soybean." Economic Botany **24**(4): 408-421.

IITA. (2014). "Soybean." from http://www.iita.org/soybean.

James, A. and R. Lawn (2011). "Application of physiological understanding in soybean improvement. II. Broadening phenological adaptation across regions and sowing dates." Crop and Pasture Science **62**(1): 12-24.

Kong, F., B. Liu, Z. Xia, S. Sato, B. M. Kim, S. Watanabe, T. Yamada, S. Tabata, A. Kanazawa, K. Harada and J. Abe (2010). "Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean." Plant Physiol **154**(3): 1220-1231.

Kong, F., H. Nan, D. Cao, Y. Li, F. Wu, J. Wang, S. Lu, X. Yuan, E. R. Cober, J. Abe and B. Liu (2014). "A new dominant gene *E9* conditions early flowering and maturity in soybean." Crop Sci.

Langewisch, T., J. Lenis, G.-L. Jiang, D. Wang, V. Pantalone and K. Bilyeu (2017). "The development and use of a molecular model for soybean maturity groups." BMC plant biology **17**(1): 91.

Lawn, R. and A. James (2011). "Application of physiological understanding in soybean improvement. I. Understanding phenological constraints to adaptation and yield potential." Crop and Pasture Science **62**(1): 1-11.

Li, X., C. Fang, M. Xu, F. Zhang, S. Lu, H. Nan, T. Su, S. Li, X. Zhao and L. Kong (2017). "Quantitative Trait Locus Mapping of Soybean Maturity Gene E6." Crop science **57**(5): 2547-2554.

LIMA, W. F., J. F. F. D. TOLEDO, C. A. A. ARIAS and M. F. D. OLIVEIRA (2000). "Stability of soybean yield through different sowing periods." Pesquisa Agropecuária Brasileira **35**: 2181-2189.

Liu, B., T. Fujita, Z.-H. Yan, S. Sakamoto, D. Xu and J. Abe (2007). "QTL mapping of domestication-related traits in soybean (Glycine max)." Annals of Botany **100**(5): 1027-1038.

Liu, B., A. Kanazawa, H. Matsumura, R. Takahashi, K. Harada and J. Abe (2008). "Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene." Genetics **180**(2): 995-1007.

Liu, B., S. Watanabe, T. Uchiyama, F. Kong, A. Kanazawa, Z. Xia, A. Nagamatsu, M. Arai, T. Yamada, K. Kitamura, C. Masuta, K. Harada and J. Abe (2010). "The soybean stem growth habit gene *Dt1* is an ortholog of Arabidopsis TERMINAL FLOWER1." Plant Physiol **153**(1): 198-210.

Lu, S., Y. Li, J. Wang, P. Srinives, H. Nan, D. Cao, Y. Wang, J. Li, X. Li and C. Fang (2015). "QTL mapping for flowering time in different latitude in soybean." Euphytica **206**(3): 725-736.

Lu, S., X. Zhao, Y. Hu, S. Liu, H. Nan, X. Li, C. Fang, D. Cao, X. Shi and L. Kong (2017). "Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield." Nature Genetics **49**(5): 773-779.

Masuda, T. and P. D. Goldsmith (2009). "World soybean production: area harvested, yield, and long-term projections." International Food and Agribusiness Management Review **12**(4): 143-162.

Mbanya, W. (2011). "Assessment of the Constraints in Soybean Production: A Case of Northern Region, Ghana." Journal of Developments in Sustainable Agriculture **6**(2): 199-214.

McBlain, B. and R. Bernard (1987). "A new gene affecting the time of flowering and maturity in soybeans." Journal of Heredity **78**(3): 160-162.
McBlain, B. A. and R. L. Bernard (1987). "A new gene affecting time of flowering and matuirty in soybeans." J Hered **78**(3): 160-162.

Pessoa Filho, M., D. Torkamaneh, A. d. L. Passianotto, M. de Oliveira, R. Abdelnoor, R. Nelson, F. Belzile and B. Diers (2016). Genotyping-by-sequencing of soybean breeding lines from Africa, Brazil, Canada, and the USA. Embrapa Cerrados-Resumo em anais de congresso (ALICE), In: Plant & Animal Genome Conference, 24., 2016, San Diego, CA.[Abstracts...]. San Diego:[sn], 2016.

Ping, J., Y. Liu, L. Sun, M. Zhao, Y. Li, M. She, Y. Sui, F. Lin, X. Liu and Z. Tang (2014). "Dt2 is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean." The Plant Cell Online **26**(7): 2831-2842.

Ray, J., K. Hinson, E. Mankono and M. Malo (1995). "Genetic Control of a Lon-Juvenile Trait in Soybean." Crop Science **35**: 1001-1006.

Robson, F., M. M. R. Costa, S. R. Hepworth, I. Vizir, P. H. Reeves, J. Putterill and G. Coupland (2001). "Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants." The Plant Journal **28**(6): 619-631.

Samach, A., H. Onouchi, S. E. Gold, G. S. Ditta, Z. Schwarz-Sommer, M. F. Yanofsky and G. Coupland (2000). "Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis." Science **288**(5471): 1613-1616.

Samanfar, B., S. J. Molnar, M. Charette, A. Schoenrock, F. Dehne, A. Golshani, F. Belzile and E. R. Cober (2017). "Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean." Theoretical and Applied Genetics **130**(2): 377-390.

Saxe, L., C. Clark, S. Lin and T. Lumpkin (1996). "Mapping the pod-shattering trait in soybean." Soybean genetics newsletter (USA).

Schmutz, J., S. B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D. L. Hyten, Q. Song, J. J. Thelen, J. Cheng, D. Xu, U. Hellsten, G. D. May, Y. Yu, T. Sakurai, T.

Umezawa, M. K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X. C. Zhang, K. Shinozaki, H. T. Nguyen, R. A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R. C. Shoemaker and S. A. Jackson (2010). "Genome sequence of the palaeopolyploid soybean." Nature **463**(7278): 178-183.

Sinclair, T., S. Kitani, K. Hinson, J. Bruniard and T. Horie (1991). "Soybean flowering date: linear and logistic models based on temperature and photoperiod." Crop Science **31**(3): 786-790.

Sinclair, T. R. and K. Hinson (1992). "Soybean flowering in response to the long-juvenile trait." Crop science **32**(5): 1242-1248.

Spehar, C. (1995). "Diallel analysis for mineral element absorption in tropical adapted soybeans [Glycine max (L.) Merrill]." Theoretical and Applied Genetics **90**(5): 707-713.
Suzuki, M., K. Fujino and H. Funatsuki (2009). "A major soybean QTL, qPDH1, controls pod dehiscence without marked morphological change." Plant production science **12**(2): 217-223.

Swaminathan, K., K. Peterson and T. Jack (2008). "The plant B3 superfamily." Trends in plant science **13**(12): 647-655.

Tasma, I. M., L. L. Lorenzen, D. E. Green and R. C. Shoemaker (2001). "Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean." Mol Breeding **8**: 25-35.

Tefera, H., B. Asafo-Adjei and K. E. Dashiell (2010). "Breeding progress for grain yield and associated traits in medium and late maturing promiscuous soybeans in Nigeria." Euphytica **175**(2): 251-260.

Thompson, J. A., R. L. Bernard and R. L. Nelson (1997). "A third allele at the soybean dt1 locus." Crop science **37**(3): 757-762.

Tian, Z., X. Wang, R. Lee, Y. Li, J. E. Specht, R. L. Nelson, P. E. McClean, L. Qiu and J. Ma (2010). "Artificial selection for determinate growth habit in soybean." Proceedings of the National Academy of Sciences **107**(19): 8563-8568.

Voldeng, H. and G. Saindon (1991). "Registration of seven long-daylength insensitive soybean genetic stocks." Crop Sci **31**: 1399.

Watanabe, S., K. Harada and J. Abe (2012). "Genetic and molecular bases of photoperiod responses of flowering in soybean." Breed Sci **61**(5): 531-543.

Watanabe, S., R. Hideshima, Z. Xia, Y. Tsubokura, S. Sato, Y. Nakamoto, N. Yamanaka, R. Takahashi, M. Ishimoto and T. Anai (2009). "Map-based cloning of the gene associated with the soybean maturity locus E3." Genetics **182**(4): 1251-1262.

Watanabe, S., R. Hideshima, Z. Xia, Y. Tsubokura, S. Sato, Y. Nakamoto, N. Yamanaka, R. Takahashi, M. Ishimoto, T. Anai, S. Tabata and K. Harada (2009). "Map-based cloning of the gene associated with the soybean maturity locus *E3*." Genetics **182**(4): 1251-1262.

Watanabe, S., T. Tajuddin, N. Yamanaka, M. Hayashi and K. Harada (2004). "Analysis of QTLs for reproductive development and seed quality traits in soybean using recombinant inbred lines." Breeding science **54**(4): 399-407.

Watanabe, S., Z. Xia, R. Hideshima, Y. Tsubokura, S. Sato, N. Yamanaka, R. Takahashi, T. Anai, S. Tabata, K. Kitamura and K. Harada (2011). "A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering." Genetics **188**(2): 395-407.

Weiss, M. G. (1970). "Genetic linkage in soybeans: Linkage group I." Crop Science **10**(1): 69-72.

Whigham, D. K. and H. C. Minor (1978). "Agronomic characteristics and environmental stress." Soybean physiology, agronomy, and utilization. Academic Press, New York: 77-112.

Woodworth, C. (1923). "Inheritance of growth habit, pod color, and flower color in soybeans." Agronomy journal.

Xia, Z., S. Watanabe, T. Yamada, Y. Tsubokura, H. Nakashima, H. Zhai, T. Anai, S. Sato, T. Yamazaki and S. Lü (2012). "Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering." Proceedings of the National Academy of Sciences **109**(32): E2155-E2164.

Xia, Z., S. Watanabe, T. Yamada, Y. Tsubokura, H. Nakashima, H. Zhai, T. Anai, S. Sato, T. Yamazaki, S. Lu, H. Wu, S. Tabata and K. Harada (2012). "Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering." Proc Natl Acad Sci U S A **109**(32): E2155-2164.

Xia, Z., H. Zhai, B. Liu, F. Kong, X. Yuan, H. Wu, E. R. Cober and K. Harada (2012). "Molecular identification of genes controlling flowering time, maturity, and photoperiod response in soybean." Plant Syst Evol **298**: 1217-1227.

Xu, M., Z. Xu, B. Liu, F. Kong, Y. Tsubokura, S. Watanabe, Z. Xia, K. Harada, A. Kanazawa, T. Yamada and J. Abe (2013). "Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean." BMC Plant Biol **13**: 91.

Yamanaka, N., S. Ninomiya, M. Hoshi, Y. Tsubokura, M. Yano, Y. Nagamura, T. Sasaki and K. Harada (2001). "An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion." DNA research **8**(2): 61-72.

Yamanaka, N., S. Watanabe, K. Toda, M. Hayashi, H. Fuchigami, R. Takahashi and K. Harada (2005). "Fine mapping of the FT1 locus for soybean flowering time using a residual heterozygous line derived from a recombinant inbred line." Theoretical and Applied Genetics **110**(4): 634-639.

Yue, Y., N. Liu, B. Jiang, M. Li, H. Wang, Z. Jiang, H. Pan, Q. Xia, Q. Ma and T. Han (2017). "A single nucleotide deletion in J encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean." Molecular plant **10**(4): 656-658.

Zagotta, M. T., K. A. Hicks, C. I. Jacobs, J. C. Young, R. P. Hangarter and D. R. Meeks-Wagner (1996). "The Arabidopsis ELF3 gene regulates vegetative photomorphogenesis and the photoperiodic induction of flowering." The Plant Journal **10**(4): 691-702.

Zhai, H., S. Lu, Y. Wang, X. Chen, H. Ren, J. Yang, W. Cheng, C. Zong, H. Gu, H. Qiu, H. Wu, X. Zhang, T. Cui and Z. Xia (2014). "Allelic variations at four major maturity *E* genes and transcriptional abundance of the *E1* gene are associated with flowering time and maturity of soybean cultivars." PLoS One **9**(5): e97636.

Zhang, L., S. Kyei-Boagen, J. Zhang, M. Zhang, T. Freeland and C. Watson, Jr. (2007). "Modifications of optimum adaptation zones for soybean maturity groups in the USA." Crop Management **6**.

Zhao, C., R. Takeshima, J. Zhu, M. Xu, M. Sato, S. Watanabe, A. Kanazawa, B. Liu, F. Kong and T. Yamada (2016). "A recessive allele for delayed flowering at the soybean maturity locus E9 is a leaky allele of FT2a, a FLOWERING LOCUS T ortholog." BMC plant biology **16**(1): 20.

Zhou, Z., Y. Jiang, Z. Wang, Z. Gou, J. Lyu, W. Li, Y. Yu, L. Shu, Y. Zhao and Y. Ma (2015). "Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean." Nature Biotechnology **33**(4): 408-414.

CHAPTER TWO

Molecular Tools for Detecting *Pdh1* Can Improve Soybean Breeding Efficiency by

Reducing Yield Losses Due to Pod Shatter

SUMMARY

Pod shattering is an ancestral trait that is necessary for seed dispersal, however can have substantial yield losses in cultivated soybean. During the domestication process, American breeders virtually eliminated the shatter phenotype from released varieties, but in other countries, such as Ghana, shatter persists. The objective of our research is to find a molecular tool to identify shatter, validate its usefulness, and apply this knowledge to identify shattering potential in parental lines. Funatsuki et al. discovered a gene, *Pdh1,* that plays a crucial role in determining the shatter phenotype. From these data, we developed a marker to detect alleles of the *Pdh1* gene. In addition, we performed a Genome Wide Association Analysis Study using the *Pdh1* alleles as a phenotype and identified an associated marker in the SoySNP50K array: ss715624199. After proving its accuracy, we evaluated soybean accessions from the GRIN National Plant Germplasm System (GRIN-NPGS) with recorded shatter scores and determined the impact of the *Pdh1* gene on early and late shattering. After analyzing the 16,250 soybean accessions in the GRIN-NPGS with SoySNP50K data we predict that nearly 50% have the shatter allele of *Pdh1*. After conducting preliminary yield tests in Ghana of a population segregating for both *E1*, an important maturity gene, and *Pdh1* we determined that the ability to shatter had a more significant effect on yield than maturity. Analysis of *Pdh1* in Ghanaian released varieties shows that ~30% contain the shatter allele. Finally, we analyzed 288 lines from the African Germplasm collection and determined ~20% of all lines have the potential to shatter. We recommend that this marker be used to predict shatter potential in parental lines in breeding programs to prevent possible yield losses.

INTRODUCTION

Soybean [*Glycine max* Merr. (L.)] production is expanding worldwide and predicted to continue increasing by 2.2% until 2030 (Masuda and Goldsmith 2009). This rising demand is creating economic opportunities to open or expand soybean production. Africa has increased soybean production over the past several decades. From 1970 to 2014, production in East Africa has increased from 28,711 metric tons to 535,779 metric tons, and in West Africa production has increased from 59,200 metric tons to 801,421 metric tons and has also seen a 300% increase in yield during that time (FAOSTAT 2014). However, these numbers are still below the highest producers in the world. As Africa strives to be a major producer, they are using exotic germplasm to improve the adaptation of soybean to their environments. Breeders in major producing countries such as the United States are returning to breeding with *Glycine soja* and soybean landraces to attempt to improve yield gains more quickly. Both of these breeding techniques carry the risk of reintroducing unfavorable ancestral traits.

Soybean was first domesticated in northern China ~5,000 years ago (Hymowitz 1970, Carter et al. 2004). There were numerous traits that were selected by humans from its ancestor *Glycine soja* to facilitate cultivation and yield gains such as upright stem architecture, larger seed size, and pod shatter resistance (Liu et al. 2007, Zhou et al. 2015). Pod shatter is a means for seed dispersal which is advantageous for wild plants but can be devastating to yield in cultivated grain crops. Shatter resistance is considered to be a “sine qua non” in cultivated crops and evolved faster in legumes than other domestication traits such as larger seed size (Fuller et al. 2014). As *G. soja* and other less

domesticated lines are being utilized in breeding programs, it is useful to have a molecular tool to identify pod shatter resistance preemptively in parental lines or detect it in progeny.

The genetic mechanism behind pod shatter resistance had been elusive although several QTLs were discovered (Saxe et al. 1996, Bailey et al. 1997, Liu et al. 2007, Suzuki et al. 2009, Gao and Zhu 2013). A gene *SHAT 1-5* was found that controls secondary cell wall biosynthesis and plays a role in preventing shatter, but the importance of selection for pod shatter resistance alleles of this gene during domestication and modern soybean variety development has not been demonstrated (Dong et al. 2014). *Pdh1*, which encodes a dirigent (DIR)-like protein, was recently cloned and is responsible for a large effect on the shattering phenotype by controlling pod wall torsion after dehiscence (Funatsuki et al. 2014). Pod shattering is observed when the wild type, functional allele of *Pdh1* is present. When the gene is nonfunctional (*pdh1*), the pod remains intact. This gene has been noted as having a ~45% influence on the shattering phenotype (Bailey et al. 1997).

Here we report the development of two molecular tools to ascertain the resistant and susceptible alleles of *Pdh1*. One is a perfect molecular marker and the other is an associated marker from the SoySNP50K array (Song et al. 2013). Using these tools, we determined the frequency and status of the shatter-susceptible allele *Pdh1* in the soybean germplasm collections maintained by the USDA GRIN and the International Institute for Tropical Agriculture (IITA) (Tefera et al. 2010). The impact of the *Pdh1* allele status was correlated with publicly available shatter score data from the USDA GRIN

collection. The *Pdh1* allele data is available on the SoyBase GRIN Data Explorer (https://soybase.org/grindata/).

MATERIALS AND METHODS

Pdh1 SimpleProbe Assay

A SimpleProbe assay was developed to distinguish *Pdh1* and *pdh1* alleles by identifying the T/A SNP (Gm16: 29,601,807 Wm82.a1.v1) with a melting curve analysis. *Pdh1*SNP PCR primers (F: 5'-GCCCTCGTTGTGTTCTTCAT-3', R: 5'-GCGTTGCTTCCGTTGTAGAT-3') were designed by Funatsuki et al (Funatsuki et al. 2014) and amplify a 125-bp region where the T/A SNP is found. The SimpleProbe oligonucleotide (Fluorescein-SPC-CATGCACCATGC**A**AGCACTTAGTC-Phosphate) was designed to the *Pdh1* sequence on the sense strand using the LightCycler Probe Design software (Roche Applied Science, Indianapolis, IN). PCR reactions were 20 µl and included the DNA template, 0.5 µM reverse primer *Pdh1*SNPr, 0.2 µM forward primer *Pdh1*SNPf, 0.2 µM SimpleProbe, buffer (40 mM Tricine- KOH [pH 8.0], 16 mM $MgCl_2$, 3.75 µg ml^{-1} BSA), 5% DMSO, 200 µM dNTPs, and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). PCR reactions were run on the LightCycler 480 real-time PCR instrument (Roche Applied Science, Indianapolis, IN). Reactions were denatured at 95°C for 3 minutes, and then in each cycle denatured at 95°C for 20 seconds, primers annealed at 60°C for 20 seconds, and products elongated at 72°C for 20 seconds for 45 cycles. After amplification was completed, a melting curve

was conducted from 55-70°C. The *pdh1* shatter-resistant peak was observed at ~61°C, and the *Pdh1* shatter-susceptible peak was observed at ~66°C. Heterozygous *Pdh1/pdh1* samples produced both peaks.

Genome Wide Association Study (GWAS) with *Pdh1* allele as phenotype

We found a marker in the SoySNP50K Beadchip (Song et al. 2013), ss715624199, that is able to accurately predict the allele status of *Pdh1* in 19,344 entries in the USDA Soybean Germplasm Collection. First, we determined the allele status of the *Pdh1* causative T/A SNP (Gm16: 29601807 Wm82.a1.v1/Gm16: 29944393 Wm82.a2.v1) in 474 of the whole-genome sequenced (WGS) lines from the Zhou 302 resequencing data set (Zhou et al. 2015) and the USB data sets (Appendix 1/Supplemental Table 1).

To generate the sequence information for the USB datasets, 350 soybean re-sequencing lines were analyzed using the Pegasus genomic variations workflow (PGen) (Liu et al. 2016) running on the Extreme Science and Engineering Discovery Environment (XSEDE). SNPs were called using HaplotypeCaller from GATK 3.0 and filtered with 'QD < 26.0 || FS > 60.0 || MQ < 40.0'. Passed SNPs were then annotated using SnpEff 3.0 and causative SNPs were extracted within the 5-kb upstream and downstream regions around *Pdh1* using SnpSift. Whole-genome Zhou 302 and USB variants data with annotations are also available through the SoyKB NGS resequencing browser (http://soykb.org/NGS_Resequence/NGS_index.php) (Joshi et al. 2012, Joshi et al. 2013, Joshi et al. 2017).

To create a phenotype file, these genotypic data were coded numerically for each Plant Introduction (PI) line, where the functional *Pdh1* allele A was coded as 1 and the nonfunctional *pdh1* allele T was coded as 2. To create a genotype file, 474 WGS lines also had SoySNP50K Beadchip data that was downloaded from SoyBase (https://www.soybase.org/dlpages/index.php). Both the phenotype and the genotype files were uploaded into Tassel 5.0 (Trait Analysis by aSSociation, Evolution, and Linkage) (Bradbury et al. 2007), and the non-compressed mixed-linear model analysis included PCA with 5 components and centered-IBS kinship matrix to account for population structure and relatedness (Bradbury, Zhang et al. 2007). A Manhattan plot was drawn to visualize any markers that were associated to the *Pdh1* causative SNP. The most highly associated marker was ss715624199 on chromosome 16, position 29,940,504 (Wm82.a2.v1)/29,567,918 (Wm82.a1.v1) (p= 4.32E-51).

Associated Marker Validation

To ensure that the associated marker was accurate at predicting a certain allele of *Pdh1*, we conducted F tests for each allele of the associated marker to determine the variance of predicting the correct *Pdh1* allele. The associated marker allele ‘T’ was coded as one, and its associated *pdh1* ‘A’ allele as well. The ss715624199 ‘A’ and *Pdh1* ‘G’ alleles were both coded as two. F tests were separately conducted for each allele of the marker to determine the variance of each marker allele to associate with the correct *Pdh1* allele. Variances were then subtracted from 1 and converted to percentages to determine accuracy.

DNA preparation for IITA lines

DNA extraction was performed on dried leaf punches (~10 mg) using the Qiagen DNeasy 96 Plant kit according to the manufacturer's protocol.

RESULTS

Prevalence of *Pdh1* in African Soybean Germplasm

In version Wm82.a1.v1 of the Williams 82 soybean genome reference sequence (https://soybase.org; Grant et al. 2010) *pdh1* is annotated as Glyma16g25580. The annotation does not accurately reflect the gene as described upon its cloning (Funatsuki et al. 2014). At the time of publication, *pdh1* was not shown on the 2nd genome assembly. The thymine to adenine causative SNP that creates a nonsense mutation from *Pdh1* to *pdh1* is on chromosome 16 at position 29,601,807 (Wm82.a1.v1) and position 29,944,393 (Wm82.a2.v1). We developed a real-time PCR-based perfect molecular marker assay for detecting *Pdh1* or *pdh1* by using primers for amplification of the region surrounding the causative SNP and a SimpleProbe (Funatsuki et al. 2014).

Using the *Pdh1* marker assay, we directly genotyped DNA from soybean germplasm from the International Institute of Tropical Agriculture (IITA) based in Ibadan, Nigeria. This institute has created and dispersed the majority of soybean germplasm currently

used on the African continent (Tefera et al. 2010). Of the 260 IITA soybean lines successfully assayed with the *Pdh1* marker, 20.7% contained the shatter allele (Appendix 1, Table 2).

Since IITA soybean germplasm is the source for many of the varieties throughout Africa, we also tested the seven released soybean varieties in Ghana for *Pdh1*. Two of the seven varieties have the shatter allele of *Pdh1* (Table 1).

Discovery of a *Pdh1*-associated marker

The USDA National Plant Germplasm System and Germplasm Resources Information Network (GRIN) database collection (www.ars-grin.gov) consists of publicly available soybean germplasm in the form of seeds from the wild ancestor *Glycine soja* as well as domesticated *Glycine max* landraces and cultivars collected or developed all over the world. One recent feature of the USDA soybean germplasm collection is genotype data from the Illumina Infinium SoySNP50K BeadChip available for 19,343 accessions (Song et al. 2013). Using a genome wide association study (GWAS), we identified marker ss715624199 from the SoySNP50K set as highly associated with the *Pdh1* allele (p= 4.32E-51). For the genotype file, we used the Illumina Infinium SoySNP50K BeadChip data for the 474 whole-genome sequenced lines that had those data available at SoyBase. Using the causative SNP position for *Pdh1*, we performed SNP calling for the same 474 whole-genome sequenced lines and used these data as the phenotype file. The most significantly associated marker was ss715624199 (position Chr16 29,597,918

Wm82.a1.v1 and Chr16 29,940,504 Wm82.a2.v1), located 3,889 base pairs from the causal *Pdh1* SNP on chromosome 16. (Figure 1).

To confirm ss715624199 as an associated marker that can accurately predict the *Pdh1* allele status, we evaluated the correspondence of the ss715624199 alleles aligned to the *Pdh1* alleles obtained from the SNP calling of the set of whole-genome sequenced lines with F tests (Table 2). When the thymine ('T') shatter-resistant allele *pdh1* was present, there was an adenine base ('A') for the ss715624199 SNP with 99% accuracy. The shatter-susceptible 'A' *Pdh1* allele corresponded to ss715624199 guanine ('G') with 93% accuracy. There were two accessions that were incorrectly predicted to be shatter-susceptible when *pdh1* alleles were determined to be present (PI548325 and PI578499A); there were eight accessions that were incorrectly predicted to be shatter-resistant when *Pdh1* alleles were determined to be present (PI157421, PI165563, PI437662, PI549018, PI567231, PI578493, PI587552, and PI095860). The complete list 474 accessions with the directly assayed *Pdh1* genotype and the ss715624199 genotype are provided (Supplemental Table 1/Appendix 1).

Prevalence of the Shatter Allele in the Soybean GRIN collection

We examined all 19,343 GRIN lines with SoySNP50K data to find the frequency of *Pdh1* alleles using the associated ss715624199 marker. We found that 9,146 lines have A present at the ss715624199 marker position, predicting that they will have the shatter-resistant allele *pdh1*. The other 10,197 lines are predicted to have the shattering allele of

Pdh1, meaning that 53% of the germplasm material in the GRIN has retained at least one of the genetic mechanisms for pod shatter.

Of the predicted shatter-susceptible lines in the GRIN, 82% of those lines are from China, Japan, or the Koreas. There are 6,290 Chinese accessions in the GRIN and 1,768 (28%) of those are predicted to have the shatter-susceptible allele of *Pdh1*. Of the 2,966 Japanese accessions, 2,542 or 86% of those lines have the predicted shatter allele of the associated marker. The Koreas, which includes data from both Koreas, have 87% of 3,633 entries in the collection with the predicted shatter allele. This is a contrast to the US entries in the GRIN where only 2% of the 2,254 accessions are predicted to contain the shatter allele of *Pdh1*.

<u>Correlation with Predicted Shatter Allele to GRIN Shatter Score Data</u>

We looked at the influence of the predicted *Pdh1* shatter allele on pod shatter by comparing the predicted alleles with reported GRIN shatter scores available on SoyBase (https://soybase.org/grindata/). Of the 19,343 GRIN lines that have SoySNP50K data for *Pdh1*, 14,376 soybean accessions also have reported shatter score data. Of these, 14,363 have early shatter score data and 12,024 have late shatter score data. "Early" shattering is assessed at harvest while "late" shattering is measured two weeks after harvest (Chen and Nelson 2004). For both early and late shattering phenotypes, the scoring is based on the estimated percentage of pods open on a five point scale, with values of "1" representing 0% or trace shattering and "5" representing 50% or more shattering. Approximately 53%

of the 19,343 GRIN accessions with data for the ss715624199 marker have the 'G' (*Pdh1*) allele. For each accession with shatter phenotype data, we categorized by the predicted status of the *Pdh1* allele and then evaluated the frequency of those lines for the early or late shatter phenotypes (Figure 2). About 80% of the lines in the *pdh1* category were scored "1" for early shatter, while over 50% of the lines in the *Pdh1* category scored "2" or higher for early shatter. The late shatter scores were distributed more broadly across the five-point scale. Nearly 75% of the lines in the *pdh1* category were scored "1" or "2" for late shatter. The distribution of late shatter scores for lines in the *Pdh1* category was most frequently "2", "3", or "4". In contrast to lines carrying the *pdh1* allele, only about 10% of lines harboring the *Pdh1* allele were scored "1" for late shatter.

Pdh1 in the GRIN Data Explorer

The predicted status of the *Pdh1* allele for the entire GRIN collection is now available on SoyBase (https://soybase.org/grindata/). Users can choose between all GRIN germplasm accessions or input a subset of desired accessions. Users can choose from *Pdh1* (imputed), *pdh1*, or Any in addition to other traits of interest. The predicted status of the *Pdh1* allele for each germplasm accession is then displayed and available for download.

DISCUSSION

Soybean is an important economic crop worldwide and is grown on ~6% of all arable land (Goldsmith 2008). Demand for soybean is increasing worldwide for both

commercial uses such as livestock feed but also for human nutrition as an inexpensive substitute for meat due to its high protein content (Singh and Singh 1992). The human nutrition aspect is important in developing countries where accessibility to animal protein is limited and malnutrition is persistent (Wansink and Cheong 2002). As soybean is introduced or expanded to new countries, breeders have used soybean lines from exotic germplasm and landraces to create locally adapted varieties.

In addition, US breeders have had difficulty making improvements in some areas of soybean production such as yield increase and disease resistance discovery due to the "bottleneck effect" where a limited number of parents were used to create all modern cultivars (Hyten et al. 2006). Soybean breeders throughout the world are returning to *Glycine soja* or landraces to increase the genetic diversity of their breeding programs. Although positive gains can be seen for certain traits, it can also bring a resurgence of negative traits that were previously eliminated. Pod shatter is a seed dispersal mechanism that originates from the *Glycine max* ancestor *Glycine soja* and has strong negative effects on yield (Hymowitz 1970). Historically, shatter was selected against through field observations, but as *Pdh1* affects late shatter predominately, an untrained or impatient breeder could select for a shatter susceptible line by harvesting too early. The effects of shatter would then be experienced by the farmer who may be unable to harvest immediately upon the crop's maturity. As soybean production becomes more prevalent in tropical environments, it is extremely important to ensure the shatter-susceptible allele of *Pdh1* is not present.

It is important to remember that the *Pdh1* gene accounts for only 42% of the pod shatter phenotype (Bailey et al. 1997). There are other genes that play minor effects, and the

environment also has a strong influence (Bandillo et al. 2017). For example, in Northern Ghana, soybean maturity occurs as the rainy season ends and temperatures can quickly rise to 40°C as pods are drying which places environmental stress on the pods to shatter. It is possible to observe total seed loss less than a week from maturity. In addition, while shattering-susceptible varieties were once favored for their ease of threshing, there is a movement away from those varieties as threshing becomes mechanized in developing countries. Fixation for *pdh1* alleles is one necessary step in the process of successful soybean cultivar development. The developed molecular tools for detecting *Pdh1* can help breeders, especially those in arid environments, preemptively protect their populations against shatter as much as possible. Good breeding practices such as correct shatter note taking for both early and late shattering are still important to ensure shatter-susceptible varieties are not released. As shown by the American breeding programs, it is possible to eliminate shattering almost entirely from cultivated soybean, and with the *Pdh1* tools described here combined with good breeding practice, it can be true for all breeding programs as well.

Here we report a tool available to all on SoyBase that allows the user to determine the predicted status of the *Pdh1* alleles for the vast majority of the USDA soybean germplasm collection. The *Pdh1* allele predictions are highly accurate (99% and 93%) with directly assayed genotypes of 474 accessions, suggesting our associated marker (ss715624199) is in strong linkage disequilibrium with the causative allele of *Pdh1*. Indeed, the physical distance between the associated marker and the causative allele is less than 4,000 bp. We did not determine the boundaries of the region of a potential selective sweep around *Pdh1*, but based on the number of highly associated 50KSNPs,

the *Pdh1* region appears to have undergone artificial selection (Figure 1) Associated markers for the *Pdh1* gene have been reported elsewhere by Bandillo (Bandillo et al. 2017) and Fang (Fang et al. 2017). Fang et al found a pod shatter associated marker, not from the SoySNP50K array, but through whole-genome sequencing; however, their most highly associated position was at Chr16 29,959,803 (Wm82.a2.v1), which is 15,410 bp away from the causal *Pdh1* SNP. Bandillo discovered an associated marker from the SoySNP50K array through mixed-model association between soybean landraces with SoySNP50K data and the corresponding climate data for each accession. They discovered a highly associated marker: ss715624379 which is at position Chr16 30,813,568 (Wm82.a1.v1) and Chr16 31,181,902 (Wm82.a2.v1) and showed a correlation between alleles of this SNP and shatter score data. We also discovered this marker in our analysis, however it was the 42nd most significant when using the casual *Pdh1* SNP as the phenotype (p= 1.72E-09). Conducting the same marker validation as described in this paper, we discovered that the Bandillo marker 'G' allele was 95.2% correct at predicting the shatter resistant 'A' *pdh1* allele, however the 'T' allele of this marker was only 77.3% correct at predicting the shatter 'G' *Pdh1* allele. Their discovered marker is located closer to the major flowering gene *FT2a* (Glyma16g26660/ Glyma.16g150700) (Kong et al. 2010), which has a starting position at Chr16 30,741,660 (Wm82.a1.v1.1)/ 31,109,999 (Wm82.a2.v1).

The molecular tools described here could be very helpful to breeders for either parental germplasm selection or progeny selection. Using the GRIN data explorer function on SoyBase, the user can select shatter-resistant parents utilizing the SoySNP50K associated marker. Due to the numerous landrace entries in the GRIN, our results show that over

50% of soybean accessions contain the shatter allele of *Pdh1*, predominately from Asian landraces, which is consistent with previous findings (Funatsuki et al. 2014). However, if a shatter susceptible accession has favorable traits, its progeny can be selected by genotype using the perfect molecular marker described here. By utilizing both of these tools, breeders have the option to select against shatter susceptibility, ensuring local farmers will not endure unnecessarily yield losses.

FIGURES

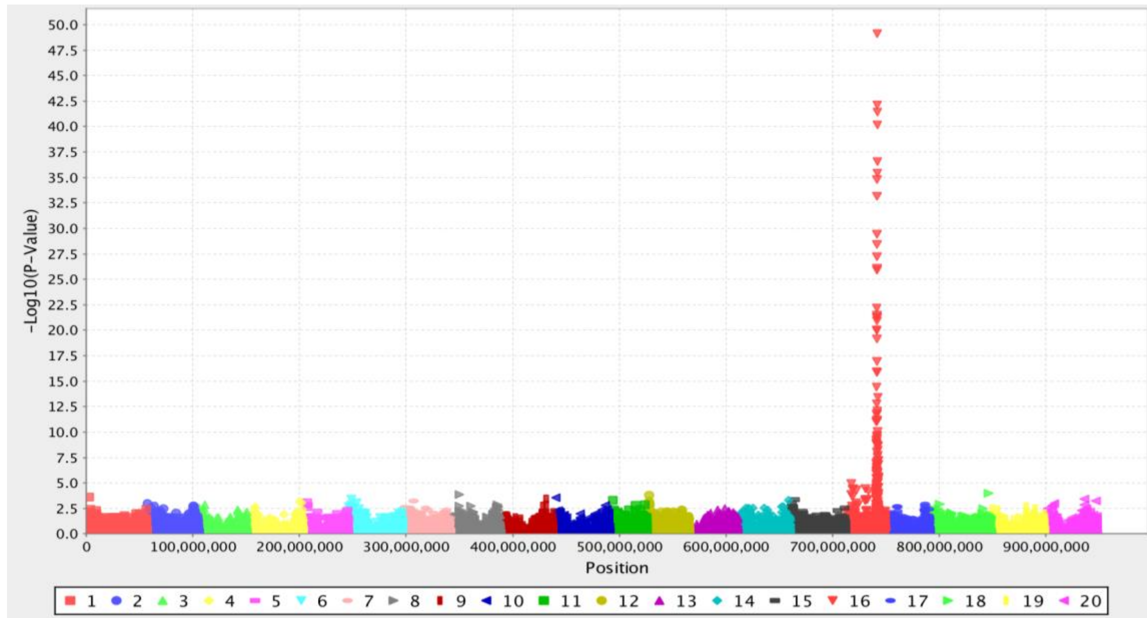

Figure 1: Manhattan plot of Genome Wide Association Analysis Study (GWAS) on a set of 474 sequenced soybean accessions using the *Pdh1* allele status as the phenotype and 50K SNP data as the genotype. The peak on chromosome 16 represents statistically associated region for the *Pdh1* alleles. The highest SNP is ss715624199 (~4 Kbp from *Pdh1)* and represents the most significantly associated SNP with *Pdh1*.

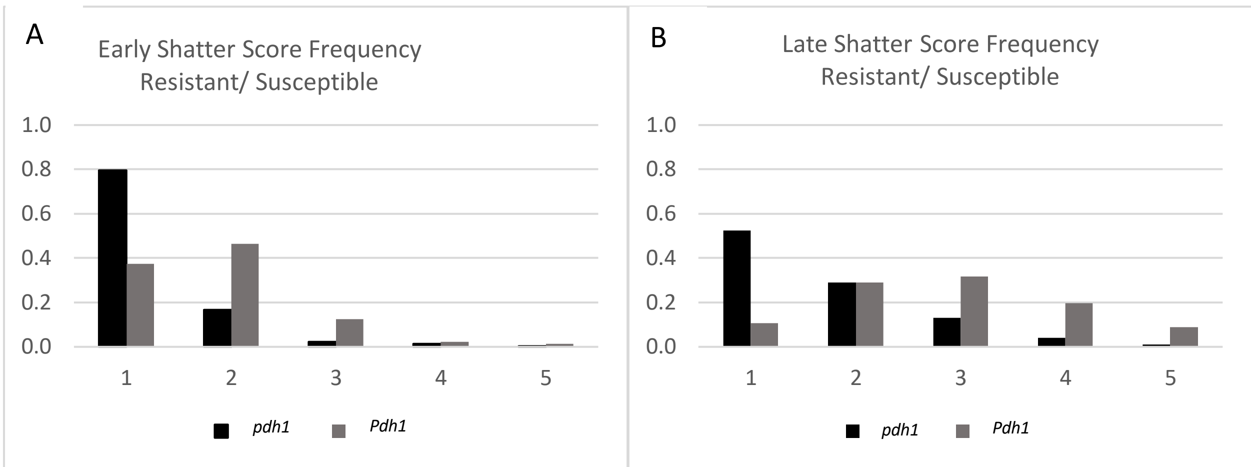

Figure 2: The frequency of shatter scores for each allele of *Pdh1*. *Pdh1* or shatter susceptible is show in grey and *pdh1* or shatter resistant is shown in black. Frequencies are shown as a percentage of total shatter score data. The shatter score scale ranges from 1-5 where 1= no shatter and 5=severe shatter. A. Early shatter scores for *Pdh1* and *pdh1*. B. Late shatter scores for *Pdh1* and *pdh1*.

TABLES

Table 1: Prevalence of *Pdh1* shatter allele in 7 released Ghanaian soybean varieties.

Variety Name	*Pdh1* Genotype	Shatter prediction based on genotype of *Pdh1*
Afayak	*pdh1*	Resistant
Jenguma	*pdh1*	Resistant
Quarshie	*pdh1*	Resistant
Salintuya-I	*pdh1*	Resistant
Salintuya-II	***Pdh1***	**Susceptible**
Songda	***Pdh1***	**Susceptible**
Suong-Pungu	*pdh1*	Resistant

Table 2: A table of F-test results for each allele of the *Pdh1* associated marker candidate ss715624199. For each allele of ss715624199 it was highly accurate for predicting the *Pdh1*

alpha =0.05	**Shatter Resistant**		**Shatter Susceptible (WT)**	
	ss715624199	***Pdh1***	**ss715624199**	***Pdh1***
Allele at SNP position	A	T	G	A
Mean	1	1.023	2	1.981
Variance	0	0.023	0	0.0183
Observations	350	350	108	108
Chance of correct *Pdh1* allele predicted		97.76%		98.17%

REFERENCES:

Bailey, M., et al. (1997). "Pod dehiscence of soybean: identification of quantitative trait loci." Journal of Heredity **88**(2): 152-154.

Bandillo, N. B., et al. (2017). "Dissecting the Genetic Basis of Local Adaptation in Soybean." Scientific reports **7**(1): 17195.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss and E. S. Buckler (2007). "TASSEL: software for association mapping of complex traits in diverse samples." Bioinformatics **23**(19): 2633-2635.

Carter, T. E., et al. (2004). "Genetic diversity in soybean." Soybeans: Improvement, production, and uses (soybeansimprove): 303-416.

Chen, Y. and R. L. Nelson (2004). "Genetic variation and relationships among cultivated, wild, and semiwild soybean." Crop Science **44**(1): 316-325.

Dong, Y., et al. (2014). "Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean." Nature communications **5**.

Fang, C., et al. (2017). "Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean." Genome biology **18**(1): 161.

Fuller, D. Q., et al. (2014). "Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record." Proceedings of the National Academy of Sciences **111**(17): 6147-6152.

Funatsuki, H., et al. (2014). "Molecular basis of a shattering resistance boosting global dissemination of soybean." Proceedings of the National Academy of Sciences **111**(50): 17797-17802.

Gao, M. and H. Zhu (2013). "Fine mapping of a major quantitative trait locus that regulates pod shattering in soybean." Molecular breeding **32**(2): 485-491.

Goldsmith, P. D. (2008). "Economics of soybean production, marketing and utilization." Soybeans Chemistry, Production, Processing, and Utilization: 117-150.

Grant, D., et al (2010). "SoyBase, the USDA-ARS soybean genetics and genomics database." Nucl. Acids Res. (2010) 38 (suppl 1): D843-D846. doi: 10.1093/nar/gkp798

Hymowitz, T. (1970). "On the domestication of the soybean." Economic Botany **24**(4): 408-421.

Hyten, D. L., et al. (2006). "Impacts of genetic bottlenecks on soybean genome diversity." Proceedings of the National Academy of Sciences **103**(45): 16666-16671.

Joshi, T., M. R. Fitzpatrick, S. Chen, Y. Liu, H. Zhang, R. Z. Endacott, E. C. Gaudiello, G. Stacey, H. T. Nguyen and D. Xu (2013). "Soybean knowledge base (SoyKB): a web resource for integration of soybean translational genomics and molecular breeding." Nucleic acids research **42**(D1): D1245-D1252.

Joshi, T., K. Patil, M. R. Fitzpatrick, L. D. Franklin, Q. Yao, J. R. Cook, Z. Wang, M. Libault, L. Brechenmacher and B. Valliyodan (2012). Soybean Knowledge Base (SoyKB): a web resource for soybean translational genomics. BMC genomics, BioMed Central.

Joshi, T., J. Wang, H. Zhang, S. Chen, S. Zeng, B. Xu and D. Xu (2017). The Evolution of Soybean Knowledge Base (SoyKB). Plant Genomics Databases, Springer**:** 149-159.

Kong, F., et al. (2010). "Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean." Plant Physiol **154**(3): 1220-1231.

Liu, B., et al. (2007). "QTL mapping of domestication-related traits in soybean (Glycine max)." Annals of Botany **100**(5): 1027-1038.

Liu, Y., S. M. Khan, J. Wang, M. Rynge, Y. Zhang, S. Zeng, S. Chen, J. V. M. dos Santos, B. Valliyodan and P. P. Calyam (2016). PGen: large-scale genomic variations analysis workflow and browser in SoyKB. BMC bioinformatics, BioMed Central.

Masuda, T. and P. D. Goldsmith (2009). "World soybean production: area harvested, yield, and long-term projections." International Food and Agribusiness Management Review **12**(4): 143-162.

Saxe, L., et al. (1996). "Mapping the pod-shattering trait in soybean." Soybean genetics newsletter (USA).

Singh, U. and B. Singh (1992). "Tropical grain legumes as important human foods." Economic Botany **46**(3): 310-321.

Song, Q., et al. (2013). "Development and evaluation of SoySNP50K, a high-density genotyping array for soybean." PLoS One **8**(1): e54985.

Suzuki, M., et al. (2009). "A major soybean QTL, qPDH1, controls pod dehiscence without marked morphological change." Plant production science **12**(2): 217-223.

Tefera, H., et al. (2010). "Breeding progress for grain yield and associated traits in medium and late maturing promiscuous soybeans in Nigeria." Euphytica **175**(2): 251-260.

Wansink, B. and J. Cheong (2002). "Taste profiles that correlate with soy consumption in developing countries." Pakistan Journal of Nutrition **1**(6): 276-278.

Zhou, Z., et al. (2015). "Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean." Nature Biotechnology **33**(4): 408-414.

CHAPTER THREE

Adaptation of Soybean to Tropical Environments for Smallholder Farmers

SUMMARY

There is a high demand for soybean in African countries, but available varieties are poor yielding. This can be partially attributed to inadequate adaptation of soybean to a tropical climate. Adaptation will require knowledge of allelic combinations of the characterized maturity genes: E1, E2, and E3; the long juvenile trait, and stem architecture. The long juvenile trait influences flowering time in short, 12 hour days, which characterize low latitudes. Stem architecture includes the determinate or indeterminate phenotypes controlled by the Dt1 gene. By understanding the influence of these genetic components on adaptation, it may be possible to control season length and improve yield greater than the currently available African varieties. To achieve the objective of understanding how these genes influence adaptation, six populations were initiated in which our genes of interest were segregating. 260 recombinant inbred lines were created across the six populations and were field tested in 5 locations in northern Ghana in 2016 and 2017. During this time phenotypes for flowering, maturity, height and shatter were noted. Our initial results from one population suggest that the long juvenile trait plays the most influential role on days to flower over E1. However, across populations segregating for the long juvenile trait these data also insinuate that that different alleles of this gene may also influence flowering phenotypes. Further analysis is being conducted to understand the effect of maturity gene allelic combinations on season. The combined knowledge of the genetic control of these traits will allow local Ghanaian breeders to produce varieties that can cater to the needs of small farmers in the north.

INTRODUCTION

Demand for soybean is increasing throughout Africa both for livestock feed and as a protein source to ameliorate malnutrition (Masuda and Goldsmith 2009), but sub-Saharan African soybean yields are lower than their potential (Abate and Orr 1981, Goldsmith 2008, Masuda and Goldsmith 2009, Abate et al. 2012, Alene et al. 2012, IITA 2014). There are many variables that are affecting soybean yields negatively, such as soil health, rainfall patterns, pod shatter potential, field preparation, and disease pressure. It is important to ensure the genetic background of tropical soybean is adapted to compensate for these environmental influences that are difficult or costly to control. Understanding the mechanisms behind agronomic traits such as days to flower and days to maturity will allow breeders to optimize the varieties they release, as photoperiod response is the most important trait influencing adaptation (Bandillo et al. 2017).

Soybean was domesticated ~5,000 years ago in northern China at latitude around 35°N (Hymowitz 1970, Carter et al. 2004). This latitude is characterized by long days >13 hours during the growing season. Soybean is a short day, photoperiod sensitive plant and flowering is induced by daylength (Garner and Allard 1920, Whigham and Minor 1978, Destro et al. 2001, Watanabe et al. 2012). When soybean is grown in a 12 hour or less daylength, it receives the cue to start flowering immediately upon emergence, making it difficult to adapt to lower latitudes (Hartwig and Edwards 1970, Hartwig and Kiihl 1979, Kiihl and Garcia 1989, Ray et al. 1995, Cober et al. 1996). This early flowering results in

a short stature plant that matures prematurely and leads to reduced yields (Sinclair and Hinson 1992). As soybean production spread worldwide, it was limited to high latitude cultivation such as the United States, Canada, Argentina, and southern Brazil. These temperate varieties were adapted to narrow bands of latitude termed "maturity groups" of 000 to VIII for soybean production in North America (Zhang et al. 2007).

Recently maturity genes controlling temperate flowering times have been cloned and characterized, and the influence of the *E* maturity genes on maturity groups is understood (Langewisch et al. 2014, Langewisch et al. 2017). *E1* is the most important maturity gene as it controls ~47% of the flowering phenotype in soybean (Bernard 1971, Xia et al. 2012). Functional *E1* is utilized in the southern United States and in maturity groups V and above (Langewisch et al. 2017). The semi-functional allele *e1-as* promotes slightly earlier flowering and maturity than *E1* and is used in the Midwest of the United States in maturity groups I to IV (Xia et al. 2012, Langewisch et al. 2017). In all earlier maturity groups, the nonfunctional *e1 null* allele is utilized to provide the earliest flowering time (Cober and Voldeng 2001). *E1* is known to be a transcription factor, but it is novel to the legume family making comparisons to the *Arabidopsis* flowering pathway a challenge (Watanabe et al. 2012, Xia et al. 2012). *E2* is also a major maturity gene and is an ortholog of the *Arabidopsis* flowering gene *GIGANTEA* (Bernard 1971, Watanabe et al. 2011). *E3* and *E4* also influence the flowering pathway as phytochrome receptors. It is important to note that *E2, E3,* and *E4* are similar to *E1* that their nonfunctional alleles: *e2, e3,* and *e4* also promote earlier flowering. (Buzzell 1971, Buzzell and Voldeng 1980, Watanabe et al. 2004, Liu et al. 2008, Watanabe et al. 2009, Xu et al. 2013).

It was discovered that it was possible to expand soybean production to ~20° latitude by increasing its maturity group by delaying flowering in short days. Manipulation of the *E* gene alleles allowed soybean growth to extend to slightly lower latitudes, although it did not allow for production to reach low, equatorial latitudes that were less than 20° (Spehar 1995, Carpentieri-Pipolo et al. 2002). A trait was discovered, named the long juvenile trait, in a plant introduction PI 159925 from Peru which did allow delayed vegetative growth in a short day (Hartwig and Kiihl 1979, Ray et al. 1995). This phenotype was observed again in Brazil through a natural variation of a cultivar Parana which was then named Paranagoiana (Bonato and Vello 1999). Paranagoiana allowed Brazil to expand its soybean production to their low latitude Matto Grosso region (Destro et al. 2001).

At this time, two separate names were assigned for the two sources of the long juvenile trait, *J* from the PI 159925 parent and *E6* in Paranagoiana, where the recessive allele of each gene controls the long juvenile trait (Ray et al. 1995, Bonato and Vello 1999). It was unclear if the phenotypes were caused by separate genes or alleles of the same gene (Destro et al. 2001). The genetic mechanism behind the long juvenile trait in PI 159925 was only discovered recently (Lu et al. 2017, Yue et al. 2017). Previous to that, numerous studies suggested that in certain backgrounds the long juvenile trait was under the control of a single gene demonstrated by a 3:1 Mendelian segregation ratio (Ray et al. 1995, Destro et al. 2001). However, delayed flowering was shown in a 1:15 segregation ratio in other studies (Carpentieri-Pipolo et al. 2002, Cober 2011) suggesting that another gene was able to influence the long juvenile phenotype. The gene controlling the long juvenile trait in PI 159925 was discovered to be the *Arabidopsis* flowering gene ortholog *ELF3*

(Lu et al. 2017, Yue et al. 2017) that contained a single nucleotide deletion causing a frameshift mutation in the 4th exon named *j-1*(Lu et al. 2017); however there was not a causative polymorphism discovered in the coding sequence of the *ELF3* gene in Paranagoiana (Li et al. 2017). Mapping data shows that *E6* is also located on chromosome 4 and may be either tightly linked or a complex mutation in *ELF3* (Li et al. 2017), so herein this allele is referred to as *j-x*. In addition, when soybean varieties from Ghana were sequenced for *ELF3*, no mutations that could affect the flowering phenotype were discovered, suggesting that there are other genes that may influence tropical soybean (Miranda et al, unpublished). The culmination of this data shows that the long juvenile trait may be influenced by multiple genes besides *ELF3,* which are still yet to be confirmed. The effect of *E1* and the long juvenile trait is only beginning to be understood recently as well (Lu et al. 2017).

The objective of this research is to understand the influence of the *E* maturity genes and alleles of the long juvenile trait on days to flower and days to maturity in a tropical environment.

MATERIALS AND METHODS

Plant materials

Six recombinant inbred line populations were created for this study, where each had one conventional juvenile parent (*J*) and one long juvenile parent (*j*). Five parents were chosen to create RIL populations. Jake is a high yielding MG V American variety with the genotype *E1, E2, J* (Shannon et al. 2007). X97-0101 (referred to as X97 for the duration of this paper) is a lectin-free, trypsin inhibitor-free isogenic experimental variation of Williams 82. It is MG III and has the genotype *e1-as, E2, J* (Palacios et al. 2004). 534545 is a food grade soybean variety, utilized for its high protein content. It is MG III and has the genotype *e1-as, E2, J* (Bilyeu and Wiebold 2016). PI 159925 is a plant introduction line from Peru. It was the first line in which the long juvenile trait was observed. It has the genotype *E1, E2, j-1* (Ray et al. 1995). Paranagoiana (PI 628880) is from natural variation of the Brazilian released variety Paraná (PI 628879) that contains the long juvenile trait. It is maturity group VI and has the genotype *E1, E2, j-x* (Bonato and Vello 1999). X5683-1-18 F718 (referred to as Canadian X for the duration of the paper) is an experimental line created by using the early maturing OT94-47 as a recurrent parent in a backcross with Paranagoiana. It has the genotype *E1, e2, j-x* (Cober 2011). A total of 256 lines were created from crosses X97-0101 x PI 159925 (X97-15), 543545 x Canadian X (534-Can), X97 x Canadian X (X97-Can), Jake x Paranagoiana (Jake-Pa) and Jake x PI 159925 (Jake-15). The list of genotypes for each parent can be found in Table 1 and the list of crosses is in Table 2.

RIL populations and field experimental design

All populations were initiated in Columbia, Missouri in summer 2014. The population X97-0101 x Jenguma was created from the self-pollinated F_1 plants used as donors from an independent backcrossing project in Upala, Costa Rica (10.8979°N, 85.0155°W) in collaboration with Costa Rica seeds. The F_2 seeds were advanced two additional generations by single seed descent then bulked and increased to create $F_{3:5}$ lines. In Jake-Pa and Jake-15, only F_2 plants that exhibited delayed flowering were selected to continue advancement, 20 lines in Jake-15 out of ~80 and 18 lines in Jake-Pa out of ~80. All other populations were advanced in Upala by single seed descent method for three additional generations then bulked and increased. Lines were selected for testing in Ghana based on amount of seed produced: 1 kg. This created artificial selection against unadapted lines, so at least 5 poor performing lines from each population (except Jake-15 and Jake-Pa) were also tested in Ghana. $F_{4:6}$ ($F_{3:5}$ X97-0101 x Jenguma) seed for all populations was shipped to Tamale, Ghana in spring 2016.

Yields trials were conducted in five fields throughout northern Ghana in 2016 and 2017. The fields were either a Savannah Agricultural Research Institute research field (Nyankpala SARI [NyS, 9.403°N,-1.008°W], Yendi SARI [YeS, 9.495°N,0.128°W], and Wa SARI [WaS, 9.799°N, -2.499°W] or a local farmer's field (Nyankpala Farmer [NyF, 9.396°N,-1.019°W] and Yendi Farmer [YeF, 9.412°N,-0.102°W]). Planting date was determined by the start of continuous seasonal rainfall and field conditions/availability. In 2016 soybeans were planted on 9 and 11 July in YeF, 13 July in NyF, 15 July in NyS, 16

July in YeS, and 20 July in WaS. In 2016, the YeF maturity and yield data were not collected due to soybean sudden death syndrome devastation. The experimental design was a single experimental line bordered by the local variety Jenguma in randomized complete block design with two replications, where one row of a RIL was bordered by a local check (Jenguma) on both sides. In 2016, blocking was done by population. All rows were hand planted 75 cm apart per IITA's recommendation (www.iita.org). Plots were ~300 cm (10 feet) long with a ~122 cm (4 foot) alley above. Granular inoculant was used and applied directly to open furloughs immediately before seeds were planted and covered. No fertilizer was used to represent local farmer practices and to replicate farmer agronomic and yield results. In 2016, 120 seeds were planted in each plot to compensate for predicted poor germination. Glyphosate was sprayed after planting and before emergence. Weed control was manual after emergence. Plots exceeding 100 plants per row were thinned to 100 during emergence note taking. Flowering date (R1) was determined when 2+ plants had opened flowers in the center of a plot to eliminate environmental influence on individual plants on plot ends. Plots were considered mature when 95% of pods were dried (Fehr and Caviness 1977). Height from the ground to the apical meristem of random individuals in each plot was taken immediately before harvest. Harvest was done by hand and threshed mechanically using a single plot thresher. Seeds were cleaned using sieves and by hand and then weighed for yield. Seed yield was calculated as grams per 10-foot row. YeS and NyF produced the highest quality seed and was stored in a 4°C cold room for planting in 2017.

The 2017 field and experimental design was identical to 2016 with some exceptions. Lines were eliminated from field tests in 2017 if they did not produce enough seed to be planted in 5 locations or if they exhibited a segregating phenotype in 2016. Populations that had PI 159925 as a parent suffered yield losses due to shatter. PI 159925 contains the *Pdh1* (Funatsuki et al. 2014) shatter-prone allele. The X97-15 population experienced heavy seed loss in Ghana, where the population size for the multi-location field test was reduced from 47 RIL in 2016 to lines to 5 in 2017 due to insufficient seed produced by the other 42 lines. In Jake-15, only 9 RILs of 20 were tested in 2017 due to low seed production. In 2017, fields were planted 8 July in YeF, 10 July in YeS, 11 July in NyF, 18 July in NyS (replanted 2 August), and 21 July in WaS.
200 seeds were planted per plot for to compensate for predicted low germination. In 2017, NyS no data were collected due to flooding damage that resulted in poor emergence.

The daylength was calculated based on civil twilight times. In northern Ghana in July, the daylength is 13 hours and the daylength in December is 12.33 hours (www.timeanddate.com).

Genotyping

DNA extraction

Initial genotyping was done with leaf presses on FTA cards (Whatman, Clifton, NJ) taken in Ghana in 2016 from trifoliates in R1 and shipped to Columbia, Missouri as described in (Beuselinck et al. 2006). Missing data was genotyped again in 2017 in Columbia, Missouri using F_7 seed that was shipped from Ghana. DNA was extracted from 2-5 seeds using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA) and followed the protocol described in (Langewisch et al. 2017).

E *gene genotyping assays*

E1 and *E2* genotyping assays were conducted as described in (Langewisch et al. 2017). *E3* genotyping assay was conducted as described in (Langewisch et al. 2014).

Dt1 *genotyping assays*

Dt1/ dt1 R166W

A SimpleProbe melting curve assay was developed to determine the adenine to thymine *dt1 R166W* missense allele from the wild type *Dt1* (Glyma.03g194700, Wm82.a2.v1). The primers Dt1in31f (5'-CATGAGAGAGATCACTGAC-3') and Dt1endr1 (5'-

GCAAAACCAGCAGCTACTT-3') amplify a 292-bp region, which includes the T/A SNP. The SimpleProbe oligonucleotide (5'- Fluorescein-SPC-TGCACAG<u>A</u>GGGAAACGGCT-Phosphate -3') was designed using the LightCycler Probe Design software (Roche Applied Science, Indianapolis, IN) and anneals to the sense strand. PCR reactions were 20 µl and included the DNA template, 0.5 µM reverse primer Dt1endr1, 0.2 µM forward primer Dt1in31f, 0.2 µM SimpleProbe, buffer (40 mM Tricine- KOH [pH 8.0], 16 mM $MgCl_2$, 3.75 µg ml^{-1} BSA), 5% DMSO, 200 µM dNTPs, and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). PCR reactions were run on the LightCycler 480 real-time PCR instrument (Roche Applied Science, Indianapolis, IN). Reactions were denatured at 95°C for 3 minutes, and then in each cycle denatured at 95°C for 20 seconds, primers annealed at 60°C for 20 seconds, and products elongated at 72°C for 20 seconds for 45 cycles. After amplification was completed, a melting curve was conducted from 50-70°C. The *dt1 R166W* mutant allele peak was observed at 57°C, and the *Dt1* wild type peak was observed at 63°C. Heterozygous *Dt1/dt1* samples produced both peaks.

Dt1/dt1 P113L

For detection of the P113L missense *dt1* alleles, a cleaved amplified polymorphic sequence assay was developed based on the introduction of a *HindIII* restriction enzyme site in the P113L *dt1* alleles (Liu et al. 2015). PCR products of 292 bp were amplified in 20 µl reactions containing DNA template with Dt1in31f and Dt1endr1 primers (as above) at 0.5 µM and buffer (40 mM Tricine- KOH [pH 8.0], 16 mM $MgCl_2$, 3.75 µg ml^{-1} BSA),

5% DMSO, 200 μM dNTPs, and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). Reactions were denatured at 95°C for 3 minutes, and then in each cycle denatured at 95°C for 20 seconds, primers annealed at 60°C for 20 seconds, and products elongated at 72°C for 20 seconds for 45 cycles. After amplification was completed 5 μl of each sample was removed to check for product formation on the FlashGel system (Lonza, Basel, Switzerland). To the remaining 15 μl of each sample, an enzyme mixture (15 μl) was added that contained 1.5 μl New England BioLabs (NEB, Ipswich, MA) buffer 2, 1.5 μl NEB *HindIII* (30,000 units), and 12 μl of ddH_20. Reactions were incubated overnight at 37°C, and products were separated on the FlashGel system. The *Dt1* genotype produced a 215 bp band, while *dt1* P113L genotypes produced bands of 215 bp and 77 bp, and heterozygous samples produced bands of 292, 215, and 77 bp.

ELF3 genotyping assays

j-1: cytosine deletion (C-del) found in PI 159925

For detection of the long juvenile trait C-del in the PI 159925 version of *ELF3* *(*Glyma04g05280, Wm82.a2.v1), a SimpleProbe assay was created. The primers Cdelfor (5'-TGTTCTGCAGAGAATGCGGT-3') and Cdelr (5'-CCTCCTCCACAACCAGTTCC-3') produce a 254-bp PCR product that contains the C/- SNP described in by Lu et al, 2017 (Lu et al. 2017). The SimpleProbe oligonucleotide (5'-Fluorescein-SPC-GACGGTAGCCACCTTTCAAAATGCA-Phosphate-3') was designed on the sense strand using the LightCycler Probe Design software (Roche

Applied Science, Indianapolis, IN). PCR was identical as the *Dt1/dt1 R166W* assay with the exception that the melting curve was from 50-75°C. The C-del mutant allele peak was observed at 61°C, and the *ELF3* wild type peak was observed at 68°C. Heterozygous samples produced both peaks.

j-x: unknown mutation in Paranagoiana

The exact polymorphism controlling the long juvenile trait in Paranagoiana is not known, but it is believed to be tightly associated with *ELF3* (Li et al. 2017). Our sequencing of *ELF3* in Paranagoiana also did not produce any polymorphisms except for our difficultly to amplify and sequence the junction between intron 3 and exon 4. We developed a gel-based assay with *ELF3* primers and control primers to ensure PCR was successful. We used the primers ljkf. (5'- CGAGTATTGTGCAATTTTCTTGATCC-3') and Cdelr: (5'-CCTCCTCCACAACCAGTTCC-3') to amplify a 652-bp region that includes the intron 3 to exon 4 junction. The control primer set lx1f (5'- ACCGACATCTTAGCGTGCTT-3') and lx1r (5'-AAAAGGTTGTCTCTATTATGCCAT-3') amplifies a region of the lipoxygenase gene on chromosome 13.

PCR reactions were 20 μl and included the DNA template (this assay did not work with DNA from leaf presses), 0.5 μM *ELF3* reverse primer Cdelr, 0.5 μM *ELF3* forward primer ljkf, control primers: 0.25 μM lx1f and 0.25 μM lx1r, buffer (40 mM Tricine-KOH [pH 8.0], 16 mM $MgCl_2$, 3.75 μg ml^{-1} BSA), 5% DMSO, 200 μM dNTPs, and

0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). PCR reactions were run on a thermocycler and were denatured at 95°C for 3 minutes, and then in each cycle denatured at 95°C for 20 seconds, primers annealed at 60°C for 20 seconds, and products elongated at 72°C for 60 seconds for 45 cycles. After amplification was completed, PCR products were run on a 1.5% agarose gel containing SYBR Safe DNA gel stain diluted 1:10,000 at (145 V) for 20 minutes. Products were visualized using a blue-light transilluminator. Only lines that produced product 129 bp for the lx1 primers were assigned a genotype for *J*. If an upper band was present such as in the *J* control, the line was considered conventional, if no 652 bp band was present, it was considered *j-x*.

<u>Statistical analysis</u>

Days to flower notes were taken three times a week in the Nyankpala fields, once per week in the Yendi fields, and once per week in the Wa field on average in 2016. Days to flower 2017, and days to maturity, 2016 and 2017 were recorded twice per week in Nyankpala fields, twice per week in Yendi, and once per week in Wa. ANOVAs for all data collected were analyzed using PROC GLM procedure in SAS software version 9.4 (SAS Institute. 2012. The SAS 9.4 system for Windows. SAS Inst., Cary, NC).
Data from lines containing the same genotype were grouped together and analyzed by ANOVA for genotype, location, rep(location), and genotype*location effect. Outliers from each genotype group were removed only after verifying that they were a note taking error. Data from lines with incomplete genotype data (either missing or heterozygous for at least one gene) were omitted from analysis. Days to flower data from Wa from 2016

and 2017 were not used in the analysis due to the imprecise data collected from only once weekly note taking. After data was cleaned based on these standards, Fisher's least significant differences (LSDs) were generated using SAS software 9.4 where p=0.05. Boxplots were constructed in Excel.

RESULTS

Analysis of Variance of Days to Flower, Days to Maturity, and Yield for Six RIL Populations

Six RIL populations were created to test the effects of our maturity genes of interest: *E1, E2, E3, Dt1*, and *ELF3* and their mutant alleles (Table 2) on soybean phenology (days to flower and days to maturity) in the low latitude environment of Ghana (Appendix 2, Table 1). All populations have one conventional juvenile parent (*J*) and one long juvenile parent (*j*) (Table 1). Two populations, Jake-15 (Jake x PI 159925) and Jake-Pa (Jake x Paranagoiana) were segregating for alleles of *ELF3: J, j-1, or j-x* and were fixed for *E1, E2,* and *dt1 R166W.* The F_2 plants were selected in those populations that flowered past 40 days to continue generation advancement. The presence of the *j-1* alleles in the Jake-15 selected lines and the *j-x* alleles in the Jake- Pa selected lines was later confirmed by genotyping assays (data not shown). Two populations, X97-15 (X97-0101x PI 159925) and X97-Jen (X97-0101 x Jenguma), were segregating for *E1* or *e1-as, J* or *j-1* or *j-x,* and *Dt1* or *dt1 R166W* in X97-15 or *dt1 P113L* in X97-Jen. Both X97-15 and X97-Jen populations were fixed for *E2*. The last two populations, 534-Can (534545

x Canadian X [OT94-47 x Paranagoiana]) and X97-Can (X97 x Canadian X) were segregating for *E1* or *e1-as, E2* or *e2, E3* or *e3, J* or *j-x*. They were both fixed for *Dt1*.

Populations were grown for 2 years (2016, 2017) in 5 locations in northern Ghana (9°N) and days to flower and maturity and yield were recorded. Analysis of variance was performed on the 2-year results (Appendix 2, Table 2-7). All six populations produced useful models ($r^2 > 0.80$) where genotype and location had significant effects on days to flower, maturity, and yield. The coefficient of variation for yield in all populations was too large to be considered useful data.

Frequencies for Days to Flower and Days to Maturity of Six RIL Populations

RIL line frequencies of all populations for days to flower and maturity in 2016 and 2017 combined are shown in Figure 1. The mean days to flower for Jake-15 lines containing the *j-1* allele was 45.6 (Figure 1a) and was 47.2 in the Jake-Pa lines that contain the *j-x* allele (Figure 1b) and the mean days to maturity in Jake-15 (Figure 1c) was 110.2 and in Jake-Pa was 115.9 days (Fig 1d). These results of the RIL progeny are consistent with days to flower and maturity for the parents (Table 1).

The X97-15 and X97-Jen populations were both segregating for *E1*, *Dt1*, and *J* and their mutant alleles (Table 2). Forty-four RILs were tested from the X97-15 populations and 60 lines were tested in X97-Jen population. Since the population was segregating for multiple maturity genes and had a larger range of days to flower and days to maturity, no

means are reported here. In X97-15 the majority of lines flowered between 37 and 43 days (Figure 1e) which is between the range of the X97 and PI 159925 parents (Table 1). The majority of RILs in the X97-Jen population flowered between 32-38 days (Figure 1f), more lines flower closer to parent X97 which flowers in 29 days than the long juvenile parent Jenguma that flowers in 44 days (Table 1). Days to maturity for X97-15 is later than the conventional parent X97, where only 5 lines mature before 100 days and most mature between 107-114 days (Figure 1g). This range is beyond the season length of the long juvenile parent which matures in 111 days (Table 1). X97-Jen shows a normal distribution of days to maturity where most mature between 102-107 days (Figure 1h), which is much earlier than the long juvenile parent Jenguma which matures in ~116 days (Table 1).

The 534-Can and X97-Can populations were segregating for *E1, E2, E3* and *J* and their mutant alleles and were fixed for *Dt1* (Table 2). Forty-seven RILs were tested from 534-Can and 39 RILs were tested in X97-Can. In 534-Can the majority lines of lines flowered around 38 days (Figure 1i) which is close the Canadian X parent which flowers in 39 days (Table 1) but there are 11 RIL lines that show delayed flowering similar to the long juvenile donor of Canadian X, Paranagoiana, at 45-46 days (Table 1). The X97-Can line shows a bimodal distribution where there is a peak of RILs that flower around the same time as the long juvenile parent (37-38 days) and then another peak that is more similar to the long juvenile donor parent Paranagoiana at 45 days (Figure 1j, Table 1). The 534-Can population shows peaks at days 106, 109-111, and at 115-117 days of maturity (Figure 1k). The first peak is similar to the Canadian X days to maturity which

is 105 days, and the third peak is similar to the long juvenile donor Paranagoiana at 113 days (Table 1). The X97-Can population shows a peak of maturity at 108-112 days (Fig 1l) which is between the range of maturity for the Canadian X parent and the long juvenile donor Paranagoiana (Table 1).

In a Jake Background, 2 Variants of the Long Juvenile Trait Exhibit Differing Means for DTF and DTM

To understand the different effects of polymorphisms of *ELF3* on days to flower and maturity, we created 2 RIL populations that were segregating for different alleles of *ELF3*, and had fixed alleles of *E1, E2, E3,* and *dt1* (Table 2). Means for days to flower and days to maturity for each population and parents were analyzed using Fisher's LSD (p=0.05) (Figure 2). When contrasted to lines with *j-1*, the conventional juvenile parent Jake, flowered 14 days earlier. Both the parent Paranagoiana and RILs derived from Paranagoiana with the *j-x* allele, showed a significant difference in days to flower compared to the PI 159925 *j-1* allele of 2 days (Figure 2a). The conventional parent Jake reaches maturity 19.3 days before the Jake-15 RILs containing the *j-1* allele, and RILs with the *j-1* allele mature 5 days before RILs with *j-x* allele from Paranagoiana. However, the parents PI 159925 and Paranagoiana do not follow show this same difference in days to maturity (Figure 2b).

The Missense Allele of the Major Maturity Gene *E1, e1-as*, Influences Days to Flower but Does Not Affect Days to Maturity

To test the effect of the allelic combinations of *e1-as*, *E1,* conventional juvenile (*J*) and the PI 159925 long juvenile trait *(j-1),* we utilized the X97-15 RIL population (Table 2). Means of each genotype for days to flower and maturity were compared (Figure 3). No RILs were present with the *e1-as*, *J* genotype in this population. Contrasting the parent X97 with the *e1-as, J* genotype to RIL lines with the *E1, J* genotype there was a significant 3.2 day difference in days to flower between the two groups. When comparing *e1-as* with the long juvenile trait to *E1* in a conventional juvenile background there is a 4.3 difference in days to flower. Finally, there is a 6.3 days to flower difference in *E1* versus *e1-as* in a long juvenile background, which is a similar result as the PI 159925 parent (Figure 2a). Interestingly, these differences are not seen in days to maturity. The only significant difference is between the genotype groups that are conventional or long juvenile, regardless of the *E1* status, with a difference of 14 days (Figure 3b).

Effects of the *j-1* and *j-x* Alleles of the Long Juvenile Trait in Different Genetic Backgrounds

To confirm that the phenotypes observed in *j-1* and *j-x* alleles can be applied for breeding purposes, we compared days to flower and days to maturity with those alleles in different genetic backgrounds. We performed a multiple means comparison test across five RIL

populations: Jake-15, X97-15, Jake-Pa, 534-Can, and X97-Can, where the *E1* and *J* genotype of each line was used for grouping within populations. Two populations, Jake-15 and X97-15, were segregating for the *j-1* long juvenile trait allele from PI 159925. There were also three populations segregating for the *j-x* long juvenile allele from Paranagoiana, Jake-Pa, 534-Can, and X97-Can. A comparison was made for days to flower and days to maturity for *E1, j-1* and *E1*, *j*-x RILs along with several control lines (Figure 4). The conventional parent (*E1*, *J*) had an 11 days to flower difference from *E1, j-1* lines in the X97-Can population. X97-Can (*j*-x) flowered after 42 days, 6 days earlier than its long juvenile donor parent, Paranagoiana. X97-Can was similar to X97-15. E1, *j-1* lines from the X97-15 population were not significantly different compared to lines in the Jake-15 population with the same genotype or from the long juvenile parent PI 159925. The Jake-Pa and 534-Can RILs with *E1, j-x* backgrounds did not show significant difference in days to flower, but both *E1, j-x* categories were significantly later than *E1, j-1* categories by at least 2 days. Jake-Pa and 534-Can also flowered later than the X97-Can population that has the same *j-x* allele (Figure 4a).

In days to maturity, the Jake-15 and X97-15 lines with *E1, j-1* did not have significant difference in days to maturity between each other or their parent PI 159925 but were significantly different from the conventional parent by 20 days. Jake-Pa and 534-Can with *E1, j-x* were not significantly different in days to maturity but matured 2.5 days later than their parent Paranagoiana and were different from *E1, j-1* by 5 days. X97-Can with the *j-x* allele matured 3.5 days before other populations with the *j-x* allele but was statistically similar to its long juvenile donor Paranagoiana (Figure 4b).

E2 Affects Days to Flower and Days to Maturity in the 534-Can Population, but Does Not Have an Effect in the X97-Can Population

To understand how *E2* affects days to flower and days to maturity in a tropical climate, two populations were created that were segregating for *E1/e1-as, E2/e2, and J/j-x.* They had the same long juvenile donor parent Canadian X (*E1, e2, j-x*) and varied in the conventional parent, either the food grade soybean 534545 (*e1-as, E2, J*) or X97 (*e1-as, E2, J*). We categorized days to flower and days to maturity data based on genotype and performed a multiple means comparison test for genotypes of each population separately. The 534-Can population had 5 different genotypes available (Figure 5). There was one conventional genotype group *E1, E2, J* which flowered the earliest at 34 days. All genotype groups significantly increased days to flower in a stepwise fashion as alleles that delay flowering were added. All genotype groups were significantly different from each other. The Canadian X parent (*E1, e2, j-x*) had a similar mean to the *e1-as, E2, j-x* group, and the long juvenile donor Paranagoiana had similar days to flower as the *E1, E2, j-x* genotypes (Figure 5a). Days to maturity increased significantly as alleles were added that delay flowering. All genotype groups were significantly different for days to maturity. The Canadian X parent (*E1, e2, j-x*) had a similar maturity to the *e1-as, e2, j-x* genotype group. The long juvenile donor Paranagoiana (*E1, E2, j-x*) has a similar maturity to the *E1, e2, j-x* genotype group. The RIL genotypes *E1, E2, j-x* have 5.5 longer days to maturity compared to Paranagoiana (Figure 5b).

The X97-Can population was the only population to have the *e1-as, J* genotype. This genotype group was not significantly different in days to flower compared to the *E1, E2, J* genotype, but it did flower significantly earlier than the *e1-as, j-x* genotype regardless if *E2* or *e2* was present. *e1-as, j-x* also flowered significantly earlier than the *E1, j-x* genotype. There was no difference in days to flower in the *E1, j-x* genotypes even if *E2* was functional or nonfunctional. The Canadian X parent (*E1, e2, j-x*) had similar days to flower compared to several genotypes. Paranagoiana (*E1, E2, j-x*) showed more delayed flowering than the same RIL genotype group by 5 days (Figure 6a). In days to maturity, the conventional juvenile genotypes did not show a difference in days to maturity, regardless if *e1-as* or *E1* was present. The conventional juvenile groups differed significantly from genotypes that had the *j-x* allele. Except for *e1-as, e2, j-x* and *E1, E2, j-x* there was not a significant difference in days to maturity in the genotypes that had the *j-x* allele. The parent Canadian X (*E1, e2, j-x*) had similar days to maturity as several genotypes, and Paranagoiana (*E1, E2, j-x*) had similar days to maturity as the same RIL genotype (Figure 6b).

<u>*E2* and *E3* Have an Additive Effect to Delay Flowering and Maturity in a *E1* background in 534-Can but Not in X97-Can</u>

To test the effect of *E1, E2, E3* and *J* allelic combinations, we compared the means of different genotype groups in 534-Can and X97-Can (Figures 7 and 8). In 534-Can, two conventional genotype groups did not exhibit delayed flowering when *E2* was added in an *E3* background. Days to flower are significantly increased from a *e1-as, e2, e3, j-x*

background by the addition of *E2, E3,* or both *E2* and *E3*. *E1*, *e2, e3, j*-x significantly delays flowering compared to any *e1-as, j-x* genotype, however the *E1, e2, E3* genotype is not significantly different. There is another increase in days to flower when all *E* genes are functional (Figure 7a). In a conventional background, the *E1, E2, E3* genotype has delayed maturity contrasted to *E1, e2, E3* but has a similar maturity as *e1-as, e2, e3*, *j-x.* When *E2* is added to the nonfunctional long juvenile background, there is a 5 day delay in maturity. The *e1-as, e2, E3* genotype has the same maturity as *e1-as, E2, e3*, but when the genotype is *e1-as, E2, E3* there an added 11 days of maturity compared to *e1-as, e2, e3, j-x.* The next significant delay in maturity is when *E1* and *E3* are functional. Finally, the latest maturing lines have all functional *E* genes and the long juvenile trait (Figure 7b).

In the X97-Can population, none of the *E* alleles influence days to flower in a conventional background. Days to flower increase with the addition of *j-x* but there isn't another significant delay in flowering until *E1* and *E2* are added. All combinations of the functional *E* alleles are not significantly different to influence days to flower (Figure 8a). In days to maturity, the only significant differences are between the conventional juvenile groups, the addition of the *j-x* allele, and the addition of all functional *E* alleles with *j-x* (Figure 8b).

Genotype Data and Mean Days to Flower and Maturity for Other Inbred Lines of Interest

Other inbred lines were also tested with the RIL populations in northern Ghana to understand if they flower and mature similarly as the experimental lines. Lines tested include three conventional juvenile lines from the United States, two experimental Canadian lines with the *j-x* allele, two varieties from Australia that have the *j-1 ELF3* allele, three IITA lines that were released in Mozambique, and five Ghanaian varieties. Their genotypes and mean days to flower and maturity are in Table 3. The American conventional varieties flower and mature the earliest. Interestingly, the Australian Melrose variety flowers and matures early regardless of functional *E* alleles and *j-1* being present. Of the African varieties, all have functional *E* genes with the exception of Walima, which has *e2*. Walima has a SNP in *ELF3* that was discovered by Lu et al (2017), in a Brazilian variety. Some of the African varieties have an arginine to glycine mutation at position 73 (R73G), however the effect of this SNP is not known. Some of the African varieties do not have polymorphisms in *ELF3* although they exhibit delayed flowering.

DISCUSSION

Soybean production is expanding to equatorial areas of the world allowing subsistence farmers access to this economically important crop (Mbanya 2011, Abate et al. 2012). Soybean is an invaluable crop for the developing world as it offers resiliency: farmers can choose to sell their seed to livestock feed markets or can eat the soybean to benefit from the high protein and calories (Masuda and Goldsmith 2009). However, there are still many obstacles that must be overcome for soybean to be accepted such as accessibility to high quality seed and profitability (Dogbe et al. 2013). Both of these challenges can be met with skilled breeding practices that strive for achieving maximum yields in a low latitude environment. One aspect of breeding soybean in this new environment is understanding the genetic mechanisms behind days to flower and days to maturity as soybean is a photoperiod sensitive plant that is not adapted to the characteristic 12-hour days near the equator, resulting in low yields (Sinclair and Hinson 1992). Our results can help facilitate breeders' efforts to breed for the correct season length to ensure the local farmer has an optimally adapted variety.

Our study aimed to understand the role of *E* genes and alleles of the long juvenile trait by conducting field tests of populations that were segregating for different allelic combinations of our genes of interest. Most importantly, we found that addition of the long juvenile trait delayed flowering a minimum of 13 days and delayed maturity by 19 days, proving that the long juvenile trait is a critical feature for adaptation to tropical environments (Bonato and Vello 1999) (Figure 2). We found that in a Jake background

the two different alleles of *ELF3*: *j-1* and *j-x* have significantly different days to flower and maturity (Figure 2). In addition, we determined that *E1* and *e1-as* influence days to flower but not days to maturity (Figure 3). These data suggest that it is possible to control soybean season length solely through the choice of the long juvenile allele and the vegetative to reproductive length ratio can be adjusted through the selection of *e1-as* or *E1*. These results are consistent with other studies that show that the *E1 or e1*-as alleles influence different days to flower in a long juvenile background (Lu et al. 2017). These results seem to be consistent in different genetic backgrounds unless *E2* and E3 are manipulated. The RILs in the X97-Can population with *E1, j-x* genotype flowered earlier than the long juvenile donor Paranagoiana which also is *E1, j-x,* even though the 534-Can population with the identical genotype behaved the same as Paranagoiana (Figure 4a). However, the X97-Can *E1, j-x* RILs had the same maturity as Paranagoiana but the Jake-Pa and 534-Can RILs with *E1, j-x* matured significantly later than both X97-Can and Paranagoiana (Figure 5b).

Interestingly, when comparing days to flower and days to maturity for *E1, E2, J* alleles, the two populations X97-Can and 534-Can do not produce the same results. The 534-Can population shows a stepwise increase in days to flower and maturity as functional *E* alleles are added. However, the X97-Can population shows *E2* does not influence days to flower. In days to maturity, the strongest influence is due to the presence of the long juvenile trait, or when all *E* genes are functional in a long juvenile trait background (Figure 6). This same trend is observed when *E3* is manipulated. 534-Can experiences significant delays in flowering and maturity when functional alleles of *E2* or *E3* are

present (Figure 7). X97-Can shows the same results as X97-15 (Figure 3), where days to flower is influenced by the long juvenile trait and alleles of *E1* (Figure 8a), but maturity is only influenced by the long juvenile trait or all functional *E* genes (Figure 8b). It is also interesting to observe the Mozambique variety Walima which has an *ELF3* allele *j-2* that was discovered in the Brazilian variety BR-121 (Lu et al, 2017). This variety typically has ~45 days to flower and ~120 days to maturity in a 12 hour day (Lu et al, 2017) however, Walima flowers in 43 days and matures in 99 days (Table 4). These results could be due to the presence of the nonfunctional *e2* allele, however if this is true then *e2* would have an effect on early maturity and not affecting days to flower. An experimental population would need to be created to understand the effect of *E2/e2*.

It is also important to note that there is also natural selection against unadapted varieties. There were a very low number of RILs with the genotype *e1-as, LJ* or *E1, LJ* that survived. Pod shatter is also devastating to yields, and the gene controlling a large percentage of the shatter phenotype, *Pdh1* (Funatsuki et al. 2014), was present in populations with PI 159925 as a parent. These plants would shatter near the same day as maturity. There also seemed to be natural selection for functional alleles of the *E* genes and the long juvenile trait based on the number of lines that survived that had delayed flowering and maturity (Figure 1). This can be seen in the two populations with Canadian X as a parent. The two parents used did not have all functional *E* alleles and the long juvenile trait (Table 1) yet there were many RILs that were produced with the *E1, E2, j-x* genotype, suggesting a preference for this genotype.

After sequencing the released African varieties for the *ELF3* gene, it is interesting to notice that some do not have their source of delayed flowering in short days from this gene. Four have a polymorphism in *ELF3* that may or may not affect the phenotype (Table 4). At the time this paper is written, the long juvenile genetic mechanism in most African varieties is not known. The question arises, were alleles of *ELF3* selected against in this environment, or were they simply not introduced? There is a possibility that breeding with alleles of *ELF3* could have yield benefits, although this would need to be evaluated in a field setting. This research has shown it is possible to manipulate the vegetative to reproductive stage ratio through the *E1* allele chosen in a *j-1* background, and it may be possible to add finer regulation of days to flower and days to maturity with *E2* and *E3* alleles in a *j-x* background. This knowledge and these alleles should be implemented in African breeding programs as is needed in certain environments, to test for possible yield benefits. There is a possibility that the source of the long juvenile trait currently being used could have yield drag, or other negative growth attributes, associated with it.

Taken together it is possible to control tropical soybean season length through the selection of the long juvenile alleles and also the days to flower through selection of *E1* or *e1-as*, and possibly *E2* and *E3* in certain backgrounds. As has been mentioned in previous studies, there are still background effects that influence long juvenile trait maturity phenotypes (Ray et al. 1995). This research will allow breeders to evaluate the impact on yield by consciously manipulating season length and the vegetative to reproductive stage ratio.

FIGURES

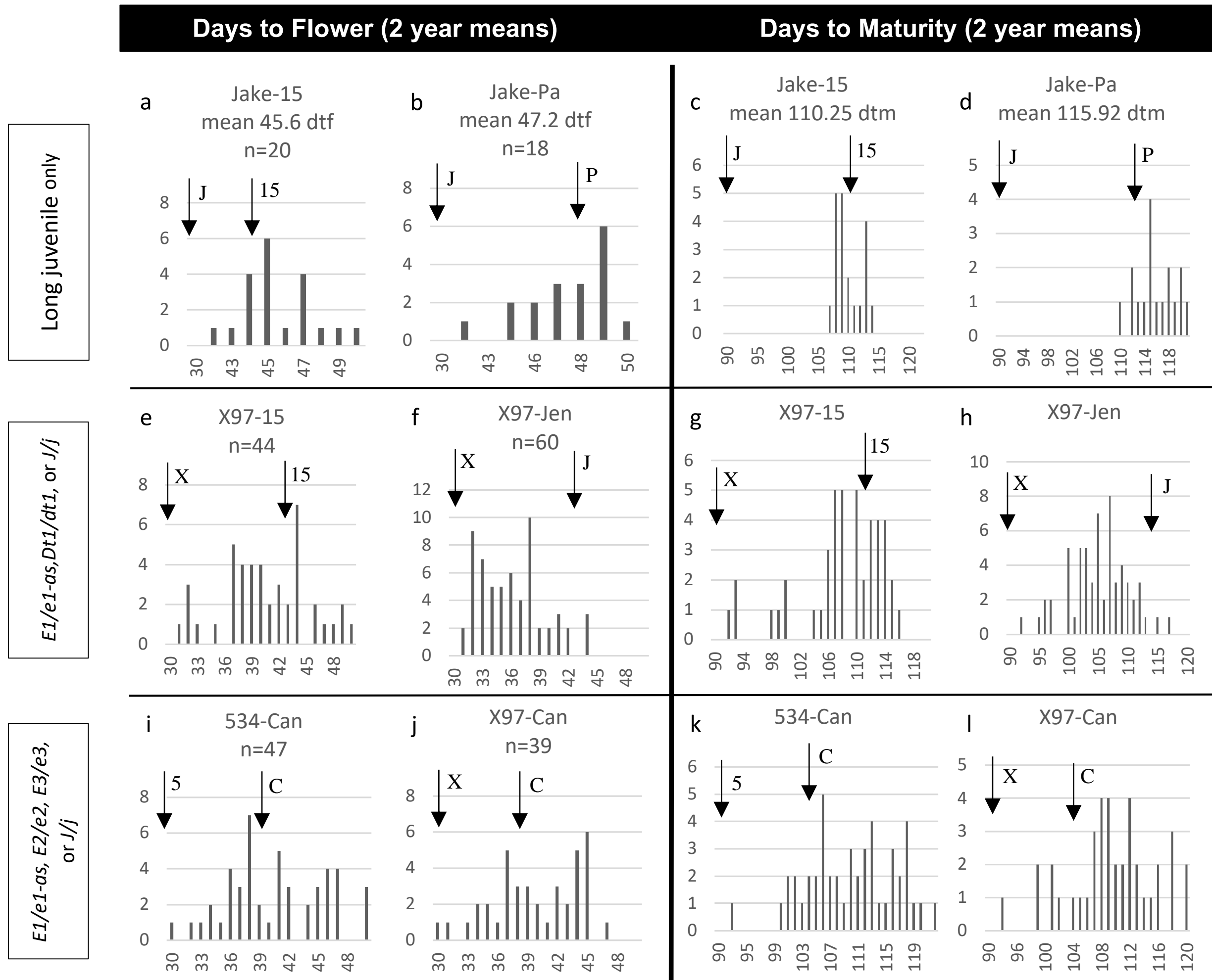

Figure 1: Frequencies of agronomic traits from six RIL populations. Number of RILs is on the y-axis and days is shown on the x-axis. Data for parents of each population are shown with an arrow with the first letter of the parent name to the right. a: Days to Flower of Jake-15 b: Days to Flower of Jake-Pa c: Days to Maturity of Jake-15 d: Days to Maturity Jake-Pa. a-d: Both populations were selected for the long juvenile trait. e: Days to Flower of X97-15 f: Days to Flower of X97-Jen g: Days to Maturity of X97-15 h: Days to Maturity of X97-Jen e-h: Both populations were segregating for *E1/e1-as* and different alleles of *J/j.* i: Days to Flower of 534-Can j: Days to Flower of X97-Can k: Days to Maturity of 534-Can l: Days to Maturity of X97-Can i-l: Both populations were segregating to *E1/e1-as, E2/e2, E3,e3* or *J/j-x*

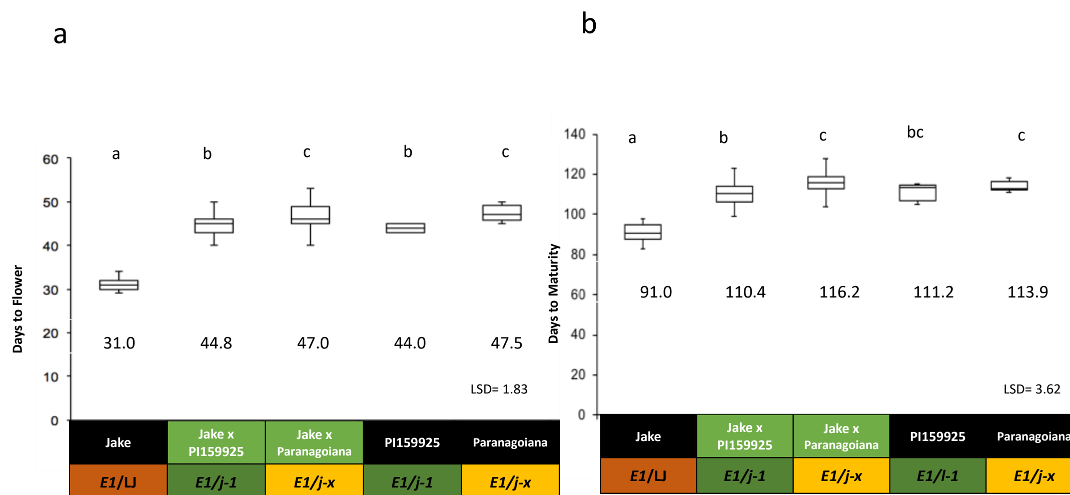

Figure 2. Days to flower and days to maturity in Jake x long juvenile RIL populations. Means for each genotype are shown under the boxplot. Parents are in a black background a: Days to flower for Jake-15 and Jake-Pa and parents. b: Days to maturity.

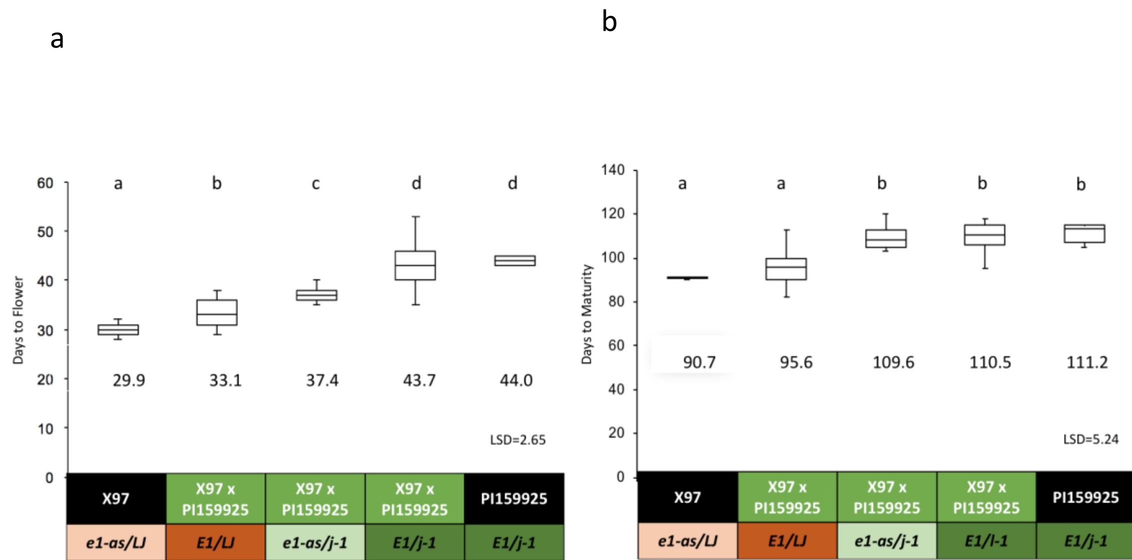

Figure 3. Days to flower and days to maturity in a RIL population that was segregating for *e1-as, E1, LJ, and j-1*. Parents have a black background. Data from the individual RILs were analyzed together based on their genotype. Means are shown under boxplots. a: Days to flower b: Days to maturity

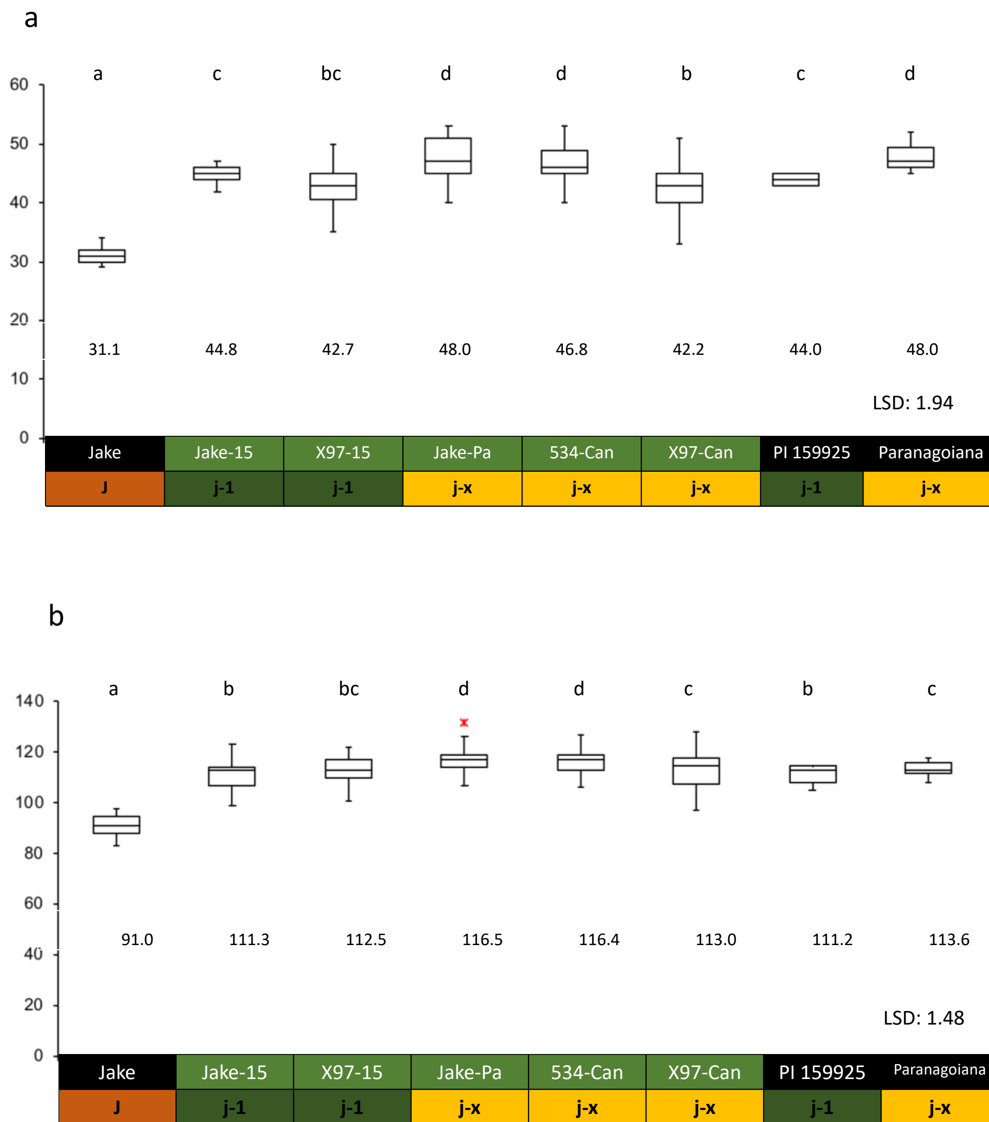

Figure 4. Days to flower and days to maturity for all RILs and parents with a fixed *E1* background. Parents have a black background. Data from the individual RILs were analyzed together based on their genotype. Means are shown under boxplots. a: Days to flower b: Days to maturity

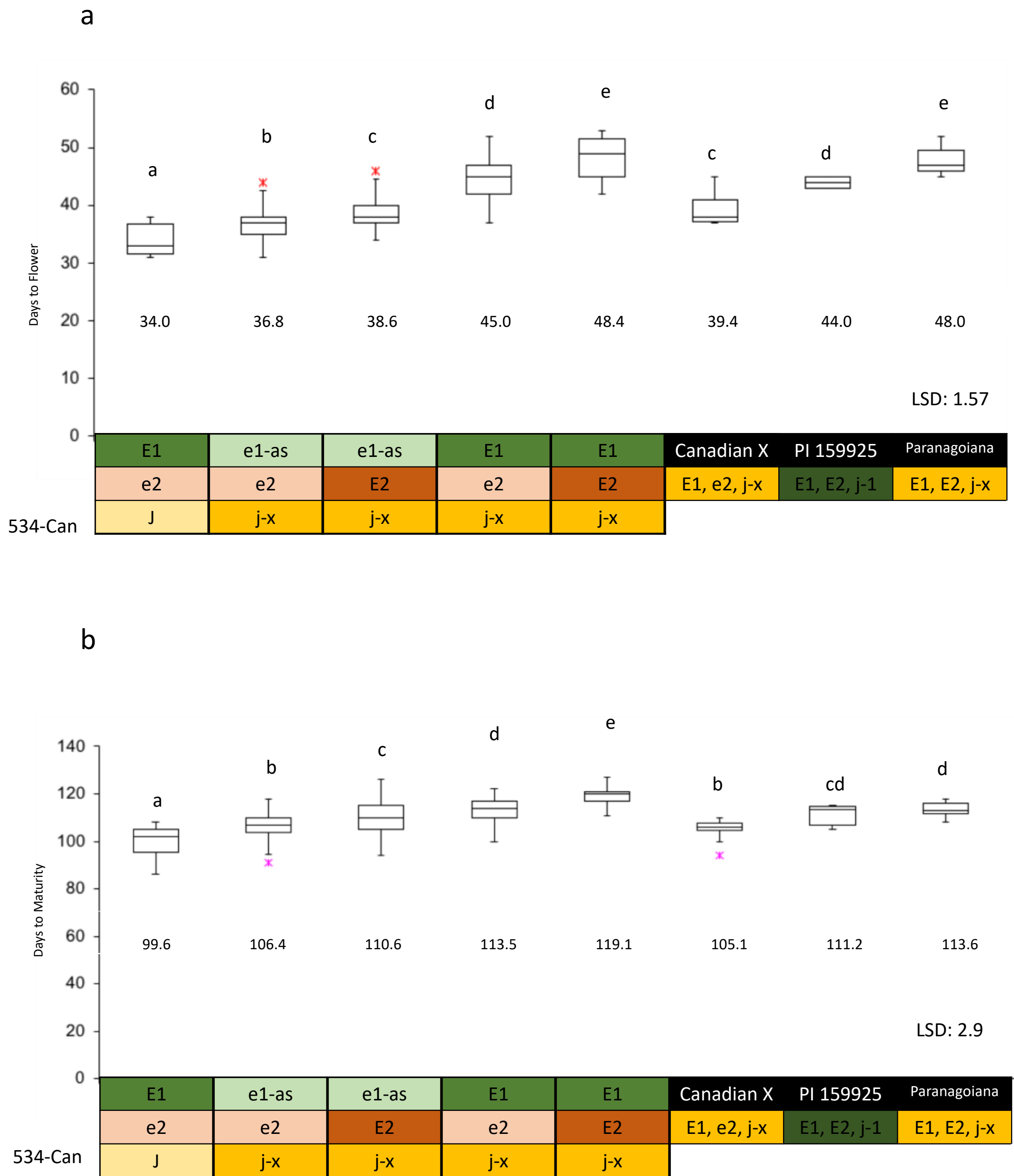

Figure 5. Days to flower and days to maturity for RILs from the 534-Can population. Lines were segregating for *E1/e1-as, E2/e2,* and *J/j-x*. Parents have a black background. Data from the individual RILs were analyzed together based on their genotype. Means are shown under boxplots. a: Days to flower b: Days to maturity

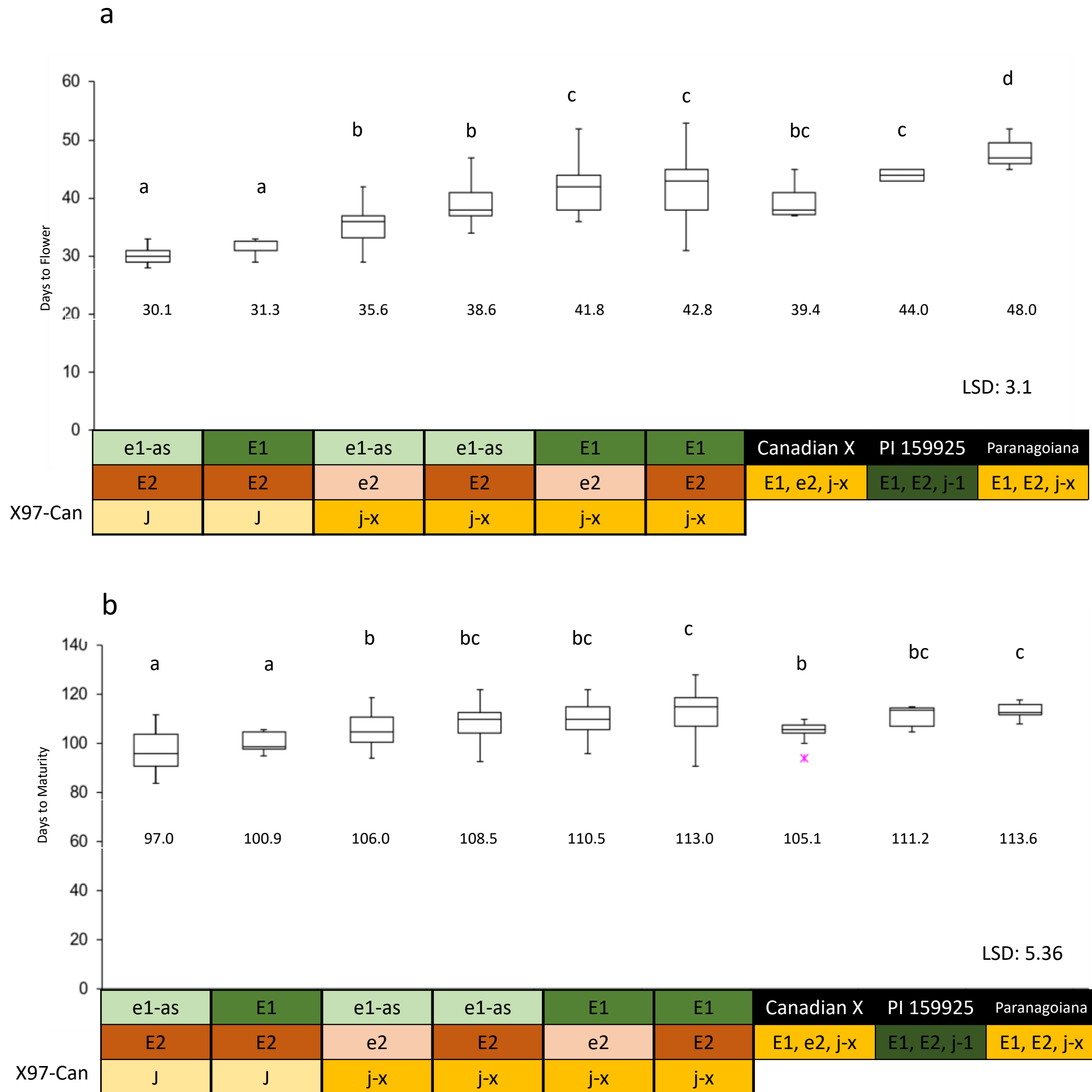

Figure 6. Days to flower and days to maturity for RILs from the X97-Can population. Lines were segregating for *E1/e1-as*, *E2/e2*, and *J/j-x*. Parents have a black background. Data from the individual RILs were analyzed together based on their genotype. Means are shown under boxplots. a: Days to flower b: Days to maturity

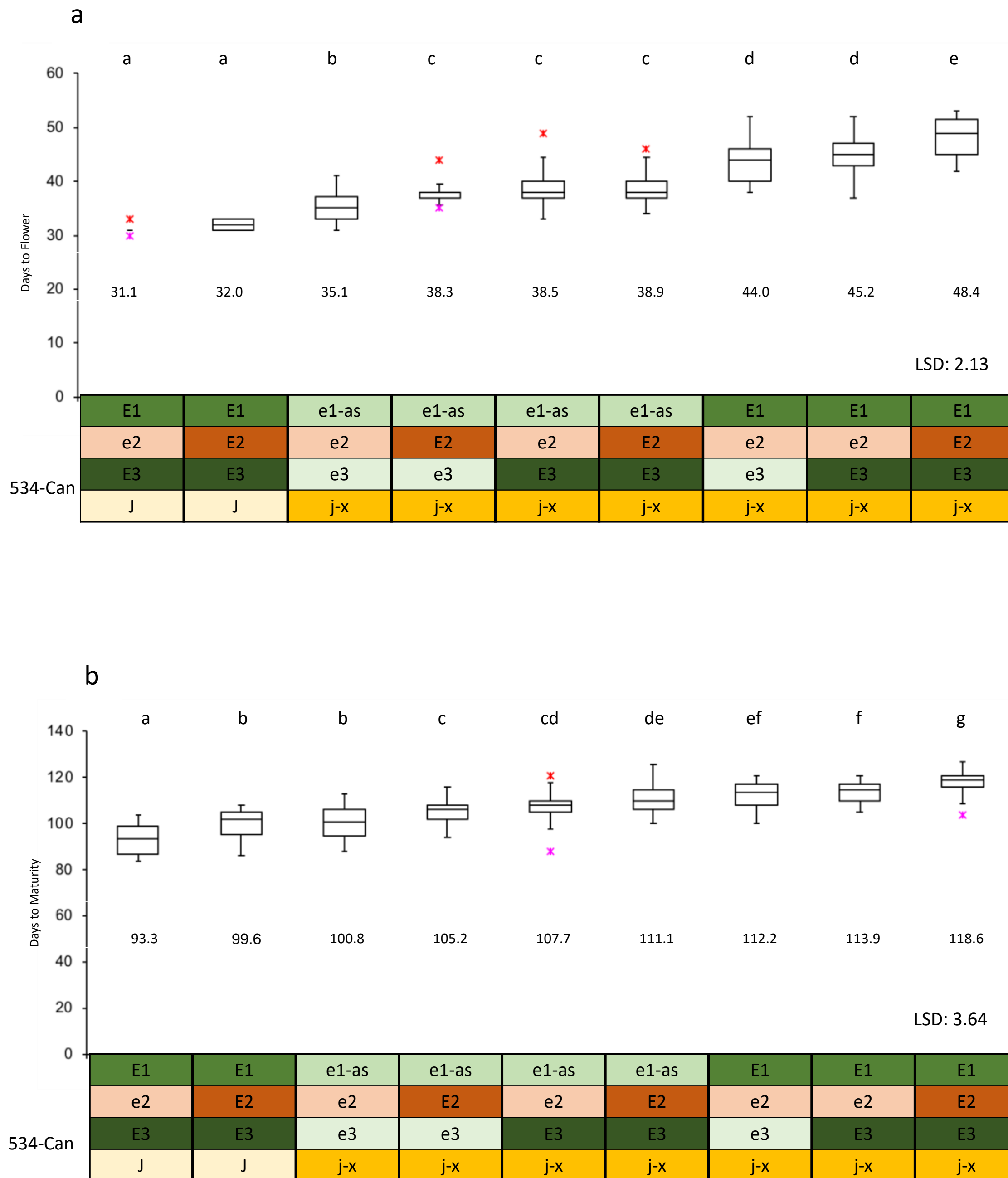

Figure 7. Days to flower and days to maturity for RILs from the 534-Can population. Lines were segregating for *E1/e1-as, E2/e2, E3/e3* and *J/j-x*. Parents have a black background. Data from the individual RILs were analyzed together based on their genotype. Means are shown under boxplots. a: Days to flower b: Days to maturity

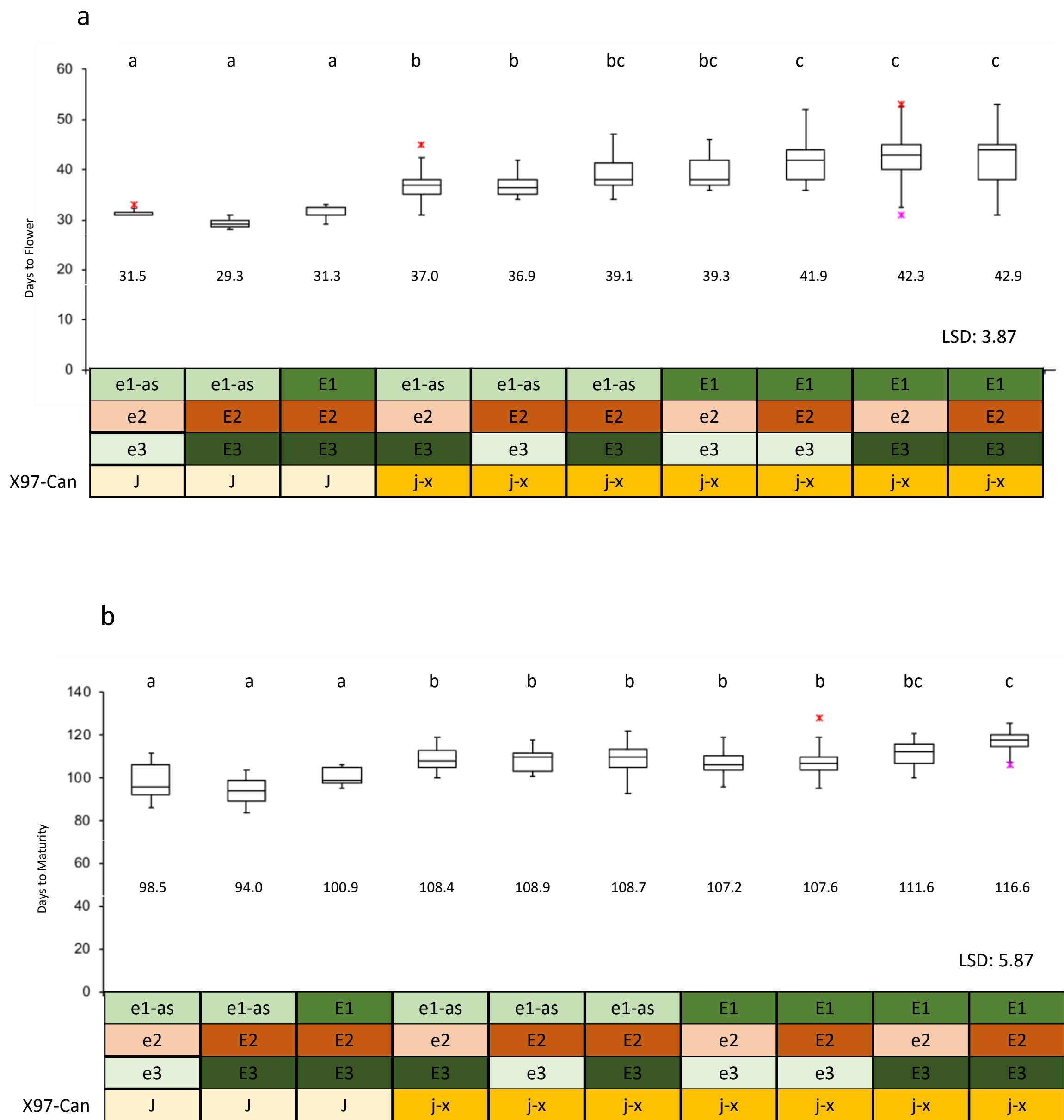

Figure 8. Days to flower and days to maturity for RILs from the X97-Can population. Lines were segregating for *E1/e1-as, E2/e2, E3/e3* and *J/j-x*. Parents have a black background. Data from the individual RILs were analyzed together based on their genotype. Means are shown under boxplots. a: Days to flower b: Days to maturity

TABLES

Table 1: Genotype data and days to flower and maturity for RIL parents.

Parental line	E1	E2	E3	Dt1	Elf3	Pdh1	Days to Flower*	Days to Maturity**
Paranagoiana	*E1*	*E2*	*E3*	*dt1 R166W*	*j-x* unknown	*pdh1*	48.00	113.56
PI 159925	*E1*	*E2*	*E3*	*dt1 R166W*	*j-1* C del exon 4	*Pdh1*	44.00	111.17
Jenguma	*E1*	*E2*	*E3*	*dt1 P113L*	*ELF3/J?* No poly. in *ELF3*	*pdh1*	44.67	115.88
Canadian X	*E1*	*e2*	*e3*	*Dt1*	*j-x* unknown	*pdh1*	39.44	105.07
Jake	*E1*	*E2*	*E3*	*dt1 R166W*	*J*	*pdh1*	31.81	92.82
X97	*e1-as*	*E2*	*E3*	*Dt1*	*J*	*pdh1*	29.89	92.29
534545	*e1-as*	*E2*	*E3*	*Dt1*	*J*	*pdh1*	29.29	87.17

*Days to flower are means of two-year data collected at 4 locations in northern Ghana (excludes WaS).
**Days to maturity are means of two-year data collected at 5 locations in northern Ghana.

Table 2: RIL population names, parents, and segregating genes of interest.

Population name	LJ parent	Long Juvenile Parent	Seg genes of interest						# of RILS planted in 2016	# of RILS planted in 2017
Jake-15*	Jake	PI 159925	J/ j-1					Pdh1/pdh1	20	9
Jake-Pa*	Jake	Paranagoiana	J/ j-x						18	14
X97-15	X97	PI 159925	J/ j-1	E1/e1-as	Dt1/ dt1 R166W			Pdh1/pdh1	47	5
X97-Jen	X97	Jenguma	N/A**	E1/e1-as	Dt1/ dt1 P113L				60	41
X97-Can	X97	Canadian X	J/ j-x	E1/e1-as		E2/e2	E3/e3		39	25
534-Can	534545	Canadian X	J/ j-x	E1/e1-as		E2/e2	E3/e3		47	33

** RILs in Jake-15 and Jake-Pa were selected for the long juvenile trait*

***The source of delayed flowering in short days has not been determined in Jenguma, but it does not have polymorphisms in ELF3*

Table 3: Genotype data and days to flower and maturity for other inbred lines of interest.

Parental line	Country	E1	E2	E3	Dt1	Elf3	Pdh1	Days to Flower*	Days to Maturity**
S12-3187	USA	*E1*	*E2*	*E3*	*Dt1*	*J*	*pdh1*	32.56	93.14
S12-1403	USA	*E1*	*E2*	*E3*	*Dt1*	*J*	*pdh1*	33.47	94.50
S12-5127	USA	*E1*	*E2*	*E3*	*Dt1*	*J*	*pdh1*	33.86	97.20
Walima	Mozambique	*E1*	*e2*	*E3*	*NA*	*j-2*	*pdh1*	43.31	98.88
Melrose	Australia	*E1*	*E2*	*E3*	*dt1 R62S*	*j-1*	*pdh1*	37.94	100.44
X5683-1-33 F718	Canada	*E1*	*e2*	*E3*	*dt1 R166W*	*j-x*	*pdh1*	37.56	101.14
Wima	Mozambique	*E1*	*E2*	*E3*	*dt1 P113L*	*R73G*	*pdh1*	41.88	104.25
x5683-1-18 F718	Canada	*E1*	*e2*	*e3*	*Dt1*	*j-x*	*pdh1*	39.44	105.07
Suong-Pungu	Ghana	*E1*	*E2*	*E3*	*dt1 R166W*	No mutation	*pdh1*	42.44	105.81
Zamboani	Mozambique	*E1*	*E2*	*E3*	*dt1 P113L*	*R73G, R308M*	*pdh1*	42.67	106.17
X5683-1-18 F728	Canada	*E1*	*E2*	*E3*	*Dt1*	*j-x*	*pdh1*	38.33	108.31
Vernal	Australia	*E1*	*E2*	*E3*	*dt1 R166W*	*j-1*	*pdh1*	47.00	111.13
Songda	Ghana	*E1*	*E2*	*E3*	*dt1 P113L*	*R73G*	*Pdh1*	48.63	116.00
Afayak	Ghana	*E1*	*E2*	*E3*	*dt1 R166W*	No mutation	*pdh1*	46.73	116.73
Quarshie	Ghana	*E1*	*E2*	*E3*	*dt1 P113L*	No mutation	*pdh1*	46.50	117.25
Sal-II	Ghana	*E1*	*E2*	*E3*	*dt1 P113L*	*R73G, R308M*	*Pdh1*	50.90	122.94

*Days to flower are means of two-year data collected at 4 locations in northern Ghana (excludes WaS).
**Days to maturity are means of two-year data collected at 5 locations in northern Ghana.

REFERENCES

Abate, T., A. D. Alene, D. Bergvinson, B. Shiferaw, S. Silim, A. Orr and S. Asfaw (2012). Tropical grain legumes in Africa and south Asia: knowledge and opportunities, International Crops Research Institute for the Semi-Arid Tropics.

Abate, T. and A. Orr (1981). "Research and development for tropical legumes: Towards a knowledge-based strategy." Journal of SAT Agricultural Research **10**: 1-12.

Alene, A., O. Coulibaly and T. Abdoulaye (2012). "The world cowpea and soybean economies: Facts, trends, and outlook." Lilongwe, Malawi: International Institute of Tropical Agriculture.

Bandillo, N. B., J. E. Anderson, M. B. Kantar, R. M. Stupar, J. E. Specht, G. L. Graef and A. J. Lorenz (2017). "Dissecting the Genetic Basis of Local Adaptation in Soybean." Scientific reports **7**(1): 17195.

Bernard, R. L. (1971). "Two major genes for time of flowering and maturity in soybeans." Crop Sci **11**: 242-244.

Beuselinck, P., D. Sleper and K. Bilyeu (2006). "An assessment of phenotype selection for linolenic acid using genetic markers." Crop science **46**(2): 747-750.

Bilyeu, K. D. and W. J. Wiebold (2016). "Environmental stability of seed carbohydrate profiles in soybeans containing different alleles of the raffinose synthase 2 (RS2) gene." Journal of agricultural and food chemistry **64**(5): 1071-1078.

Bonato, E. R. and N. A. Vello (1999). "E6, a dominant gene conditioning early flowering and maturity in soybeans." Genetics and Molecular Biology **22**(2): 229-232.

Buzzell, R. (1971). "Inheritance of a soybean flowering response to fluorescent-daylength conditions." Canadian Journal of Genetics and Cytology **13**(4): 703-707.

Buzzell, R. and H. Voldeng (1980). "Inheritance of insensitivity to long daylength." Soybean Genet. Newsl **7**: 26-29.

Carpentieri-Pipolo, V., L. Alves de Almeida and R. Afonso de S. Kiihl (2002). "Inheritance of a long juvenile period under short-day conditions in soybean." Genet Mol Biol **25**(4): 463-469.

Carter, T. E., R. L. Nelson, C. H. Sneller and Z. Cui (2004). "Genetic diversity in soybean." Soybeans: Improvement, production, and uses(soybeansimprove): 303-416.

Cober, E., J. Tanner and H. Voldeng (1996). "Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines." Crop science **36**(3): 601-605.

Cober, E. R. (2011). "Long juvenile soybean flowering responses under very short photoperiods." Crop science **51**(1): 140-145.

Cober, E. R. and H. D. Voldeng (2001). "A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*." Crop Sci **41**: 698-701.

Destro, D., V. Carpentieri-Pipolo, R. Afonso de S. Kiihl and L. Alves de Almeida (2001). "Photoperiodism and genetic control of the long juvenile period in soybean: a review." Crop Breed Appl Biot **1**(1): 72-92.

Dogbe, W., P. M. Etwire, E. Martey, J. C. Etwire, I. I. Baba and A. Siise (2013). "Economics of Soybean Production: Evidence from Saboba and Chereponi Districts of Northern Region of Ghana." Journal of Agricultural Science **5**(12): 38.

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

Funatsuki, H., M. Suzuki, A. Hirose, H. Inaba, T. Yamada, M. Hajika, K. Komatsu, T. Katayama, T. Sayama and M. Ishimoto (2014). "Molecular basis of a shattering resistance boosting global dissemination of soybean." Proceedings of the National Academy of Sciences **111**(50): 17797-17802.

Garner, W. W. and H. A. Allard (1920). "Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants1." Monthly Weather Review **48**(7): 415-415.

Goldsmith, P. D. (2008). "Economics of soybean production, marketing and utilization." Soybeans Chemistry, Production, Processing, and Utilization: 117-150.

Hartwig, E. E. and C. J. Edwards (1970). "Effects of morphological characteristics upon seed yield in soybeans." Agronomy Journal **62**(1): 64-65.

Hartwig, E. E. and R. A. Kiihl (1979). "Identification and utilization of a delayed flowering character in soybeans for short-day conditions." Field Crops Research **2**: 145-151.

Hymowitz, T. (1970). "On the domestication of the soybean." Economic Botany **24**(4): 408-421.

IITA. (2014). "Soybean." from http://www.iita.org/soybean.

Kiihl, R. and A. Garcia (1989). The use of long juvenile trait in breeding soybean cultivars. th Conferencia Mundial de Investigación en Soja, Buenos Aires, Argentina.

Langewisch, T., J. Lenis, G.-L. Jiang, D. Wang, V. Pantalone and K. Bilyeu (2017). "The development and use of a molecular model for soybean maturity groups." BMC plant biology **17**(1): 91.

Langewisch, T., H. Zhang, R. Vincent, T. Joshi, D. Xu and K. Bilyeu (2014). "Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes." PLoS One **9**(4): e94150.

Li, X., C. Fang, M. Xu, F. Zhang, S. Lu, H. Nan, T. Su, S. Li, X. Zhao and L. Kong (2017). "Quantitative Trait Locus Mapping of Soybean Maturity Gene E6." Crop science **57**(5): 2547-2554.

Liu, B., A. Kanazawa, H. Matsumura, R. Takahashi, K. Harada and J. Abe (2008). "Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene." Genetics **180**(2): 995-1007.

Liu, G., L. Zhao, B. J. Averitt, Y. Liu, B. Zhang, R. Chang, Y. Ma, X. Luan, R. Guan and L. Qiu (2015). "Geographical distribution of GmTfl1 alleles in Chinese soybean varieties." The Crop Journal **3**(5): 371-378.

Lu, S., X. Zhao, Y. Hu, S. Liu, H. Nan, X. Li, C. Fang, D. Cao, X. Shi and L. Kong (2017). "Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield." Nature Genetics **49**(5): 773-779.

Masuda, T. and P. D. Goldsmith (2009). "World soybean production: area harvested, yield, and long-term projections." International Food and Agribusiness Management Review **12**(4): 143-162.

Mbanya, W. (2011). "Assessment of the Constraints in Soybean Production: A Case of Northern Region, Ghana." Journal of Developments in Sustainable Agriculture **6**(2): 199-214.

Palacios, M., R. Easter, K. Soltwedel, C. Parsons, M. Douglas, T. Hymowitz and J. Pettigrew (2004). "Effect of soybean variety and processing on growth performance of young chicks and pigs 1." Journal of animal science **82**(4): 1108-1114.

Ray, J., K. Hinson, E. Mankono and M. Malo (1995). "Genetic Control of a Lon-Juvenile Trait in Soybean." Crop Science **35**: 1001-1006.

Shannon, J., J. Wrather, D. Sleper, R. Robbins, H. Nguyen and S. Anand (2007). "Registration of 'Jake'soybean." Journal of plant registrations **1**(1): 29-30.
Sinclair, T. R. and K. Hinson (1992). "Soybean flowering in response to the long-juvenile trait." Crop science **32**(5): 1242-1248.

Spehar, C. (1995). "Diallel analysis for mineral element absorption in tropical adapted soybeans [Glycine max (L.) Merrill]." Theoretical and Applied Genetics **90**(5): 707-713.

Watanabe, S., K. Harada and J. Abe (2012). "Genetic and molecular bases of photoperiod responses of flowering in soybean." Breed Sci **61**(5): 531-543.

Watanabe, S., R. Hideshima, Z. Xia, Y. Tsubokura, S. Sato, Y. Nakamoto, N. Yamanaka, R. Takahashi, M. Ishimoto and T. Anai (2009). "Map-based cloning of the gene associated with the soybean maturity locus E3." Genetics **182**(4): 1251-1262.

Watanabe, S., T. Tajuddin, N. Yamanaka, M. Hayashi and K. Harada (2004). "Analysis of QTLs for reproductive development and seed quality traits in soybean using recombinant inbred lines." Breeding science **54**(4): 399-407.

Watanabe, S., Z. Xia, R. Hideshima, Y. Tsubokura, S. Sato, N. Yamanaka, R. Takahashi, T. Anai, S. Tabata, K. Kitamura and K. Harada (2011). "A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering." Genetics **188**(2): 395-407.

Whigham, D. K. and H. C. Minor (1978). "Agronomic characteristics and environmental stress." Soybean physiology, agronomy, and utilization. Academic Press, New York: 77-112.

Xia, Z., S. Watanabe, T. Yamada, Y. Tsubokura, H. Nakashima, H. Zhai, T. Anai, S. Sato, T. Yamazaki, S. Lu, H. Wu, S. Tabata and K. Harada (2012). "Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering." Proc Natl Acad Sci U S A **109**(32): E2155-2164.

Xia, Z., H. Zhai, B. Liu, F. Kong, X. Yuan, H. Wu, E. R. Cober and K. Harada (2012). "Molecular identification of genes controlling flowering time, maturity, and photoperiod response in soybean." Plant Syst Evol **298**: 1217-1227.

Xu, M., Z. Xu, B. Liu, F. Kong, Y. Tsubokura, S. Watanabe, Z. Xia, K. Harada, A. Kanazawa, T. Yamada and J. Abe (2013). "Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean." BMC Plant Biol **13**: 91.

Yue, Y., N. Liu, B. Jiang, M. Li, H. Wang, Z. Jiang, H. Pan, Q. Xia, Q. Ma and T. Han (2017). "A single nucleotide deletion in J encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean." Molecular plant **10**(4): 656-658.

Zhang, L., S. Kyei-Boagen, J. Zhang, M. Zhang, T. Freeland and C. Watson, Jr. (2007). "Modifications of optimum adaptation zones for soybean maturity groups in the USA." Crop Management **6**.

CHAPTER FOUR

The Effects of *Dt1* on Days to Flower, Days to Maturity, and Height in a Tropical Environment

SUMMARY

Height is an important trait for adaptation to a specific environment for the ability to affect yield and also to prevent lodging. There is a careful balance between yield and lodging that creates an optimal height for each environment. Height is a quantitative trait that is also affected by the environment, however there are two genes, *Dt1* and *Dt2,* that play a large role in controlling terminal stem elongation. *Dt1* is the wild type allele that allows for stem growth after flowering, *dt1* is the polymorphic allele that ceases terminal stem growth after the first flower appears. There are 4 known polymorphisms of *Dt1* that may have different effects on the height phenotype. In addition, *Dt1* encodes a florigen protein which is known to be an important factor in the *Arabidopsis* flowering pathway and may also have a similar role in soybean. The objective of this research is to understand the role of *Dt1* and two of its alleles: *R166W* and *P133L* on height, days to flower, and days to maturity in a tropical environment.

INTRODUCTION

Height is an important trait that affects adaptation and yield in soybean [*Glycine max* (L.) Merr.] (Cober and Morrison 2010). Two hundred and fifty-five minor QTLs have been published to date as associated with the soybean height phenotype (SoyBase, www.soybase.org), however there is still little understanding about their role (Zhang et al. 2015). Two genes have been discovered that play a major role in stem elongation and ultimately height: *Dt1* and *Dt2*. Both of these genes influence indeterminate, determinate, or semi determinate growth habits (Bernard 1972, Tian et al. 2010). To date, knowledge does not exist if alleles of *Dt1* or these growth habits behave the same way in a tropical climate.

Dt1 is a well characterized gene that regulates stem elongation in soybean. It is an ortholog of the Arabidopsis (*Arabidopsis thaliana*) *TERMINAL FLOWER1* and encodes *GmTFL1b*, which promotes stem elongation (Liu et al. 2010, Tian et al. 2010). When *Dt1* is mutated, *dt1*, *GmTFL1b* expression is similar to the wild type *Dt1* until the inductive phase of flowering begins then expression is lost in the shoot apical meristem (Liu et al. 2010). This results in termination of main stem growth when the first flowers begin and is called the determinate growth habit (Bernard 1972). Determinate soybean types have larger diameter main stems that provide lodging resistance. The indeterminate growth habit is the ancestral phenotype where main stem growth continues past flowering (Bernard 1972). The semi-determinates (*Dt2*), which express in the indeterminate genetic background only, a phenotype of intermediate stature with a terminal raceme

(Bernard 1972). The *Dt2* gene was recently shown to be a gain of function MADS-Domain Factor gene that was thought to regulate the *Dt1* gene (Ping et al. 2014).

Of the three known variations of stem termination, the most common in North America and ancestrally are the indeterminates, (*Dt1),* which are grown from the Midwest all through Canada (Bernard 1972, Thompson et al. 1997). Determinates, (*dt1*), are most common in the southern USA (Thompson et al. 1997, Tian et al. 2010). Semi determinates are not common in the United States (Bernard 1972). The preference for each growth habit in different environments is due to its effect on lodging. In the northern United States the shorter season length allows for the indeterminate growth type, but the longer seasons of the southern United States need the determinate growth type to prevent lodging (Bernard 1972).

There are four identified missense mutations of *Dt1* that can cause the *dt1* genotype. An arginine to serine mutation at position 62 (R62S), a proline to leucine mutation at position 113 (P113L), an arginine to lysine mutation at position 130 (R130K), and an arginine to tryptophan mutation at position 166 (R166W) (Tian et al. 2010). It is not understood if these alleles have different effects on height (Thompson et al. 1997).

There is also speculation the *Dt1* and *dt1* may influence days to flower or maturity (Bernard 1972) although other work suggests it may not (Tian et al. 2010).
The objective of this research was to understand if the *Dt1* alleles influence days to flower and days to maturity in a tropical climate. We also investigated the effect on

height of *Dt1* and two alleles of *dt1*: R166W and P113L on height in the tropics. These results will help breeders understand what height phenotypes are possible in low latitudes, and if manipulation of those alleles will affect flowering or season length.

MATERIALS AND METHODS

Please see pages 66-74.

RESULTS

Determinate Varieties Are the Most Accessible Varieties in Ghana and Mozambique

To understand the current preferred alleles of *Dt1* in Ghana and Mozambique, we genotyped 9 released, commonly grown varieties for *Dt1*. Only determinate, *dt1* alleles were found in the African varieties tested. The most common allele was *dt1* P113L, which was found in 7 varieties and then *dt1* R166W, which was in 2 varieties (Table 1).

Population and Location Effects and Their Interaction Influence Days to Flower; Genotype, Population, Location and Population x Location Affect Days to Maturity

Analysis of variance was conducted for days to flower (dtf) and days to maturity (dtm) for all RILs and categorized by *Dt1* or *dt1* allele in all populations regardless of genetic background (Jake-15, Jake-Pa, X97-15, X97-Jen, 534-Can, and X97-Can). The analysis considered the *Dt1/dt1* genotype, population, location, and interactions between rep and

location, genotype and location, and population and location (Table 2). In the type III error analysis, genotype was not significant for days to flower, but population and location were and also for the population by location interaction. However, the means comparison for dtf between *Dt1* and *dt1* did show there was a significant difference of two days (Figure 1a).

Analysis of variance for dtm with the same conditions as described for dtf was conducted. The *Dt1/dt1* genotype was significant as was population, location, and the interaction between population and location (Table 2). In the means comparison to dtm, there was no significant difference (Figure 1b).

To minimize population effect, we conducted an analysis of variance on the two populations that were segregating for *Dt1/dt1*, X97-15 and X97-Jen. Genotype was not significant for dtf, but population, location and their interaction were. (Table 3). There was no difference in the mean days to flower for *Dt1* and *dt1* (Figure 2a). Genotype, population, location, and population x location were significant for dtm (Table 3) and the mean days to maturity were significantly different for *Dt1* and *dt1* although the difference was only one day (Figure 2b).

Dt1 Does Not Affect Days to Flower or Days to Maturity in a Tropical Environment

Analysis of variance was conducted for dtf and dtm in the X97-15 population to eliminate the population effect observed. The ANOVA was not able to produce a useful model for

dtf, suggesting that *Dt1* and the other variables tested were not important for affecting days to flower (Table 4). Analysis of variance for dtm shows that genotype was not significant for influencing that trait, but location was (Table 4). Dtm means were significantly different for *Dt1* and *dt1* alleles by 3 days even though the genotype was not significant (Figure 3).

The same analysis was conducted in the X97-Jen population. The *Dt1* genotype was not significant with either dtf or dtm, only location was (Table 5).

E1 and *J* Affect dtf But Their Interaction with *Dt1* does not; *Dt1* Affects dtm in a Conventional Background

To test if an interaction between the maturity genes: *E1* and *J* and *Dt1* affected dtf and dtm, data was categorized by the genotype of those three genes in the X97-15 population resulting in 6 genotype groups: *E1/J/Dt1, E1/J/dt1, e1-as/j/Dt1, e1-as/j/dt1, E1/j/Dt1,* and *E1/j/dt1*. Analysis of variance of dtf shows a significant genotype effect and location was not significant (Table 6). A means comparison of dtf shows there are significant differences among genotype groups with different *E1* and *J* alleles, regardless of their *Dt1* status (Figure 4a). The ANOVA for dtm shows that genotype and location significantly affect the trait (Table 6). A means comparison of dtm shows the *j-1* allele affects maturity; however in a *J* background, alleles of *Dt1* influence a significant difference (Figure 4b).

Dt1, Environment, and Genetic Background Influence Height

To understand the influence of *Dt1* on height, height data from all populations (Jake-15, Jake-Pa, X97-15, X97-Jen, 534-Can, and X97-Can) was grouped together based on the *Dt1* or *dt1* allele status of each line. Analysis of variance was performed on all data (Table 7). *Dt1* genotype, population, and location were all significant variables affecting the height phenotype. A means comparison of height of *Dt1* and *dt1* show a significant difference, where determinate lines were 9 cm shorter than indeterminate lines (Figure 5). Determinates on average achieved 83% of the height of indeterminates.

Determinate Alleles *R166W* and *P113L* Height Means are Not Significantly Different

To test if two *dt1* mutant alleles have different effects on height, data from populations were grouped according to their allele status of *Dt1, dt1 R166W,* and *dt1 P113L.* ANOVA results show that there is a genotype, population, and location effect on height (Table 8). A means comparison of *Dt1, dt1 R166W,* and *dt1 P113L* shows that there is a significant difference between *Dt1* and *dt1* alleles, however there is no difference between the *dt1 R166W* and *dt1 P113L* alleles (Figure 6).

Dt1 and Environment Influence Height

To test if the environmental effect is eliminated if the population variable is removed, we tested the effects of height and *Dt1* in the X97-Jen population. Analysis of variance

shows that genotype and location affect the height in this one population (Table 9). The means between *Dt1* and *dt1* are still significantly different, and the determinate lines were 80% the height of the indeterminates (Figure 7).

DISCUSSION

To maximize adaptation, it is important to understand how major traits affecting yield perform in a tropical climate. Height is an important trait that can influence yield not only by allowing area for pod producing nodes to be added, but an optimal height can also prevent lodging, which can cause yield loss (Bernard 1972). Height has often been discussed simultaneously with maturity for this reason. We were surprised to detect *dt1* alleles in the nine released African soybean varieties examined because it seemed there would be selection for indeterminate types in tropical environments where selection for delayed flowering to support optimum yields was a key adaptation feature.

We have shown that *Dt1* does not affect days to flower or days to maturity in a tropical environment. Analysis of dtf and dtm by *Dt1* alone and in combination with the other maturity genes *E1* and *J* across several populations and in single populations did not show significant effects. These results allow for tropical breeders to experiment with different *dt1* alleles and with the semideterminate and tall determinate growth habits without affecting season length (Bernard 1972). It is important to understand what the optimal stem height is to maximize yield while reducing lodging probability.

We have also demonstrated that there is a significant difference in height between the indeterminate and determinate experimental lines, although analysis with a single population showed that environment influenced height as well. Interestingly, we did not find a difference in height between two alleles of *dt1:* R166W and P113L. P113L is the most common *dt1* allele in the African varieties we tested and R166W is the most common determinate allele found in North American soybean. The reason for the lack of the indeterminate trait in African varieties is unknown. It may not have been available in the imported germplasm necessary for tropical adaptation or simply it could have been selected against. Currently the effects of the indeterminate and determinate traits on yield are not known. We collected yield data; however due to inconsistent plant stands, the data quality was too low for full analyses. Lodging was minimal in indeterminate RILs suggesting that the indeterminate allele may be useful to increasing yield; however, the effect of the indeterminate trait and internode length in a tropical climate is not known either.

FIGURES

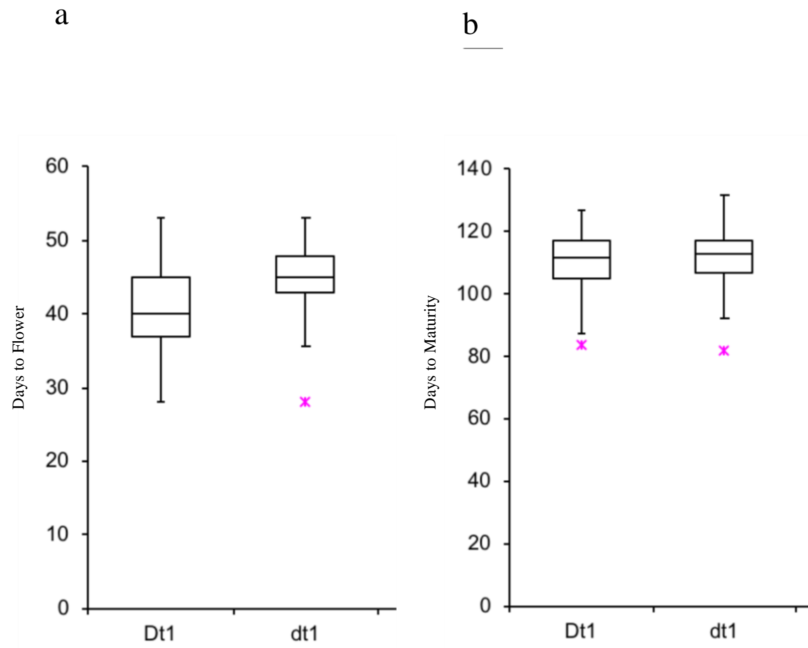

Figure 1: Days to flower and days to maturity of *Dt1* and *dt1* alleles in all populations. a: Days to flower. b: Days to maturity.

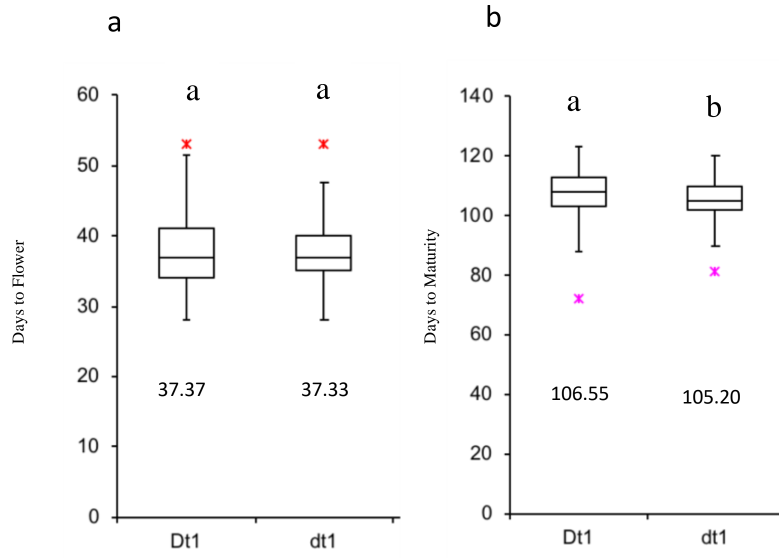

Figure 2: Days to flower and days to maturity in two populations that were segregating for *Dt1/dt1*, X97-15 and X97-Jen. Means for each genotype are shown under the boxplot. a: Days to flower. b: Days to maturity.

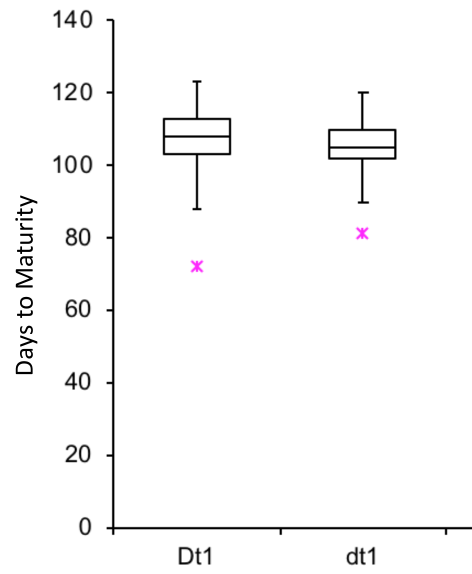

Figure 3: Days to maturity in the X97-15 population that was segregating for *Dt1/dt1*.

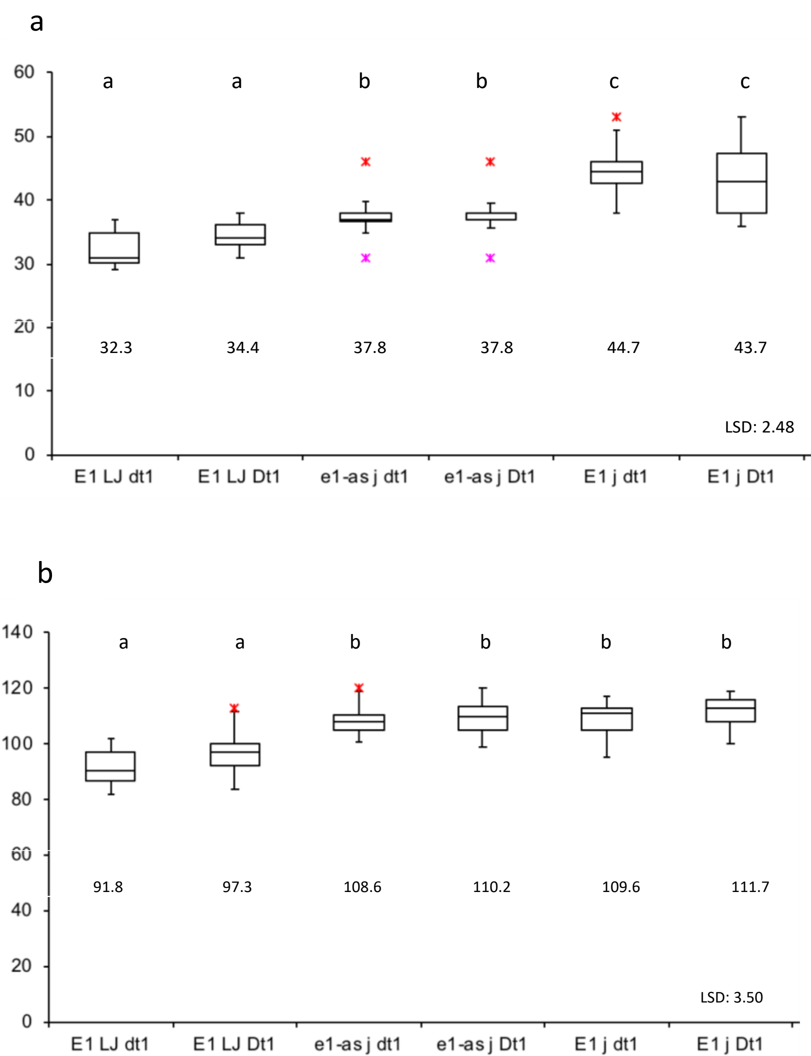

Figure 4. Days to flower and days to maturity X97-15 that was segregating for *E1/e1-as, J/j-1,* and *Dt1/dt1*. Means for each genotype are shown under the boxplot. a: Days to flower. b: Days to maturity.

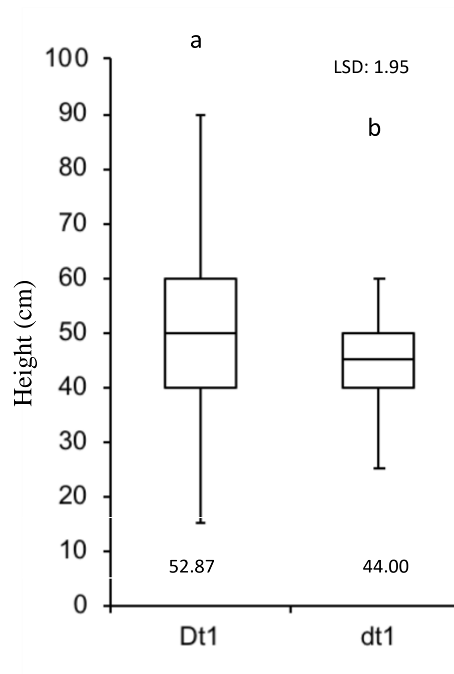

Figure 5. Height for alleles *Dt1* and *dt1* in all populations. Means are shown under the boxplots.

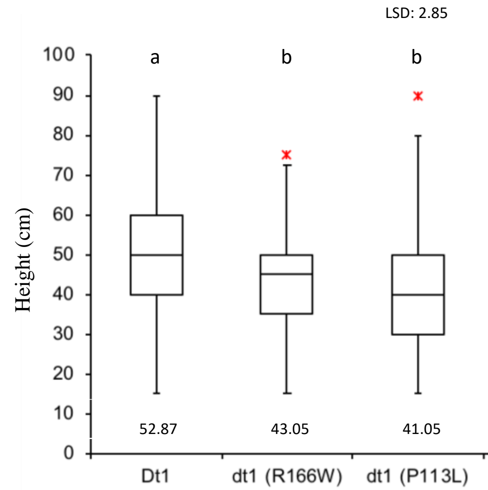

Figure 6. Height for alleles *Dt1* and *dt1* (R166W) and *dt1* (P133L) in all populations. Means are shown under the boxplots.

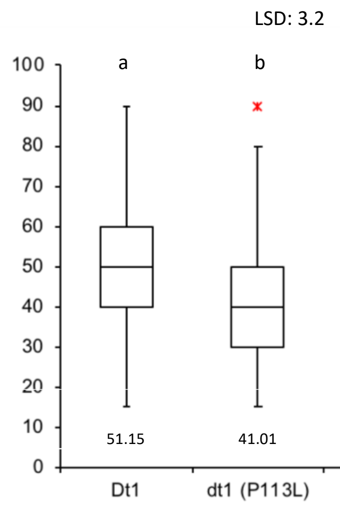

Figure 7. Height for alleles *Dt1* and *dt1* (P133L) in the X97-Jen population. Means are shown under the boxplots.

TABLES

Table 1: *Dt1* alleles of released African varieties and the country in which they were released

Dt1 alleles of Released African Varieties		
Variety Name	**Country**	***Dt1 allele***
Wima	Mozambique	*dt1* P113L
Zamboani	Mozambique	*dt1* P113L
Afayak	Ghana	*dt1* R166W
Jenguma	Ghana	*dt1* P113L
Quarshie	Ghana	*dt1* P113L
Salintuya-I	Ghana	*dt1* P113L
Salintuya-II	Ghana	*dt1* P113L
Songda	Ghana	*dt1* P113L
Suong-Pungu	Ghana	*dt1* R166W

Table 2: ANOVA results for *Dt1* alleles on two traits in all six RIL populations: days to flower and days to maturity. P values are shown for all variables of each trait.
DTF = days to flower DTM= days to maturity

		Dt1/ dt1 in all populations	
	df	DTF*** R-sq: 0.38 CV: 11.4	DTM*** R-sq: 0.28 CV: 6.36
Genotype	1	0.39	0.0016
Population	5	<0.0001	<0.0001
Location	4	<0.0001	<0.0001
Rep(Location)	5	0.923	0.37
Genotype x Location	4	0.989	0.1273
Population x Location	20	0.0019	<0.0001

*** Model significance at <0.0001

Table 3: ANOVA results for days to flower and days to maturity in two populations that were segregating for *Dt1/dt1*, X97-15 and X97-Jen. P values are shown for all variables of each trait.
DTF = days to flower DTM= days to maturity

		Dt1/ dt1 in X97-15 and X97-Jen combined	
	df	DTF*** R-sq: 0.18 CV: 12.13	DTM*** R-sq: 0.26 CV: 6.92
Genotype	1	0.3837	0.0030
Population	5	<0.0001	<0.0001
Location	4	<0.0001	<0.0001
Rep(Location)	5	0.8265	0.6814
Genotype x Location	4	0.9983	0.1750
Population x Location	20	0.0019	0.0336

*** Model significance at <0.0001

Table 4: ANOVA results for days to flower and days to maturity in the X97-15 population that was segregating for *Dt1/dt1*. P values are shown for all variables of each trait.
DTF = days to flower DTM= days to maturity

		Dt1/ dt1 in X97-15	
		DTF[NS]	DTM***
	df	R-sq: 0.10 CV: 13.97	R-sq: 0.28 CV: 7.2
Genotype	1	0.4057	0.1467
Location	4	0.028	<0.0001
Rep(Location)	5	0.774	0.6677
Genotype x Location	4	0.6475	0.6248

*** Model significance at <0.0001
NS Model not significant

Table 5: ANOVA results for days to flower and days to maturity in the X97-Jen population that was segregating for *Dt1/dt1*. P values are shown for all variables of each trait.
DTF = days to flower DTM= days to maturity

		Dt1/ dt1 in X97-Jen	
	df	DTF*** R-sq: 0.09 CV: 11.34	DTM*** R-sq: 0.23 CV: 6.82
Genotype	1	0.8701	0.0544
Location	4	<0.0001	<0.0001
Rep(Location)	5	0.8443	0.2619
Genotype x Location	4	0.9447	0.2755

*** Model significance at <0.0001

Table 6: ANOVA results for days to flower and days to maturity for X97-15 on allelic combinations of *E1, J,* and *Dt1*. P values are shown for all variables of each trait.
DTF = days to flower DTM= days to maturity

		E1, J, and *Dt1* in X97-15	
	df	DTF*** R-sq: 0.61 CV: 9.04	DTM*** R-sq: 0.68 CV: 4.71
Genotype	5	<0.0001	<0.0001
Location	4	0.0709	<0.0001
Rep(Location)	4	0.1077	0.0767
Genotype x Location	12	0.2609	0.1320

*** Model significance at <0.0001

Table 7: ANOVA results for height in all six RIL populations with the *Dt1* or *dt1* allele. P values are shown for all variables of each trait.

		Dt1/dt1 in X97-Jen and X97-15
	df	Height*** R-sq: 0.25 CV: 27.08
Genotype	1	<0.0001
Population	5	<0.0001
Location	3	<0.0001
Rep(Location)	4	0.3918
Genotype x Location	3	0.5914
Population x Location	15	0.0947

*** Model significance at <0.0001

Table 8: ANOVA results for height in all populations with the *Dt1* or *dt1* (R166W) or dt1 (P113L) allele: X97-15 and X97-Jen. P values are shown for all variables of each trait.

		Dt1/ dt1 in all populations
	df	Height*** R-sq: 0.26 CV: 27.12
Genotype	1	<0.0001
Population	5	<0.0001
Location	3	<0.0001
Rep(Location)	4	0.3929
Genotype x Location	6	0.7588
Population x Location	15	0.1621

*** Model significance at <0.0001

Table 9: ANOVA results for height in the X97-Jen population with the *Dt1* and dt1 (P113L) alleles. P values are shown for all variables of each trait.

		Dt1 X97-Jen
	df	Height*** R-sq: 0.23 CV: 28.72
Genotype	5	<0.0001
Location	4	<0.0001
Rep(Location)	4	0.8830
Genotype x Location	12	0.6418

*** Model significance at <0.0001

REFERENCES

Bernard, R. (1972). "Two genes affecting stem termination in soybeans." crop Science **12**(2): 235-239.

Beuselinck, P., D. Sleper and K. Bilyeu (2006). "An assessment of phenotype selection for linolenic acid using genetic markers." Crop science **46**(2): 747-750.

Bilyeu, K. D. and W. J. Wiebold (2016). "Environmental stability of seed carbohydrate profiles in soybeans containing different alleles of the raffinose synthase 2 (RS2) gene." Journal of agricultural and food chemistry **64**(5): 1071-1078.

Bonato, E. R. and N. A. Vello (1999). "E6, a dominant gene conditioning early flowering and maturity in soybeans." Genetics and Molecular Biology **22**(2): 229-232.

Cober, E. R. (2011). "Long juvenile soybean flowering responses under very short photoperiods." Crop science **51**(1): 140-145.

Cober, E. R. and M. J. Morrison (2010). "Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean." Theoretical and applied genetics **120**(5): 1005-1012.

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

Funatsuki, H., M. Suzuki, A. Hirose, H. Inaba, T. Yamada, M. Hajika, K. Komatsu, T. Katayama, T. Sayama and M. Ishimoto (2014). "Molecular basis of a shattering resistance boosting global dissemination of soybean." Proceedings of the National Academy of Sciences **111**(50): 17797-17802.

Langewisch, T., J. Lenis, G.-L. Jiang, D. Wang, V. Pantalone and K. Bilyeu (2017). "The development and use of a molecular model for soybean maturity groups." BMC plant biology **17**(1): 91.

Li, X., C. Fang, M. Xu, F. Zhang, S. Lu, H. Nan, T. Su, S. Li, X. Zhao and L. Kong (2017). "Quantitative Trait Locus Mapping of Soybean Maturity Gene E6." Crop science **57**(5): 2547-2554.

Liu, B., S. Watanabe, T. Uchiyama, F. Kong, A. Kanazawa, Z. Xia, A. Nagamatsu, M. Arai, T. Yamada and K. Kitamura (2010). "The soybean stem growth habit gene Dt1 is an ortholog of Arabidopsis TERMINAL FLOWER1." Plant Physiology **153**(1): 198-210.

Liu, G., L. Zhao, B. J. Averitt, Y. Liu, B. Zhang, R. Chang, Y. Ma, X. Luan, R. Guan and L. Qiu (2015). "Geographical distribution of GmTfl1 alleles in Chinese soybean varieties." The Crop Journal **3**(5): 371-378.

Lu, S., X. Zhao, Y. Hu, S. Liu, H. Nan, X. Li, C. Fang, D. Cao, X. Shi and L. Kong (2017). "Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield." Nature Genetics **49**(5): 773-779.

Palacios, M., R. Easter, K. Soltwedel, C. Parsons, M. Douglas, T. Hymowitz and J. Pettigrew (2004). "Effect of soybean variety and processing on growth performance of young chicks and pigs 1." Journal of animal science **82**(4): 1108-1114.

Ping, J., Y. Liu, L. Sun, M. Zhao, Y. Li, M. She, Y. Sui, F. Lin, X. Liu and Z. Tang (2014). "Dt2 is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean." The Plant Cell Online **26**(7): 2831-2842.

Ray, J. D., K. Hinson, E. B. Mankono and M. F. Malo (1995). "Genetic control of a long-juvenile trait in soybean." Crop Sci **35**: 1001-1006.

Shannon, J., J. Wrather, D. Sleper, R. Robbins, H. Nguyen and S. Anand (2007). "Registration of 'Jake'soybean." Journal of plant registrations **1**(1): 29-30.

Thompson, J. A., R. L. Bernard and R. L. Nelson (1997). "A third allele at the soybean dt1 locus." Crop science **37**(3): 757-762.

Tian, Z., X. Wang, R. Lee, Y. Li, J. E. Specht, R. L. Nelson, P. E. McClean, L. Qiu and J. Ma (2010). "Artificial selection for determinate growth habit in soybean." Proceedings of the National Academy of Sciences **107**(19): 8563-8568.

Zhang, J., Q. Song, P. B. Cregan, R. L. Nelson, X. Wang, J. Wu and G.-L. Jiang (2015). "Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (Glycine max) germplasm." BMC genomics **16**(1): 217.

APPENDIX 1

Supplementary data for Chapter Two

FIGURES

TABLES

Table 1:

IITA lines and their Pdh1 allele status
Using the SimpleProbe perfect marker, we determined the allele of *pdh1* present in the IITA germplasm collection.
A blank square indicates that the shatter resistant *pdh1* allele was present. Pdh1 written means that line has
the shatter prone *Pdh1* allele. NS means no data was determined.

Line name	Pdh1 allele status
TGx 1670-5F	
TGx 1925-1F	
TGx 1927-5F	
TGx 1984-8F	
TGx 1985-4F	
TGx 1987-11F	Pdh1
TGx 1987-15F	Pdh1
TGx 1987-17F	Pdh1
TGx 1989-129F	NS
TGx 1990-12F	NS
TGx 1990-43F	NS
TGx 1990-111F	Pdh1
TGx 1990-114F	
TGx 1990-140F	Pdh1
TGx 1990-142F	Pdh1
TGx 1991-12F	
TGx 1953-1F	
TGx 1984-22F	NS
TGx 1988-2F	Pdh1
TGx 1988-4F	
TGx 1988-5F	NS
TGx 1989-15F	
TGx 1990-29F	
Storm	
TGx 1989-4F	Pdh1
TGx 1989-13F	
TGx 1989-18F	Pdh1
TGx 1989-53F	Pdh1
TGx 1989-59F	
TGx 1989-70F	

TGx 1990-68F	Pdh1
TGx 1990-123F	Pdh1
TGx 1866-7F	
TGx 1904-6F	
TGx 1910-8F	
TGx 1910-13F	
TGx 1924-2F	
TGx 1949-7F	
TGx 1955-4F	
TGx 1965-5F	
TGx 1935-7F	
TGx 1945-1F	
TGx 1949-5F	
TGx 1951-3F	
TGx 1963-3F	
TGx 1975-2F	
TGx 1976-1F	Pdh1
TGx 1978-3F	
TGx 1987-18F	Pdh1
TGx 1987-19F	Pdh1
TGx 1987-32F	
TGx 1987-62F	
TGx 1987-117F	
TGx 1988-9F	
TGx 1988-14F	
TGx 1988-15F	Pdh1
TGx 1989-24F	NS
TGx 1989-29F	Pdh1
TGx 1989-45F	Pdh1
TGx 1989-72F	
TGx 1990-105F	Pdh1
TGx 1990-106F	
TGx 1990-121F	
TGx 1990-139F	
TGx 1989-12F	
TGx 1991-2F	
TGx 1909-3F	
TGx 1935-4F	
TGx 1949-10F	

TGx 1949-13F	
TGx 1951-1F	
TGx 1951-8F	
TGx 1989-60F	Pdh1
TGx 1990-15F	
TGx 1990-38F	
TGx 1990-56F	
TGx 1990-108F	Pdh1
TGx 1990-112F	
TGx 1990-120F	Pdh1
TGx 1990-130F	
TGx 1910-16F	NS
TGx 1989-27F	Pdh1
TGx 1990-45F	NS
TGx 1990-116F	NS
TGx 1990-129F	
TGx 1990-137F	
TGx 1991-15F	
TGx 1448-2E	NS
Santa	
TGx 1935-2F	
TGx 1956-1F	
TGx 1988-6F	Pdh1
TGx 1989-10F	Pdh1
TGx 1957-5F	NS
TGx 1965-7F	
TGx 1895-50F	NS
TGx 1911-8F	
TGx 1977-2F	
TGx 1986-2F	
TGx 1987-8F	
TGx 1987-14F	
TGx 1987-28F	Pdh1
TGx 1987-31F	
TGx 1987-34F	
TGx 1904-3F	
TGx 1910-14F	
TGx 1985-7F	
TGx 1990-109F	Pdh1

TGx 1905-2F	
Nasoko (MW1)	
TGx 1740-2F	
TGx 1703-3F	
TGx 1972-1F	
1935-3F	
1989-30F	Pdh1
TGx 1984-24F	
1988-25F	
1989-23F	
MAGOYE (MW2)	
1880-3F	
TGx 1990-5F	
TGx 1990-67F	
TGx 1990-110F	Pdh1
TGx 1990-135F	
TGx 1990-136F	
TGx 1990-141F	HET
TGx 1991-23F	
TGx 1991-24F	
Solitaire	
TGx 1951-9F	
TGx 1989-1F	
TGx 1989-21F	Pdh1
TGx 1838-5E	
TGx 1866-12F	
TGx 1871-12E	
TGx 1873-16E	
1987-6F	
1987-23F	
1987-40F	Pdh1
1987-64F	
1988-24F	
1988-28F	
1989-25F	Pdh1
1989-37F	Pdh1
TGx 1987-20F	
TGx 1987-35F	
TGx 1987-37F	

TGx 1987-65F	
TGx 1987-124F	Pdh1
TGx 1988-1F	
TGx 1988-26F	Pdh1
TGx 1988-27F	
TGx 1991-22F	
TGx 1991-26F	Pdh1
TGx 1019-2EB	
TGx 1440-1E	
TGx 1483-1E	
TGx 1485-1D	
TGx 1805-31F	
TGx 1830-20E	NS
TGx 1740-2F	NS
TGx 1903-7F	
TGx 1903-8F	
TGx 1904-5F	
TGx 1905-5F	
TGx 1932-1F	
TGx 1932-3F	
TGx 1933-2F	
TGx 1991-13F	
TGx 536-02D	
TGx 1805-8F	
TGx 1844-18E	
TGx 1895-33F	
TGx 1985-10F	
TGx 1987-3F	Pdh1
TGx 1987-5F	
TGx 1984-28F	
TGx 1985-3F	
TGx 1985-9F	
TGx 1985-12F	
TGx 1986-1F	
TGx 1988-19F	Pdh1
TGx 1989-9F	
TGx 1989-19F	Pdh1
TGx 1984-1F	
TGx 1984-5F	

TGx 1984-10F	
TGx 1984-11F	
TGx 1984-19F	
TGx 1984-23F	
TGx 1988-3F	Pdh1
TGx 1989-11F	Pdh1
TGx 1988-16F	
TGx 1988-17F	Pdh1
TGx 1988-22F	Pdh1
TGx 1989-2F	Pdh1
TGx 1989-6F	
TGx 1989-20F	Pdh1
TGx 1989-36F	
TGx 1989-56F	
TGx 1991-18F	
TGx 1991-20F	Pdh1
TGx 1835-10E	
TGx 1904-6F	
Ocepara- 4 (MW3)	
TGx 923-2E	
TGx 1834-1E	
TGx 1904-2F	
TGx 1954-4F	
TGx 1957-6F	
TGx 1984-17F	
TGx 1987-129F	
TGx 1989-14F	
TGx 1989-28F	Pdh1
TGx 1990-1F	
TGx 1990-28F	
TGx 1990-131F	
TGx 1991-11F	
TGx 1991-21F	Pdh1
UG5	Pdh1
Makwacha (MW4)	
Ocepara- 4 (MW3)	
TGx 1448-2E	
TGx 1903-3F	
TGx 1019-2EN	

TGx 1990-134F	
MAGOYE (MW2)	
TGx 1910-11F	
TGx 1954-1F	
TGx 1989-26F	Pdh1
TGx 1989-33F	
TGx 1990-107F	
TGx 1949-8F	
1985-8F	
TGx 1990-3F	
TGx 1990-13F	
TGx 1939-2F	
TGx 1950-4F	
1989-55F	Pdh1
1990-127F	Pdh1
TGx 1987-38F	Pdh1
TGx 1988-11F	
TGx 1988-12F	Pdh1
TGx 1988-23F	
TGx 1988-29F	Pdh1
TGx 1989-17F	
TGx 1989-42F	Pdh1
TGx 1989-54F	
TGx 1907-1F	
Lukanga	
MRI-Dina	
TGx 1910-2F	
Hernon 147	
Kaleya	
Safari	
Soprano	
1894-3F	
1908-3F	
1961-1F	
1971-1F	
1977-4F	
1985-2F	
1985-11F	
1986-3F	

Nasoko (MW1)	
SOY104	
TGx 1844-4E	
TGx 1908-8F	
TGx 1926-4F	
TGx 1927-1F	
TGx 1937-1F	NS
TGx 1983-37F	NS
TGx 1876-4E	
TGx 1880-3E	
TGx 1910-5F	
TGx 1910-6F	
TGx 1938-1F	
TGx 1939-1F	
TGx 1950-7F	
TGx 1951-4F	
1989-46F	
1989-50F	
1989-51F	Pdh1
1989-52F	
1989-63F	Pdh1
1989-58F	Pdh1
1989-68F	Pdh1
1989-69F	NS

Table 2: Pdh1 causative SNP allele and associated SNP allele from the 50kSNP marker set.

Line name	Pdh1 allele	50kSNPdata
FC029333	T	A
FC031697	T	A
FC033243	T	A
PI103088	T	A
PI123440	A	G
PI153231	T	A
PI154189	T	A
PI157421	A	A
PI159925	A	G
PI165563	A	A
PI165675	A	G
PI166105	T	A
PI171428	T	A
PI171451	A	G
PI179935	T	A
PI180501	T	A
PI189873	T	A
PI196166	A	G
PI209332	A	G
PI209333	A	G
PI209334	T	A
PI232992	A	G
PI240664	A	G
PI243541	A	G
PI253658B	T	A
PI253661B	T	A
PI261272C	A	G
PI266806C	T	A
PI274453	A	G
PI291294	T	A
PI291309D	T	A
PI291310C	T	A
PI297505	T	A
PI297520	T	A
PI317334A	A	G

PI317336	A	G
PI322692	T	A
PI323576	T	A
PI324924	T	A
PI339734	A	G
PI342434	A	G
PI342619A	T	A
PI360957	A	G
PI361066B	T	A
PI361070	T	A
PI361080	A	G
PI361087	T	A
PI361093	T	A
PI372403B	T	A
PI372418	T	A
PI374207	T	A
PI378658	T	A
PI378663	A	G
PI378680E	T	A
PI379618	A	G
PI391577	T	A
PI391583	T	A
PI398296	A	G
PI398881	T	A
PI398965	T	A
PI399043	A	G
PI404182	T	A
PI404187	T	A
PI404188A	T	A
PI407701	T	A
PI407708A	T	A
PI407716	T	A
PI407742	T	A
PI407801	A	G
PI407849	A	G
PI416751	T	A
PI416838	A	G
PI416890	A	G
PI416971	T	A
PI417215	A	G

PI417242	T	A
PI417345B	T	A
PI417381	T	A
PI417398	T	A
PI417479	A	G
PI417500	T	A
PI417529	A	G
PI417581	A	G
PI423926	T	A
PI423954	A	G
PI424038B	A	G
PI424078	A	G
PI424195A	T	A
PI424391	A	G
PI427136	T	A
PI430595	A	G
PI436684	T	A
PI437110A	T	A
PI437112A	T	A
PI437127A	T	A
PI437160	T	A
PI437165A	T	A
PI437169B	T	A
PI437240	T	A
PI437321	T	A
PI437376A	T	A
PI437485	T	A
PI437500A	T	A
PI437505	T	A
PI437653	T	A
PI437654	T	A
PI437662	A	A
PI437679	T	A
PI437685D	T	A
PI437695A	T	A
PI437776	A	G
PI437788A	T	A
PI437793	T	A
PI437814A	T	A
PI437838	T	A

PI437944	T	A
PI437991B	T	A
PI438019B	T	A
PI438083	T	A
PI438112B	T	A
PI438230A	T	A
PI438239B	T	A
PI438309	T	A
PI438323	T	A
PI438335	A	G
PI438336	T	A
PI438347	T	A
PI438496B	T	A
PI438496C	T	A
PI438498	T	A
PI438500	T	A
PI445824A	T	A
PI458505	T	A
PI458510	T	A
PI464896	T	A
PI464912	T	A
PI464923	T	A
PI467343	T	A
PI467347	T	A
PI468408B	T	A
PI468908	A	G
PI475820	T	A
PI476352B	T	A
PI479735	T	A
PI490766	T	A
PI495020	T	A
PI497953	A	G
PI497964A	T	A
PI497967	T	A
PI504288	A	G
PI506862	A	G
PI506933	A	G
PI506942	T	A
PI507017	A	G
PI507088	A	G

PI507180	A	G
PI507293B	A	G
PI507355	A	G
PI507458	T	A
PI507467	A	G
PI507471	T	A
PI507480	T	A
PI507681B	T	A
PI508083	T	A
PI513382	T	A
PI514671	T	A
PI515961	T	A
PI518668	T	A
PI518727	A	G
PI518750	T	A
PI518751	T	A
PI532463B	T	A
PI533602	T	A
PI533655	T	A
PI536635	T	A
PI538386A	T	A
PI540552	T	A
PI542403	T	A
PI542972	T	A
PI546044	T	A
PI547409	T	A
PI547459	T	A
PI547460	T	A
PI547488	T	A
PI547562	T	A
PI547680	T	A
PI547686	T	A
PI547690	T	A
PI547716	T	A
PI547779	T	A
PI547890	T	A
PI548162	T	A
PI548169	T	A
PI548171	T	A
PI548178	A	G

PI548182	T	A
PI548190	T	A
PI548193	T	A
PI548198	T	A
PI548200	T	A
PI548256	A	G
PI548298	T	A
PI548311	T	A
PI548313	A	G
PI548316	T	A
PI548325	T	G
PI548336	T	A
PI548342	A	G
PI548348	T	A
PI548356	A	G
PI548359	T	A
PI548360	T	A
PI548364	T	A
PI548379	T	A
PI548382	A	G
PI548383	T	A
PI548391	T	A
PI548400	T	A
PI548402	T	A
PI548406	T	A
PI548411	T	A
PI548417	T	A
PI548427	T	A
PI548445	A	G
PI548447	A	G
PI548452	T	A
PI548456	A	G
PI548473	T	A
PI548474	T	A
PI548477	T	A
PI548479	A	G
PI548488	T	A
PI548490	T	A
PI548512	T	A
PI548520	T	A

PI548521	T	A
PI548524	T	A
PI548540	T	A
PI548561	T	A
PI548565	T	A
PI548571	T	A
PI548572	T	A
PI548573	T	A
PI548582	T	A
PI548593	A	G
PI548603	T	A
PI548604	T	A
PI548619	T	A
PI548622	T	A
PI548631	T	A
PI548633	T	A
PI548634	T	A
PI548643	T	A
PI548644	T	A
PI548657	T	A
PI548696	T	A
PI548978	T	A
PI548985	T	A
PI549017	T	A
PI549018	A	A
PI549021A	A	G
PI549026	T	A
PI549028	A	G
PI549040	T	A
PI549041A	T	A
PI553047	T	A
PI556511	T	A
PI559932	T	A
PI561318A	T	A
PI561370	T	A
PI561371	T	A
PI561387	A	G
PI561389B	T	A
PI561701	T	A
PI567071A	A	G

PI567074B	T	A
PI567171	A	G
PI567173	T	A
PI567189A	A	G
PI567225	T	A
PI567226	T	A
PI567231	A	A
PI567238	A	G
PI567258	A	G
PI567262A	A	G
PI567298	T	A
PI567307	T	A
PI567343	T	A
PI567346	T	A
PI567352A	T	A
PI567353	T	A
PI567361	T	A
PI567364	T	A
PI567383	T	A
PI567395	A	G
PI567407	T	A
PI567408	A	G
PI567410B	T	A
PI567415A	T	A
PI567416	T	A
PI567418A	T	A
PI567426	T	A
PI567428	T	A
PI567435B	T	A
PI567439	T	A
PI567488A	T	A
PI567489A	T	A
PI567503	T	A
PI567525	T	A
PI567532	A	G
PI567548	T	A
PI567558	A	G
PI567576	T	A
PI567604A	T	A
PI567675	T	A

PI567685	T	A
PI567698A	T	A
PI567726	T	A
PI567746	T	A
PI567780B	T	A
PI567782	T	A
PI567788	T	A
PI574477	T	A
PI578309	T	A
PI578375B	A	G
PI578412	T	A
PI578457A	A	G
PI578493	A	A
PI578495	T	A
PI578499A	T	G
PI578503	T	A
PI578504	A	G
PI587552	A	A
PI587588A	T	A
PI587588B	T	A
PI587666	T	A
PI587712B	T	A
PI587752	T	A
PI587804	T	A
PI587811A	A	G
PI587848	A	G
PI588053A	A	G
PI591431	T	A
PI591432	T	A
PI591433	T	A
PI591435	T	A
PI591495	T	A
PI591511	T	A
PI591541	T	A
PI592523	T	A
PI592937	T	A
PI592940	A	G
PI592952	T	A
PI592954	T	A
PI592960	T	A

PI593258	T	A
PI593953	T	A
PI594170B	A	G
PI594301	A	G
PI594307	A	G
PI594451	A	G
PI594456A	A	G
PI594579	T	A
PI594629	A	G
PI594777	A	G
PI594788	A	G
PI594880	T	A
PI594922	T	A
PI597464	T	A
PI597471A	A	G
PI597476	T	A
PI597478B	T	A
PI598124	T	A
PI598358	T	A
PI602502B	A	G
PI602991	T	A
PI602993	T	A
PI603162	T	A
PI603290	T	A
PI603318	T	A
PI603336	T	A
PI603345	T	A
PI603357	T	A
PI603384	T	A
PI603389	T	A
PI603397	A	G
PI603399	T	A
PI603420	T	A
PI603424A	T	A
PI603426G	T	A
PI603442	T	A
PI603463	T	A
PI603488	T	A
PI603492	A	G
PI603494	A	G

PI603495B	T	A
PI603497	A	G
PI603526	T	A
PI603549	T	A
PI603555	T	A
PI603556	T	A
PI603559	T	A
PI603675	A	G
PI603698J	A	G
PI603722	A	G
PI603756	A	G
PI605765B	T	A
PI606374	A	G
PI612730	T	A
PI612754	A	G
PI615553	T	A
PI628812	T	A
PI628913	T	A
PI628963	T	A
PI631123	T	A
PI632418	T	A
PI632650	A	G
PI633730	T	A
PI633731	T	A
PI634883	T	A
PI639283	T	A
PI639285	T	A
PI639528B	T	A
PI639543	T	A
PI639550E	T	A
PI639559B	T	A
PI639570	T	A
PI639740	T	A
PI643146	T	A
PI054591	T	A
PI054608_1	A	G
PI054614	T	A
PI054615_1	T	A
PI058955	T	A
PI062203	T	A

PI068521_1	T	A
PI068604_1	T	A
PI068732_1	T	A
PI070080	T	A
PI070466_3	T	A
PI071465	T	A
PI080822	T	A
PI080837	A	G
PI081041	A	G
PI081785	A	G
PI083881	A	G
PI083942	A	G
PI083945-3	T	A
PI084631	T	A
PI084637	T	A
PI084656	T	A
PI084946_2	A	G
PI084973	T	A
PI084987	A	G
PI084987A	A	G
PI086024	A	G
PI086904	T	A
PI086972_2	T	A
PI087620	T	A
PI088313	T	A
PI088468	T	A
PI088479	T	A
PI088788	T	A
PI089005_5	T	A
PI089138	T	A
PI089775	T	A
PI090479P	T	A
PI090763	T	A
PI091100_3	T	A
PI091159_4	T	A
PI091160	T	A
PI092651	T	A
PI094159_3	T	A
PI095860	A	A

APPENDIX 2

Supplementary Data for Chapter Three

FIGURES

TABLES

Table 1: Mean data for days to flower, days to maturity, and genotype for each RIL.

Outliers not removed

Population	Line number	dtf means	dtm means	E1	E2	E3	Dt1	J	Pdh1
Jake x 159925	5001	45.3	108.4	E1	E2	E3	R166W	j-1	Pdh1
Jake x 159925	5002	46.8	110.7	E1	E2	E3	R166W	j-1	Pdh1
Jake x 159925	5003	48.0	109.0	E1	E2	E3	R166W	j-1	pdh1
Jake x 159925	5004	50.6	114.3	E1	E2	E3	R166W	j-1	Pdh1
Jake x 159925	5005	44.4	109.4	E1	E2	E3	R166W	j-1	Pdh1
Jake x 159925	5006	43.0	109.7	E1	E2	E3	R166W	j-1	pdh1
Jake x 159925	5007	40.0	108.5	E1	E2	E3	R166W	J	Pdh1
Jake x 159925	5008	45.0	108.7	E1	E2	E3	R166W	j-1	Pdh1
Jake x 159925	5009	47.4	112.5	E1	E2	E3	R166W	j-1	Pdh1
Jake x 159925	5011	43.9	108.8	E1	E2	E3	R166W	j-1	Pdh1
Jake x 159925	5012	45.6	111.8	E1	E2	E3	R166W	j-1	pdh1
Jake x 159925	5013	43.9	108.1	E1	E2	E3	R166W	j-1	pdh1
Jake x 159925	5014	46.7	113.0	E1	E2	E3	R166W	j-1	NA
Jake x 159925	5015	48.5	113.8	E1	E2	E3	R166W	j-1	pdh1
Jake x 159925	5016	45.4	107.9	E1	E2	E3	R166W	j-1	pdh1
Jake x 159925	5017	47.4	110.4	E1	E2	E3	R166W	NA	Pdh1
Jake x 159925	5018	45.3	112.9	E1	E2	E3	R166W	j-1	Pdh1
Jake x 159925	5019	44.8	113.0	E1	E2	E3	R166W	j-1	Pdh1
Jake x 159925	5020	44.4	107.6	E1	E2	E3	R166W	j-1	pdh1
Jake x 159925	5021	44.9	106.8	E1	E2	E3	R166W	j-1	pdh1
Jake x Paranagoiana	5102	47.9	117.2	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5103	46.3	116.1	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5104	46.8	118.7	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5105	48.6	119.6	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5106	49.8	119.7	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5107	44.8	112.0	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5108	49.0	114.4	E1	E2	E3	R166W	NA	pdh1
Jake x Paranagoiana	5109	49.1	114.9	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5110	47.2	115.2	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5111	48.0	115.1	E1	E2	E3	R166W	NA	pdh1
Jake x Paranagoiana	5112	48.7	115.5	E1	E2	E3	R166W	j-x	pdh1

Jake x Paranagoiana	5113	48.6	122.0	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5114	46.2	113.5	E1	E2	E3	R166W	NA	pdh1
Jake x Paranagoiana	5115	47.6	112.1	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5116	40.6	117.7	E1	E2	E3	R166W	NA	pdh1
Jake x Paranagoiana	5117	44.6	110.3	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5118	47.0	115.0	E1	E2	E3	R166W	j-x	pdh1
Jake x Paranagoiana	5119	48.8	117.6	E1	E2	E3	R166W	j-x	pdh1
X97 x 159925	5201	39.8	100.0	e1-as	E2	E3	R166W	j-1	pdh1
X97 x 159925	5202	39.1	107.0	e1-as	E2	E3	R166W	j-1	Pdh1
X97 x 159925	5203	43.5	104.0	E1	E2	E3	Dt1	j-1	pdh1
X97 x 159925	5204	31.3	93.3	het	E2	E3	Dt1	J	pdh1
X97 x 159925	5205	37.5	106.3	e1-as	E2	E3	Dt1	j-1	pdh1
X97 x 159925	5206	37.1	105.0	e1-as	E2	E3	R166W	j-1	Pdh1
X97 x 159925	5207	36.8	107.3	e1-as	E2	E3	NA	j-1	H
X97 x 159925	5208	42.2	110.1	e1-as	E2	E3	Dt1	j-1	Pdh1
X97 x 159925	5209	38.5	111.5	E1	E2	E3	Dt1	j-1	pdh1
X97 x 159925	5210	32.3	91.8	E1	E2	E3	R166W	J	pdh1
X97 x 159925	5211	39.7	107.9	NA	E2	E3	het	j-1	pdh1
X97 x 159925	5212	40.3	106.9	e1-as	E2	E3	het	j-1	pdh1
X97 x 159925	5213	36.7	106.9	e1-as	E2	E3	R166W	j-1	Pdh1
X97 x 159925	5214	36.7	105.9	e1-as	E2	E3	R166W	j-1	Pdh1
X97 x 159925	5215	43.5	108.4	E1	E2	E3	R166W	j-1	Pdh1
X97 x 159925	5216	44.1	108.4	E1	E2	E3	het	j-1	Pdh1
X97 x 159925	5218	31.6	93.3	e1-as	E2	E3	R166W	J	Pdh1
X97 x 159925	5219	34.9	98.2	het	E2	E3	NA	J	Pdh1
X97 x 159925	5220	47.7	110.6	E1	E2	E3	R166W	j-1	Pdh1
X97 x 159925	5221	42.7	110.4	E1	E2	E3	R166W	j-1	Pdh1
X97 x 159925	5222	45.9	111.7	E1	E2	E3	Dt1	j-1	Pdh1
X97 x 159925	5223	32.0	105.5	E1	E2	E3	Dt1	J	pdh1
X97 x 159925	5224	NA	NA	E1	E2	E3	NA	J	Pdh1
X97 x 159925	5225	37.3	107.2	e1-as	E2	E3	Dt1	j-1	Pdh1
X97 x 159925	5226	38.9	114.6	E1	E2	E3	Dt1	j-1	pdh1
X97 x 159925	5227	38.3	114.3	E1	E2	E3	NA	j-1	Pdh1
X97 x 159925	5228	NA	NA	NA	E2	E3	NA	NA	NA
X97 x 159925	5229	39.6	112.7	e1-as	E2	E3	R166W	j-1	pdh1
X97 x 159925	5230	38.1	108.4	e1-as	E2	E3	R166W	het	Pdh1
X97 x 159925	5231	41.0	111.6	e1-as	E2	E3	Dt1	NA	pdh1

X97 x 159925	5232	48.8	112.8	NA	E2	E3	R166W	j-1	pdh1
X97 x 159925	5233	37.8	99.3	E1	E2	E3	Dt1	J	pdh1
X97 x 159925	5234	42.6	114.2	E1	E2	E3	Dt1	j-1	Pdh1
X97 x 159925	5235	41.3	107.9	E1	E2	E3	Dt1	j-1	pdh1
X97 x 159925	5236	50.0	113.7	E1	E2	E3	Dt1	j-1	Pdh1
X97 x 159925	5237	NA	NA	E1	E2	E3	Dt1	j-1	Pdh1
X97 x 159925	5238	32.9	99.5	E1	E2	E3	NA	J	pdh1
X97 x 159925	5239	41.5	113.5	e1-as	E2	E3	Dt1	j-1	pdh1
X97 x 159925	5240	43.8	116.4	E1	E2	E3	Dt1	het	Pdh1
X97 x 159925	5241	38.6	110.0	NA	E2	E3	NA	NA	NA
X97 x 159925	5242	41.9	113.4	e1-as	E2	E3	R166W	j-1	Pdh1
X97 x 159925	5243	44.0	115.3	e1-as	E2	E3	Dt1	het	Pdh1
X97 x 159925	5244	46.6	113.1	E1	E2	E3	NA	j-1	Pdh1
X97 x 159925	5245	48.7	109.9	NA	E2	E3	NA	NA	pdh1
X97 x 159925	5246	45.8	110.6	E1	E2	E3	R166W	j-1	pdh1
X97 x 159925	5247	43.8	112.3	E1	E2	E3	Dt1	j-1	pdh1
X97 x 159925	5249	43.6	109.8	E1	E2	E3	R166W	j-1	pdh1
KB 13-34 (534545 x Canadian X18)	5301	32.4	93.3	E1	e2	E3	Dt1	J	pdh1
KB 13-34 (534545 x Canadian X18)	5302	37.4	109.5	e1-as	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5303	38.2	105.1	e1-as	E2	e3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5304	37.8	107.1	e1-as	Het	e3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5305	38.3	105.5	e1-as	Het	e3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5306	40.8	112.3	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5307	38.5	105.3	e1-as	E2	e3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5308	40.8	112.7	e1-as	E2	E3	Dt1	J	pdh1
KB 13-34 (534545 x Canadian X18)	5309	39.8	113.0	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5310	43.8	116.5	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5311	50.6	122.5	E1	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5313	44.8	112.3	E1	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5314	41.0	111.0	E1	e2	e3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5315	36.6	105.5	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5316	36.2	106.3	e1-as	E2	NA	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5317	45.7	112.8	E1	e2	e3	Dt1	j-x	pdh1

KB 13-34 (534545 x Canadian X18)	5318	36.1	100.9	e1-as	e2	e3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5319	50.1	118.2	E1	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5320	46.7	116.2	Het	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5321	44.4	110.3	e1-as	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5322	38.3	109.6	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5323	46.3	116.1	E1	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5324	47.3	117.9	E1	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5325	41.8	118.1	e1-as	E2	NA	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5326	37.7	108.0	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5327	45.3	114.2	E1	e2	NA	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5328	36.2	105.5	e1-as	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5329	39.2	108.8	e1-as	E2	E3	Dt1	NA	pdh1
KB 13-34 (534545 x Canadian X18)	5330	40.7	106.0	het	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5331	46.2	117.2	E1	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5332	44.9	115.4	E1	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5333	38.0	103.9	e1-as	E2	e3	Dt1	J	pdh1
KB 13-34 (534545 x Canadian X18)	5334	30.4	102.2	E1	het	e3	Dt1	J	pdh1
KB 13-34 (534545 x Canadian X18)	5335	49.7	119.9	E1	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5336	46.8	118.1	E1	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5337	34.0	99.6	E1	E2	E3	Dt1	J	pdh1
KB 13-34 (534545 x Canadian X18)	5338	46.1	111.7	E1	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5339	33.5	100.6	e1-as	e2	e3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5340	36.6	102.8	e1-as	e2	E3	Dt1	J	pdh1
KB 13-34 (534545 x Canadian X18)	5341	42.0	119.3	het	e2	NA	Dt1	J	pdh1
KB 13-34 (534545 x Canadian X18)	5342	36.1	108.0	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5343	38.0	104.3	e1-as	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5344	46.9	115.9	het	het	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5345	33.1	102.0	e1-as	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5346	41.2	111.3	e1-as	e2	E3	Dt1	j-x	pdh1
KB 13-34 (534545 x Canadian X18)	5347	35.3	107.0	e1-as	e2	E3	Dt1	j-x	pdh1

KB 13-34 (534545 x Canadian X18)	5348	42.1	112.7	E1	e2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5401	31.3	99.3	het	e2	e3	Dt1	J	pdh1
KB 13-33 (X97 x Canadian X18)	5402	NA	107.9	e1-as	E2	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5403	38.7	109.5	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5404	39.6	107.8	E1	E2	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5405	40.3	119.6	E1	E2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5406	43.9	109.0	E1	E2	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5407	46.5	118.0	E1	E2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5408	44.7	120.2	E1	E2	NA	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5409	41.5	109.2	E1	e2	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5410	45.1	114.0	E1	e2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5411	38.7	101.5	E1	E2	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5412	36.6	103.6	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5413	37.3	104.7	E1	E2	e3	Dt1	NA	pdh1
KB 13-33 (X97 x Canadian X18)	5414	34.6	107.3	NA	e2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5415	40.7	115.5	het	E2	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5416	38.0	105.7	E1	e2	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5417	42.4	116.3	E1	E2	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5418	38.4	112.0	E1	e2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5419	43.1	107.3	E1	E2	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5420	37.4	108.5	e1-as	e2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5421	42.8	110.9	E1	e2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5422	33.9	98.5	e1-as	E2	e3	Dt1	J	pdh1
KB 13-33 (X97 x Canadian X18)	5423	43.8	108.2	E1	E2	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5424	44.9	112.0	E1	het	e3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5425	32.8	101.6	e1-as	e2	het	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5426	39.4	113.1	E1	E2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5427	41.6	110.2	E1	e2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5428	36.1	108.3	e1-as	e2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5429	44.1	111.6	E1	e2	E3	Dt1	j-x	pdh1

KB 13-33 (X97 x Canadian X18)	5430	37.5	106.8	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5431	45.4	118.2	E1	E2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5432	30.3	94.0	e1-as	E2	E3	Dt1	J	pdh1
KB 13-33 (X97 x Canadian X18)	5433	36.7	109.3	e1-as	het	E3	Dt1	J	pdh1
KB 13-33 (X97 x Canadian X18)	5434	33.6	100.9	E1	E2	E3	Dt1	J	pdh1
KB 13-33 (X97 x Canadian X18)	5435	44.0	113.2	e1-as	E2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5436	44.8	118.0	E1	E2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5437	44.9	112.0	E1	E2	E3	Dt1	J	pdh1
KB 13-33 (X97 x Canadian X18)	5438	35.3	110.8	E1	E2	E3	Dt1	j-x	pdh1
KB 13-33 (X97 x Canadian X18)	5440	44.1	115.3	E1	E2	E3	Dt1	j-x	pdh1
X97 x Jenguma	5501	40.9	112.0	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5502	35.5	112.3	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5503	33.8	111.0	NA	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5504	32.3	99.7	e1-as	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5505	32.9	102.1	NA	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5506	37.0	102.1	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5507	33.7	108.9	e1-as	E2	E3	NA	NA	pdh1
X97 x Jenguma	5508	36.3	107.4	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5509	36.2	107.0	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5510	38.3	107.4	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5511	32.1	99.9	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5512	31.5	92.3	e1-as	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5513	32.9	106.3	e1-as	E2	E3	NA	NA	pdh1
X97 x Jenguma	5514	40.5	108.7	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5515	33.3	110.1	NA	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5516	40.0	106.0	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5517	32.0	96.6	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5518	31.0	97.3	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5519	35.4	101.7	E1	E2	E3	NA	NA	pdh1
X97 x Jenguma	5520	37.9	103.1	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5521	35.3	105.4	NA	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5522	33.9	100.8	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5523	32.4	102.0	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5524	43.8	114.9	NA	E2	E3	Dt1	NA	pdh1

X97 x Jenguma	5525	38.6	109.0	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5526	41.1	107.7	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5527	36.8	102.5	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5528	37.7	107.3	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5529	31.7	99.8	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5530	37.6	109.8	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5531	32.7	104.8	NA	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5532	36.6	104.8	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5533	41.9	111.5	E1	E2	E3	NA	NA	pdh1
X97 x Jenguma	5534	40.4	108.4	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5535	33.1	105.1	e1-as	E2	E3	NA	NA	pdh1
X97 x Jenguma	5536	44.5	117.2	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5537	38.3	104.4	NA	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5538	43.8	113.0	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5539	38.2	111.6	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5540	39.5	110.0	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5541	37.7	104.6	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5542	41.8	108.8	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5543	36.3	104.4	E1	E2	E3	NA	NA	pdh1
X97 x Jenguma	5544	37.9	106.9	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5545	35.4	102.9	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5546	38.0	106.8	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5547	34.9	104.4	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5548	31.7	103.1	NA	E2	E3	NA	NA	pdh1
X97 x Jenguma	5549	32.9	100.1	NA	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5550	32.3	96.4	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5551	36.5	95.0	NA	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5552	32.0	99.9	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5553	35.3	104.8	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5554	33.5	103.0	e1-as	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5555	33.6	104.9	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5556	38.3	108.0	E1	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5557	35.6	106.8	e1-as	E2	E3	P113L	NA	pdh1
X97 x Jenguma	5559	36.1	107.0	E1	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5560	33.1	101.6	e1-as	E2	E3	Dt1	NA	pdh1
X97 x Jenguma	5561	31.3	95.7	NA	E2	E3	Dt1	NA	pdh1

Table 2. ANOVAs for days to flower (dtf), days to maturity (dtm), and yield (yld) for the RIL population Jake x PI 159925.

The SAS System

The GLM Procedure

Class Level Information		
Class	Levels	Values
g	19	5001 5002 5003 5004 5005 5006 5008 5009 5011 5012 5013 5014 5015 5016 5017 5018 5019 5020 5021
l	7	1 2 4 6 7 8 9
r	2	1 2
dtf	13	40 41 42 43 44 45 46 47 48 49 50 52 53
dtm	24	99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 123
yld	169	0.6016 0.7296 0.8832 0.9024 0.928 1.056 1.3184 1.3248 1.3312 1.4272 1.44 1.5552 1.7536 1.7664 1.8752 1.9136 1.9904 2.0992 2.112 2.1376 2.208 2.2144 2.2912 2.304 2.4448 2.4704 2.5984 2.7008 2.7648 2.848 2.8544 2.8736 2.9824 2.9952 3 3.04 3.104 3.3216 3.4368 3.456 3.6 3.7952 3.8144 3.9424 4.0384 4.0512 4.2624 4.288 4.4032 4.9408 4.9728 5.2928 5.4144 5.4528 5.5 5.5168 5.5552 5.7 5.7536 5.7856 6.0608 6.1184 6.2016 6.24 6.4 6.6816 6.7 6.7328 6.8608 7.2256 7.2896 7.3536 7.3792 7.6736 7.8016 7.9552 7.9616 8.0576 8.192 8.4096 8.4352 8.7232 8.8 9.2 9.2288 9.2416 9.2672 9.5 9.6192 9.6768 9.7344 9.8432 10.4192 10.6304 10.8544 10.9 11.584 11.7632 12.0384 12.2304 13.5808 14.6 15.2 15.7888 16.3264 16.5 16.6144 16.8704 17.1 17.4144 17.5 18.2144 18.2656 18.9696 19.2448 19.7 21 21.0176 21.8 22.9 25.0176 25.312 29.5 30.1 31.5 33.1 33.4 34 34.6 36 38.9 39.6 41.1 42.1 45.8 46.3 46.4 47 50.7 51.1 52.9 54.9 56.5 56.9 57.3 58.7 60.5 62.9 63.6 63.9 66.8 67 68.1 68.9 69 72.1 77.4 86.2 86.5 89.5 90.7 94.2 96.3 106.5 115.2 117.6 145.1 168 224.2

Number of Observations Read	272
Number of Observations Used	174

The SAS System

The GLM Procedure

Dependent Variable: dtf

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	97	1510.810225	15.575363	7.46	<.0001
Error	76	158.684028	2.087948		
Corrected Total	173	1669.494253			

R-Square	Coeff Var	Root MSE	dtf Mean
0.904951	3.176167	1.444973	45.49425

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	18	484.9299672	26.9405537	12.90	<.0001
l	6	689.3502994	114.8917166	55.03	<.0001
r(l)	7	18.3070816	2.6152974	1.25	0.2853
g*l	66	318.2228770	4.8215587	2.31	0.0002

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	18	514.1427211	28.5634845	13.68	<.0001
l	6	654.4456926	109.0742821	52.24	<.0001
r(l)	7	14.3159722	2.0451389	0.98	0.4525
g*l	66	318.2228770	4.8215587	2.31	0.0002

The SAS System

The GLM Procedure

Dependent Variable: dtm

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	97	3548.267967	36.580082	4.37	<.0001
Error	76	635.938930	8.367617		
Corrected Total	173	4184.206897			

R-Square	Coeff Var	Root MSE	dtm Mean
0.848014	2.615772	2.892683	110.5862

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	18	1082.071182	60.115066	7.18	<.0001
l	6	1462.745575	243.790929	29.14	<.0001
r(l)	7	54.148070	7.735439	0.92	0.4926
g*l	66	949.303140	14.383381	1.72	0.0115

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	18	1024.492588	56.916255	6.80	<.0001
l	6	1482.390719	247.065120	29.53	<.0001
r(l)	7	55.061070	7.865867	0.94	0.4811
g*l	66	949.303140	14.383381	1.72	0.0115

The SAS System

The GLM Procedure

Dependent Variable: yld

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	97	168796.7376	1740.1726	4.70	<.0001
Error	76	28141.6301	370.2846		
Corrected Total	173	196938.3677			

R-Square	Coeff Var	Root MSE	yld Mean
0.857104	79.10419	19.24278	24.32587

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	18	56814.23570	3156.34643	8.52	<.0001
l	6	65983.98690	10997.33115	29.70	<.0001
r(l)	7	546.46240	78.06606	0.21	0.9820
g*l	66	45452.05261	688.66746	1.86	0.0046

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	18	25419.54081	1412.19671	3.81	<.0001
l	6	61888.30658	10314.71776	27.86	<.0001
r(l)	7	513.72252	73.38893	0.20	0.9850
g*l	66	45452.05261	688.66746	1.86	0.0046

Table 3. ANOVAs for days to flower (dtf), days to maturity (dtm), and yield (yld) for the RIL population Jake x Paranagoiana.

The SAS System

The GLM Procedure

Class Level Information		
Class	Levels	Values
g	14	5102 5103 5104 5105 5106 5107 5108 5109 5111 5112 5114 5115 5117 5119
l	7	1 2 4 6 7 8 9
r	2	1 2
dtf	13	40 42 43 44 45 46 47 48 49 50 51 52 53
dtm	23	101 102 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 126
yld	138	0.7 1.0432 1.2352 1.5616 1.6 2.4 2.7 2.752 3.2192 3.3664 3.584 4.3072 4.3584 4.6 4.6272 4.6528 4.672 5.1 5.3 5.472 6.3552 6.464 6.7072 7.2 7.3 7.4112 7.7056 7.9424 8.1984 8.4 8.4224 8.4928 8.6 9.1712 9.4208 9.6 10.0736 10.144 10.4 10.4768 10.688 10.8672 11.2064 11.2192 11.3984 11.4 11.9 12.5184 12.6 13.792 14.0672 14.0928 14.2 14.5216 14.9504 15.3472 15.4816 15.5328 16.4352 16.576 17.024 17.6896 18.5536 19.4688 19.6096 19.8464 20 20.2 21.0816 21.152 21.312 21.7792 21.8048 22.1 23 23.104 23.5264 23.7 23.7952 24.0896 24.4 24.7 24.704 25.088 25.4 25.6 25.6448 25.8 26.7072 26.8 27.1 27.1168 27.5 28.1984 28.4096 28.6912 28.7808 29.0432 29.6 30.6 32.8 33.4 34.7 35.2 37.9 38.5 39 40.2432 40.5 42.1 44.4 44.6 48.6 49.1264 49.4 50.7 52.2 53.2 53.9 55.8 56 57.3 60.3 66.8 69 70.2 72.2 73.2 74.3 78.3 90.6 126.1 128.6 133.5 155.4 170.3 171.5 266

Number of Observations Read	292
Number of Observations Used	141

The SAS System

The GLM Procedure

Dependent Variable: dtf

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	89	1078.021530	12.112601	7.30	<.0001
Error	51	84.630952	1.659430		
Corrected Total	140	1162.652482			

R-Square	Coeff Var	Root MSE	dtf Mean
0.927209	2.757471	1.288189	46.71631

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	13	256.9644953	19.7664996	11.91	<.0001
l	6	563.3732867	93.8955478	56.58	<.0001
r(l)	7	23.2906585	3.3272369	2.01	0.0725
g*l	63	234.3930894	3.7205252	2.24	0.0017

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	13	321.4031942	24.7233226	14.90	<.0001
l	6	547.3311147	91.2218524	54.97	<.0001
r(l)	7	15.8690476	2.2670068	1.37	0.2398
g*l	63	234.3930894	3.7205252	2.24	0.0017

The SAS System

The GLM Procedure

Dependent Variable: dtm

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	89	2499.753774	28.087121	1.97	0.0048
Error	51	728.884524	14.291853		
Corrected Total	140	3228.638298			

R-Square	Coeff Var	Root MSE	dtm Mean
0.774244	3.296502	3.780457	114.6809

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	13	845.5392636	65.0414818	4.55	<.0001
l	6	661.4891779	110.2481963	7.71	<.0001
r(l)	7	142.4346965	20.3478138	1.42	0.2164
g*l	63	850.2906360	13.4966768	0.94	0.5884

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	13	827.0188191	63.6168322	4.45	<.0001
l	6	672.0301950	112.0050325	7.84	<.0001
r(l)	7	186.1154762	26.5879252	1.86	0.0958
g*l	63	850.2906360	13.4966768	0.94	0.5884

The SAS System

The GLM Procedure

Dependent Variable: yld

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	89	172706.0361	1940.5173	4.35	<.0001
Error	51	22767.6717	446.4249		
Corrected Total	140	195473.7078			

R-Square	Coeff Var	Root MSE	yld Mean
0.883526	68.95278	21.12877	30.64238

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	13	35878.17479	2759.85960	6.18	<.0001
l	6	40513.92570	6752.32095	15.13	<.0001
r(l)	7	2657.53780	379.64826	0.85	0.5515
g*l	63	93656.39780	1486.60949	3.33	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	13	33469.66334	2574.58949	5.77	<.0001
l	6	31281.45735	5213.57623	11.68	<.0001
r(l)	7	2750.24879	392.89268	0.88	0.5288
g*l	63	93656.39780	1486.60949	3.33	<.0001

Table 4. ANOVAs for days to flower (dtf), days to maturity (dtm), and yield (yld) for the RIL population X97 x PI 159925.

The SAS System

The GLM Procedure

Class Level Information		
Class	Levels	Values
g	43	5201 5202 5203 5204 5205 5206 5207 5208 5209 5210 5211 5212 5213 5214 5215 5216 5218 5219 5220 5221 5222 5223 5225 5226 5227 5229 5230 5231 5232 5233 5234 5235 5236 5237 5238 5239 5240 5242 5243 5244 5246 5247 5249
l	8	1 2 4 5 6 7 8 9
r	2	1 2
dtf	21	28 30 31 33 34 35 37 38 40 41 42 43 44 45 46 47 48 50 51 52 53
dtm	35	82 83 84 86 87 88 90 91 92 94 95 96 98 99 100 101 102 103 104 105 106 107 108 110 111 112 113 115 116 117 118 119 120 122 123
yld	232	0.2112 0.3264 0.3648 0.5312 0.5376 0.5568 0.5824 0.6848 0.6912 0.7808 0.8064 0.8256 0.8704 0.8832 0.9472 0.9664 0.992 0.9984 1.0176 1.024 1.056 1.0624 1.0816 1.088 1.1328 1.1648 1.184 1.1904 1.216 1.2288 1.2672 1.312 1.3376 1.3952 1.4016 1.44 1.4912 1.504 1.5232 1.5296 1.536 1.5424 1.5616 1.5808 1.5936 1.6192 1.6384 1.6512 1.6576 1.6768 1.7088 1.7216 1.7472 1.76 1.8176 1.8304 1.856 1.8624 1.92 1.9328 1.9392 1.9584 1.9648 1.9904 2.016 2.0288 2.0992 2.24 2.2656 2.272 2.3168 2.3232 2.3296 2.3552 2.3616 2.368 2.3808 2.3872 2.3936 2.4064 2.432 2.4704 2.4768 2.5472 2.5536 2.56 2.5664 2.592 2.5984 2.656 2.6752 2.688 2.7264 2.7456 2.7904 2.8416 2.848 2.8544 2.9056 2.9248 2.9632 2.9824 2.9888 3.04 3.072 3.0848 3.104 3.1168 3.1424 3.1488 3.2128 3.2256 3.3152 3.3216 3.328 3.3344 3.3408 3.3472 3.3728 3.4368 3.4432 3.4688 3.5648 3.6224 3.6288 3.6672 3.808 3.8336 3.904 3.9168 3.9296 3.9872 4.0384 4.0448 4.1088 4.1152 4.1856 4.1984 4.2432 4.2496 4.2688 4.3008 4.4992 4.5 4.544 4.5824 4.672 4.7552 4.7744 4.8 4.8256 4.9536 5.0112 5.1008 5.2032 5.2864 5.3312 5.376 5.4 5.4592 5.472 5.5936 5.8496 5.9392 5.9456 6.112 6.1824 6.24 6.2464 6.432 6.4384 6.5856 6.6624 6.6752 6.752 6.7584 6.912 6.9248 7.0912 7.1 7.1808 7.9168 7.9808 8.5632 8.6 8.7 8.7552 8.8256 8.9088 9.0752 9.1264 9.1392 9.6 10.1632 10.4512 10.4832 11.0016 11.072 11.1872 11.6032 11.9 12.8 13.3056 13.8 14.5728 14.5856 14.5984 16.608 18.0608 21.2 22.3 23.4048 23.4944 27 31.7 37.1 56.2 60.9 76.6 87.3 110.2 115.3 116.3 122.1 139.4 145.7 149.8 227.1 232.3 247.5 272.6 302.7

Number of Observations Read	510
Number of Observations Used	246

The SAS System

The GLM Procedure

Dependent Variable: dtf

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	158	7439.207824	47.083594	12.69	<.0001
Error	87	322.828761	3.710675		
Corrected Total	245	7762.036585			

R-Square	Coeff Var	Root MSE	dtf Mean
0.958409	4.802601	1.926311	40.10976

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	42	6128.504443	145.916772	39.32	<.0001
l	7	453.614075	64.802011	17.46	<.0001
r(l)	8	72.621260	9.077657	2.45	0.0195
g*l	101	784.468046	7.767010	2.09	0.0002

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	42	5221.954995	124.332262	33.51	<.0001
l	7	439.599496	62.799928	16.92	<.0001
r(l)	8	74.671239	9.333905	2.52	0.0165
g*l	101	784.468046	7.767010	2.09	0.0002

The SAS System

The GLM Procedure

Dependent Variable: dtm

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	158	16958.10017	107.32975	5.94	<.0001
Error	87	1573.23723	18.08319		
Corrected Total	245	18531.33740			

R-Square	Coeff Var	Root MSE	dtm Mean
0.915104	3.914600	4.252433	108.6301

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	42	10816.44619	257.53443	14.24	<.0001
l	7	2900.25923	414.32275	22.91	<.0001
r(l)	8	446.70900	55.83863	3.09	0.0041
g*l	101	2794.68574	27.67016	1.53	0.0213

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	42	8712.841696	207.448612	11.47	<.0001
l	7	2307.390694	329.627242	18.23	<.0001
r(l)	8	125.262767	15.657846	0.87	0.5484
g*l	101	2794.685744	27.670156	1.53	0.0213

The SAS System

The GLM Procedure

Dependent Variable: yld

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	158	420683.1579	2662.5516	24.44	<.0001
Error	87	9476.5747	108.9261		
Corrected Total	245	430159.7326			

R-Square	Coeff Var	Root MSE	yld Mean
0.977970	72.44914	10.43677	14.40565

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	42	207236.1247	4934.1934	45.30	<.0001
l	7	92629.8548	13232.8364	121.48	<.0001
r(l)	8	7770.5081	971.3135	8.92	<.0001
g*l	101	113046.6703	1119.2740	10.28	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	42	69706.0152	1659.6670	15.24	<.0001
l	7	83787.6465	11969.6638	109.89	<.0001
r(l)	8	4146.9129	518.3641	4.76	<.0001
g*l	101	113046.6703	1119.2740	10.28	<.0001

Table 5. ANOVAs for days to flower (dtf), days to maturity (dtm), and yield (yld) for the RIL population 534545 x Canadian X.

The SAS System

The GLM Procedure

Class Level Information		
Class	Levels	Values
g	47	5301 5302 5303 5304 5305 5306 5307 5308 5309 5310 5311 5313 5314 5315 5316 5317 5318 5319 5320 5321 5322 5323 5324 5325 5326 5327 5328 5329 5330 5331 5332 5333 5334 5335 5336 5337 5338 5339 5340 5341 5342 5343 5344 5345 5346 5347 5348
l	8	1 2 4 5 6 7 8 9
r	2	1 2
dtf	22	28 30 31 33 34 35 37 38 40 41 42 43 44 45 46 47 48 49 50 51 52 53
dtm	41	84 86 87 88 89 91 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127
yld	431	0.6784 0.8128 1.0944 1.2672 1.28 1.4592 1.5 1.5168 1.5488 1.568 1.7024 1.7536 1.7728 1.8 1.8176 1.8496 1.9072 2.0416 2.048 2.0544 2.0672 2.0928 2.1504 2.1888 2.2272 2.3 2.4256 2.4832 2.4896 2.5152 2.5408 2.5856 2.6304 2.6944 2.7 2.7008 2.72 2.7968 2.8096 2.9 2.9376 2.9568 2.9696 3.0592 3.1424 3.232 3.264 3.2704 3.3 3.3024 3.3472 3.3792 3.3856 3.392 3.4432 3.456 3.4816 3.4944 3.6736 3.6928 3.7376 3.744 3.7632 3.776 3.84 3.872 3.8912 3.9 3.9808 4.0256 4.032 4.0704 4.0768 4.0832 4.1088 4.1152 4.1792 4.2 4.2048 4.2112 4.2496 4.2624 4.3 4.3072 4.3392 4.3648 4.3776 4.384 4.4608 4.4928 4.544 4.5824 4.6208 4.6528 4.7104 4.7232 4.736 4.7744 4.8064 4.928 5.0048 5.1072 5.12 5.1584 5.2416 5.2928 5.3376 5.3824 5.3888 5.4848 5.4912 5.5104 5.5424 5.5488 5.5744 5.5808 5.6832 5.696 5.7 5.7344 5.8304 5.8432 5.8752 5.9 5.9712 6.1184 6.208 6.2464 6.2528 6.2976 6.3104 6.3424 6.3552 6.3616 6.4448 6.464 6.5 6.5408 6.6048 6.6368 6.6688 6.7 6.7648 6.8 6.848 6.9184 6.944 6.9632 6.9824 7.0016 7.0336 7.072 7.104 7.1488 7.1936 7.2 7.2576 7.2768 7.36 7.3664 7.4432 7.7184 7.8336 7.9424 7.9872 8.1536 8.2 8.224 8.288 8.3136 8.3328 8.352 8.3648 8.384 8.4 8.4096 8.4416 8.4608 8.5 8.5248 8.6272 8.6656 8.704 8.7552 8.7936 8.832 8.8448 8.9 9 9.0688 9.1008 9.2 9.2928 9.3 9.3696 9.4592 9.4848 9.4912 9.5 9.5104 9.5296 9.6512 9.7 9.792 9.8752 9.9 9.9968 10 10.0032 10.0416 10.08 10.0928 10.0992 10.1888 10.2336 10.2464 10.3 10.3104 10.432 10.4448 10.4768 10.5856 10.6 10.6624 10.7 10.752 10.7968 11.0656 11.1 11.1808 11.2 11.2256 11.2512 11.2768 11.2832 11.3 11.3408 11.36 11.4496 11.4624 11.5328 11.9936 12 12.0512 12.0832 12.1 12.1344 12.352 12.3968 12.8 12.8256 12.9 12.9856 13 13.024 13.0816 13.1 13.1776 13.184 13.312 13.4 13.4528 13.4592 13.4912 13.5872 13.664 13.8 13.8112 13.824 13.9 14.016 14.0672 14.08 14.112 14.3 14.4192 14.56 14.624 14.7 15.2 15.2384 15.3 15.3216 15.328 15.5456 15.616 15.6288 15.8656 15.936 15.9936 16.2944 16.3 16.3264 16.5 16.7 16.7104 16.9728 17.1 17.2 17.2736 17.3 17.4 17.472 17.5552 17.6512 17.728 17.7344 17.9 18.1 18.5 18.6 18.6048 18.9 19.0592 19.2 19.456 19.5264 19.5392 19.7 19.7824 19.8 19.8848 20 20.0448 20.0704 20.1 20.3 20.3264 20.5568 20.5696 20.8 20.9 21 21.1 21.3 21.6 21.7152 21.9 22.2 22.368 22.7 23.3 23.4 23.8 24.4544 24.4992 24.5952 24.6 24.7 24.8 25.7 26 26.2336 26.8 26.9 27.3 27.5 27.7 27.9 28.2304 28.6912 29 29.8 29.9 31 31.1488 31.2 31.4 32.4 32.6 32.8 33.1 33.8 34 35.7 36.7 36.8 36.9 37.3 38.6 38.9 40.1 41.3 41.4 43.4 43.8 45.4 45.8 46.2 46.9 47.7 48 54.4 54.9 55.4 55.5 55.6 55.8 60.6 61.3 62.5 65.8 67.1 67.3 67.8 69 69.4 71.9 72.6 75.4 75.5 76.5 79.7 84.5 85 86.9 88.2 89.6 91 92.1 98.2 100.4 105.7 106.3 112.8 114.8 119.4 120.4 126.7 133.3 199.8

Number of Observations Read	734
Number of Observations Used	465

The SAS System

The GLM Procedure

Dependent Variable: dtf

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	272	13079.61525	48.08682	11.71	<.0001
Error	192	788.35034	4.10599		
Corrected Total	464	13867.96559			

R-Square	Coeff Var	Root MSE	dtf Mean
0.943153	4.941218	2.026325	41.00860

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	46	9955.417200	216.422113	52.71	<.0001
l	7	1632.444683	233.206383	56.80	<.0001
r(l)	8	32.365490	4.045686	0.99	0.4488
g*l	211	1459.387881	6.916530	1.68	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	46	7232.144063	157.220523	38.29	<.0001
l	7	1482.754286	211.822041	51.59	<.0001
r(l)	8	19.149662	2.393708	0.58	0.7912
g*l	211	1459.387881	6.916530	1.68	0.0001

The SAS System

The GLM Procedure

Dependent Variable: dtm

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	272	24723.20683	90.89414	5.12	<.0001
Error	192	3405.55876	17.73729		
Corrected Total	464	28128.76559			

R-Square	Coeff Var	Root MSE	dtm Mean
0.878930	3.807629	4.211566	110.6086

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	46	17327.47322	376.68420	21.24	<.0001
l	7	2610.36946	372.90992	21.02	<.0001
r(l)	8	274.49789	34.31224	1.93	0.0570
g*l	211	4510.86626	21.37851	1.21	0.0939

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	46	13847.78794	301.03887	16.97	<.0001
l	7	2280.47883	325.78269	18.37	<.0001
r(l)	8	47.44124	5.93015	0.33	0.9519
g*l	211	4510.86626	21.37851	1.21	0.0939

The SAS System

The GLM Procedure

Dependent Variable: yld

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	272	242137.8482	890.2127	4.53	<.0001
Error	192	37743.5059	196.5808		
Corrected Total	464	279881.3541			

R-Square	Coeff Var	Root MSE	yld Mean
0.865145	74.79352	14.02073	18.74591

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	46	61321.1463	1333.0684	6.78	<.0001
l	7	76219.8024	10888.5432	55.39	<.0001
r(l)	8	3816.9118	477.1140	2.43	0.0161
g*l	211	100779.9878	477.6303	2.43	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	46	33670.0248	731.9571	3.72	<.0001
l	7	70166.6234	10023.8033	50.99	<.0001
r(l)	8	4002.4081	500.3010	2.55	0.0117
g*l	211	100779.9878	477.6303	2.43	<.0001

Table 6. ANOVAs for days to flower (dtf), days to maturity (dtm), and yield (yld) for the RIL population X97 x Canadian X.

The SAS System

The GLM Procedure

Class Level Information		
Class	Levels	Values
g	39	5401 5402 5403 5404 5405 5406 5407 5408 5409 5410 5411 5412 5413 5414 5415 5416 5417 5418 5419 5420 5421 5422 5423 5424 5425 5426 5427 5428 5429 5430 5431 5432 5433 5434 5435 5436 5437 5438 5440
l	9	1 2 3 4 5 6 7 8 9
r	2	1 2
dtf	20	28 29 30 31 33 34 35 37 38 40 41 42 43 44 45 46 47 49 51 52
dtm	36	91 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 128
yld	321	0.64 0.8384 1.248 1.3 1.6384 1.664 1.6704 1.7536 1.8816 2.2656 2.3232 2.4448 2.5216 2.5408 2.6 2.6112 2.6176 2.7 2.7136 2.752 2.7712 2.7776 2.7904 2.8352 2.9 2.9184 2.9312 3.0336 3.0656 3.0976 3.104 3.1808 3.2896 3.3 3.3408 3.3472 3.5072 3.5456 3.5648 3.5712 3.616 3.6224 3.6672 3.7056 3.7184 3.8 3.872 3.8784 3.9 3.9168 3.9296 4 4.0704 4.1024 4.1152 4.2368 4.256 4.288 4.3 4.3072 4.32 4.3648 4.3712 4.3776 4.3904 4.3968 4.4288 4.448 4.4928 4.5056 4.5568 4.5632 4.6464 4.6528 4.7 4.7872 4.8256 4.8384 4.9024 4.9344 4.9856 5.1584 5.2224 5.2352 5.3 5.3504 5.4 5.4528 5.4848 5.5296 5.5936 5.6128 5.6192 5.6448 5.6512 5.7088 5.7152 5.824 5.8304 5.8432 5.9392 5.9648 6.1 6.144 6.1696 6.2 6.2656 6.304 6.3104 6.368 6.3872 6.4128 6.4832 6.5024 6.5728 6.6624 6.7 6.752 6.784 6.88 6.9312 6.9504 7.0016 7.0592 7.0912 7.1872 7.232 7.3216 7.456 7.4688 7.5392 7.552 7.5648 7.6288 7.648 7.6544 7.712 7.7888 7.9 7.9168 7.936 7.9936 8 8.0448 8.0704 8.1024 8.1408 8.2368 8.5 8.6848 8.8 8.9 8.9472 8.9728 9.0816 9.088 9.248 9.2992 9.3056 9.312 9.3184 9.9136 10 10.016 10.0992 10.2656 10.2912 10.4 10.5 10.5216 10.56 10.8096 10.9248 11 11.0848 11.2 11.3 11.4048 11.5 11.5072 11.8592 11.9744 11.9872 12.0448 12.3456 12.7168 12.7552 12.9 12.9088 12.928 13.0176 13.1 13.6 13.9072 13.9264 14.0608 14.1248 14.1376 14.144 14.2336 14.496 14.8 14.976 15 15.2768 15.3 15.5 15.5072 15.7 15.904 16.5248 16.6656 16.6784 16.7 17.216 17.3 17.7088 18.2784 18.3 18.4 18.5 18.8 18.8672 19 19.0144 19.2 19.3 20 21.2 21.3 21.4 21.9 22.5 22.9 23.5904 23.904 24.1 25.2 26.3 26.9 27.1104 27.2704 27.9 28 28.3 28.6 29.728 29.8 30.4 30.6752 30.8 31.4 31.8 31.9 32.1 32.8 33.1 33.3 33.8 34.1 34.3 34.9 35.5 35.9 36.3 36.4 37.1 38 38.1 39 39.1 39.3 40 40.8 41.4 41.5 42.9 44.3 44.6 44.8 45 45.5 46.2 47.1 47.6 49.5 51.8 52.7 52.9 54.7 57.7 59 61.2 63.9 64.8 67.1 67.8 69.6 70.8 74.3 78.2 84.2 85.3 88 88.1 89.6 91.6 102.1 106.8 111.1 111.9 122.6 126.9 133.2 133.5 168.8 177.9 187 234.7 243.2 328.9

Number of Observations Read	590
Number of Observations Used	331

The SAS System

The GLM Procedure

Dependent Variable: dtf

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	201	7870.679633	39.157610	9.72	<.0001
Error	129	519.519763	4.027285		
Corrected Total	330	8390.199396			

R-Square	Coeff Var	Root MSE	dtf Mean
0.938080	4.960081	2.006810	40.45921

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	38	4930.399680	129.747360	32.22	<.0001
l	8	1601.083487	200.135436	49.69	<.0001
r(l)	8	49.923821	6.240478	1.55	0.1465
g*l	147	1289.272646	8.770562	2.18	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	38	3175.454853	83.564601	20.75	<.0001
l	8	1666.243468	208.280433	51.72	<.0001
r(l)	8	56.480237	7.060030	1.75	0.0923
g*l	147	1289.272646	8.770562	2.18	<.0001

The SAS System

The GLM Procedure

Dependent Variable: dtm

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	201	15445.60749	76.84382	5.78	<.0001
Error	129	1715.14175	13.29567		
Corrected Total	330	17160.74924			

R-Square	Coeff Var	Root MSE	dtm Mean
0.900054	3.299885	3.646323	110.4985

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	38	8765.080588	230.660015	17.35	<.0001
l	8	3143.950996	392.993875	29.56	<.0001
r(l)	8	93.935164	11.741896	0.88	0.5327
g*l	147	3442.640744	23.419325	1.76	0.0005

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	38	7039.947037	185.261764	13.93	<.0001
l	8	2800.148279	350.018535	26.33	<.0001
r(l)	8	84.858248	10.607281	0.80	0.6055
g*l	147	3442.640744	23.419325	1.76	0.0005

The SAS System

The GLM Procedure

Dependent Variable: yld

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	201	369167.6537	1836.6550	2.79	<.0001
Error	129	85058.9513	659.3717		
Corrected Total	330	454226.6050			

R-Square	Coeff Var	Root MSE	yld Mean
0.812739	110.8041	25.67823	23.17444

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	38	158789.6887	4178.6760	6.34	<.0001
l	8	85050.9537	10631.3692	16.12	<.0001
r(l)	8	3103.8533	387.9817	0.59	0.7860
g*l	147	122223.1579	831.4501	1.26	0.0888

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	38	95415.3663	2510.9307	3.81	<.0001
l	8	78125.3167	9765.6646	14.81	<.0001
r(l)	8	2783.1680	347.8960	0.53	0.8340
g*l	147	122223.1579	831.4501	1.26	0.0888

Table 7. ANOVAs for days to flower (dtf), days to maturity (dtm), and yield (yld) for the RIL population X97 x Jenguma.

The SAS System

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
g	44	5501 5504 5505 5506 5508 5509 5510 5511 5512 5514 5516 5517 5518 5519 5520 5521 5522 5524 5525 5526 5527 5528 5530 5532 5533 5534 5536 5537 5538 5539 5540 5541 5542 5543 5544 5545 5546 5549 5551 5553 5554 5555 5560 5561
l	8	1 2 4 5 6 7 8 9
r	2	1 2
dtf	20	28 29 30 31 32 33 34 35 37 38 40 41 42 43 44 45 46 47 49 52
dtm	41	72 81 83 84 85 86 87 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123
yld	606	0.448 1.4976 1.6 1.8752 2.3296 2.4 2.7264 2.9 2.9056 2.9952 3.3 3.3536 3.3984 3.5 3.6 3.6544 3.9808 4 4.0512 4.096 4.1 4.1024 4.352 4.4288 4.6208 4.7488 4.8 4.9024 4.9088 4.9792 5.0304 5.1 5.1712 5.3 5.312 5.3888 5.536 5.632 5.6768 5.6896 5.792 5.8624 5.8752 5.888 5.9 5.9456 6.0608 6.0928 6.1504 6.3104 6.368 6.3808 6.3936 6.6 6.6752 6.6944 6.8224 6.8352 6.8608 6.9632 6.9952 7.0272 7.1552 7.1744 7.1808 7.392 7.4368 7.4944 7.5136 7.552 7.6416 7.6672 7.7312 7.7824 7.808 7.84 7.8528 7.8592 7.8848 7.9424 8.0192 8.1 8.1216 8.1472 8.1984 8.3 8.32 8.3264 8.4672 8.4928 8.544 8.5568 8.5632 8.704 8.736 8.7808 8.8 8.8832 8.9024 8.9664 8.9792 8.992 9.1008 9.1136 9.1648 9.248 9.2544 9.3 9.3632 9.376 9.4272 9.472 9.5104 9.5168 9.5232 9.5872 9.6448 9.6832 9.6896 9.696 9.7024 9.8432 9.9 10.0096 10.0544 10.0608 10.0992 10.2 10.2336 10.2976 10.3168 10.3744 10.3936 10.4 10.4384 10.4704 10.4896 10.5472 10.656 10.8096 10.8352 10.8736 10.8928 10.9056 10.9312 10.9696 11.0528 11.0784 11.0848 11.1104 11.1232 11.136 11.1936 11.2192 11.2832 11.2896 11.3088 11.3728 11.4432 11.4816 11.5904 11.6352 11.7 11.7504 11.7696 11.8144 11.8464 11.8848 11.8912 11.9168 11.9232 11.9296 11.9488 12.0704 12.0896 12.1024 12.192 12.32 12.3456 12.4 12.4288 12.5376 12.5504 12.5952 12.64 12.6656 12.7168 12.736 12.896 12.9984 13.3376 13.4336 13.5 13.7 13.7152 13.728 13.7728 13.8688 13.9968 14 14.1 14.1056 14.112 14.2144 14.2784 14.2912 14.304 14.3232 14.368 14.3872 14.5216 14.6752 14.7 14.7136 14.816 14.848 14.88 14.9632 15.0272 15.1296 15.2128 15.264 15.2768 15.328 15.4304 15.4368 15.5 15.5072 15.5584 15.616 15.6224 15.6736 15.8016 15.8528 15.9104 15.936 15.9552 15.9936 16 16.064 16.1088 16.192 16.2048 16.2112 16.384 16.576 16.6144 16.6208 16.6784 16.7232 16.7808 16.9664 16.9792 17.0432 17.0752 17.1328 17.1712 17.2032 17.2352 17.2672 17.5 17.5808 17.6192 17.6384 17.728 18 18.1824 18.2656 18.4576 18.5 18.8224 18.8288 18.8352 18.8608 19.1296 19.1872 19.2576 19.3408 19.8208 20 20.224 20.2304 20.2816 20.384 20.4 20.4096 20.4992 20.6656 20.8 20.8448 21.1008 21.1584 21.312 21.5936 21.632 21.7536 21.8752 22.1 22.144 22.1632 22.2464 22.5 22.5152 22.6 22.7072 22.8544 23 23.04 23.3 23.3856 23.4688 23.5328 23.5776 23.6224 23.7 24.2368 24.3648 24.3712 24.384 24.4 24.5 24.512 24.7 24.8 24.8192 25.0112 25.4 25.6064 25.9904 26.0224 26.144 26.3616 26.5024 26.5344 26.9696 27.2 27.3408 27.4 28.2624 28.8 28.9 29.1 29.1392 29.2608 29.9072 30 30.0032 30.7 30.8608 31.1 31.232 31.5 31.8 31.808 32.096 32.4352 32.5504 32.6 32.6528 32.7104 32.8 32.9024 33.3056 33.5 34.3 34.4 34.88 34.9 34.912 35.2 35.4 36.1 36.4 36.6 37.1 37.5 37.7 38 38.8928 39.1296 39.3 39.4 39.4688 39.6 40.032 40.2 40.3 41.0432 41.5 41.9 42.7 42.8 42.9 43.4 45.2 45.44 46.5 47.3 49.5 52.1 53.1 53.3 53.9 54.9 55 55.3 57.5 58.7 59.4 59.6 60.5 61 61.1 61.7 62.7 62.8 62.9 63.2 63.7 64 64.2 64.5 64.9 65.3 66.1 66.3 66.4 66.5 66.9 67 68.1 68.2 68.3 69 70.3 71.4 71.6 71.9 72.6 73.5 73.9 74 75.4 76.9 78.4 78.5 78.8 79.9 81.6 82.1 82.9 84.6 86.9 88.2 88.4 90.3 90.8 91.6 92 92.1 92.7 92.9 93.4 96 97 97.4 98.9 100.2 100.8 102.5 102.7 102.9 103.3 103.9 104.2 104.5 104.6 104.7 105 105.2 107.6 109.2 110.1 110.7 110.8 113.7 114.6 116.4 116.6 118.6 118.7 119.2 119.5 121.1 122.9 125.7 126.9 127 128 128.3 128.9 129.4 130.7 132.4 133.2 135.4 135.7 136.3 138.1 140.9 145.7 145.9 146.7 146.9 147.8 148.4 149.2 150.4 150.5 155.2 156.2 158.6 159.7 163.2 163.7 165.4 165.8 166.1 166.5 167.4 169 169.3 169.6 170.2 170.6 170.8 171.7 173.7 175.4 176.4 177 178.4 179.7 180.9 181.6 182.7 183.3 184 184.2 184.5 185.1 185.4 189.4 190.6 193.4 194.4 195.8 197.4 197.5 198.1 204.1 205.5 206.6 210.4 211.3 211.4 214.9 217.3 218.2 219 220.3 222.4 223.4 230.9 232.1 235.5 237.3 241.5 246 246.2 251.1 251.4 254.8 262.9 265.3 267.3 273.8 280.1 282.9 285 285.3 288.9 295.4 303 312.6 316.2 319.4 323.8 328.6 336.1 348 351 371.2 425.5 434 477.5

Number of Observations Read	928
Number of Observations Used	622

The SAS System

The GLM Procedure

Dependent Variable: dtf

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	334	10331.60522	30.93295	8.85	<.0001
Error	287	1003.58449	3.49681		
Corrected Total	621	11335.18971			

R-Square	Coeff Var	Root MSE	dtf Mean
0.911463	5.072061	1.869976	36.86817

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	43	7235.912656	168.277039	48.12	<.0001
l	7	1083.949359	154.849908	44.28	<.0001
r(l)	8	52.369765	6.546221	1.87	0.0642
g*l	276	1959.373444	7.099179	2.03	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	43	6943.816022	161.484094	46.18	<.0001
l	7	954.727572	136.389653	39.00	<.0001
r(l)	8	45.915513	5.739439	1.64	0.1128
g*l	276	1959.373444	7.099179	2.03	<.0001

The SAS System

The GLM Procedure

Dependent Variable: dtm

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	334	35665.48663	106.78289	7.47	<.0001
Error	287	4103.33813	14.29735		
Corrected Total	621	39768.82476			

R-Square	Coeff Var	Root MSE	dtm Mean
0.896820	3.593589	3.781183	105.2203

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	43	17250.98034	401.18559	28.06	<.0001
l	7	9414.29945	1344.89992	94.07	<.0001
r(l)	8	421.01794	52.62724	3.68	0.0004
g*l	276	8579.18891	31.08402	2.17	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	43	14596.64910	339.45696	23.74	<.0001
l	7	9094.04720	1299.14960	90.87	<.0001
r(l)	8	419.16187	52.39523	3.66	0.0004
g*l	276	8579.18891	31.08402	2.17	<.0001

The SAS System

The GLM Procedure

Dependent Variable: yld

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	334	3319918.608	9939.876	4.53	<.0001
Error	287	629794.934	2194.407		
Corrected Total	621	3949713.542			

R-Square	Coeff Var	Root MSE	yld Mean
0.840547	75.28314	46.84450	62.22443

Source	DF	Type I SS	Mean Square	F Value	Pr > F
g	43	598358.526	13915.315	6.34	<.0001
l	7	1739781.134	248540.162	113.26	<.0001
r(l)	8	27525.976	3440.747	1.57	0.1340
g*l	276	954252.972	3457.438	1.58	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
g	43	430840.803	10019.554	4.57	<.0001
l	7	1618272.189	231181.741	105.35	<.0001
r(l)	8	28651.350	3581.419	1.63	0.1153
g*l	276	954252.972	3457.438	1.58	<.0001

VITA

Carrie Miranda was born July 1st, 1983 in San Francisco, California to Donna Miranda as she was pursuing her talent for fashion design. They both moved back to Donna's hometown of Cleveland, Ohio, when Carrie was three. Carrie graduated from Trinity High School in 2001. She briefly attended the Ohio State University her freshman year but transferred to Cleveland State University for a smaller campus. She graduated with her B.S. in Biology in 2005 with a research emphasis in invasive species removal. She then completed a year of volunteer work with AmeriCorp doing invasive species removal throughout the Southwestern United States. She then pursued her love of travel by teaching English in Seoul, South Korea for two years. During this time, she still worked in the scientific field by volunteering in Sunghwa Choe's plant molecular biology lab at Seoul National University. She then moved to San Diego to complete her M.S. in Molecular Biology at San Diego State University with William Stumph. Her research aimed at understanding the recruitment of two different RNA polymerases by a single transcription factor through a structural biology approach. She then moved to Columbia, Missouri to pursue a PhD researching plant breeding in an international setting with Kristin Bilyeu. She has one beloved dog, Juan, and leads an adventurous life when her research allows it.