

ABSTRACT

WALDECK, NATHAN JAMES. Characterization of the Effects of Extreme Temperature on Germination and Elevated Ozone Stress Response in Wild Soybean and Expanding Genetic Diversity for Abiotic Stress Resistance within the Soybean Gene Pool (Under the direction of Dr. Earl Taliercio).

Soybean (*Glycine max* (L.) Merr.) is an important crop for oil and protein worldwide. Considered one of the first domesticated crops, soybean has experienced a narrowing of genetic diversity with the majority of North American public breeding cultivars tracing back to only seventeen ancestral lines. With this narrow genetic base, it is feared that yield improvement as well as the plant's ability to deal with stresses will be limited. Therefore, increasing genetic diversity from different sources within the soybean gene pool has become of interest in recent years.

The ancestor of soybean (*Glycine soja* Zucc and Sieb) is a weedy species found in many areas of Asia and has not experienced the genetic bottlenecks associated with domesticated soybean. It has proven to be a source of stress resistance and yield-enhancing alleles that are not found in its cultivated relative. Because wild soybean creates fertile offspring with cultivated soybean, unique stress responses found in wild sources may be transferred to cultivated soybean through conventional plant breeding.

In this research, a diverse set of wild soybean accessions is characterized to identify genetic responses to abiotic stressors that soybean may encounter. In that regard, the ideal germination temperature for soybean is thought to be 28 – 30 °C. If global temperatures rise as expected, plants could be exposed to higher temperatures at planting. In addition, early planting schemes are becoming popular for soybean cultivation in certain areas of the United States, which could expose soybean to lower germination temperatures. Results from this

thesis study indicated that wild soybean accessions display a range of responses when subjected to suboptimal temperatures at germination. Differences in both percent germination as well as average root growth in both hot and cold temperatures were identified.

Ozone is another abiotic stress that has quadrupled in the past 150 years and can be detrimental to plant productivity. The ozone response of domesticated soybean genotypes has been reported, though wild soybean response is less well characterized. This thesis research measures the range of responses to increased ozone levels among a diverse core set of wild soybean accessions. Gene expression analysis on two ozone-sensitive accessions and two ozone-resistant accessions shows much greater response to ozone exposure in sensitive lines as compared to resistant counterparts. Defense response, oxidoreductase activity, and photosynthesis are among categories of genes most affected by ozone exposure. In addition, genes responding in the apoplast of the plant cell are affected indicating a control point of ozone damage for the plant. Differences of gene response in leaves of different maturities were also noted in this study. Older leaves generally displayed greater reactions to ozone. Responses of the diverse lines were further used to identify associations with SNP markers. A putative marker was identified for ozone response unique to wild soybean on Chromosome 19.

© Copyright 2016 by Nathan James Waldeck

All Rights Reserved

Characterization of the Effects of Extreme Temperature on Germination and Elevated Ozone
Stress Response in Wild Soybean and Expanding Genetic Diversity for Abiotic Stress
Resistance within the Soybean Gene Pool

by
Nathan James Waldeck

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

Crop Science

Raleigh, North Carolina

2016

APPROVED BY:

Earl Taliercio
Committee Chair

Kent Burkey

Gina Brown-Guedira

Thomas E. Carter, Jr.

David Dickey

DEDICATION

To my family.

This thesis is dedicated to my family members who have been encouraging in all of my ventures: first and foremost, my parents, Lynn and Lisa Waldeck, as well as my grandparents Richard and Deanna Bails and Marvin and Linda Waldeck. All of the aforementioned have shown a positive example in life as well as in agriculture, which sparked my interest in science. Without your examples and encouragement, this thesis would not have been possible.

BIOGRAPHY

Nathan James Waldeck was born in 1991 and raised just outside of Farmersville, Illinois. His experience in agriculture began at a very young age on the family farm. A graduate of Lincolnwood High School in Raymond, IL, he went on to pursue a B.S. in Crop Sciences with a concentration in Plant Biotechnology and Molecular Biology at the University of Illinois at Urbana-Champaign, which he completed in 2014. During his college career, he worked as an undergraduate research assistant in soybean breeding for Dr. Brian Diers, where he gained his first experience in applied genetics research. Upon completion of his B.S., he decided to further his studies and pursue a degree in plant breeding and genetics from North Carolina State University. His Master of Science work was conducted in the USDA Soybean & Nitrogen Fixation Unit under the direction of Dr. Earl Taliercio.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor, Dr. Earl Taliercio. His attitude toward truly training me for future success and his appropriate guidance during the duration of this research has been phenomenal. His drive to try new approaches and willingness to challenge accepted methods has given me a foundation for the continuation of my research career. It has been a great experience working for him.

I would also like to thank my committee members, Dr. Kent Burkey, Dr. Gina Brown-Guedira, Dr. Thomas Carter, and Dr. David Dickey. Each of you played a unique role in methods involved in this thesis, and I thank you for all of your assistance. I would also like to thank Dr. Ross Whetten for all of his bioinformatics assistance involved in my ozone experiment.

The current and past staff at the USDA Soybean and Nitrogen Fixation Unit also deserve recognition for their contributions to this research: Edgar Alvarez, Zac Bryan, David Eickholt, Carolyn Fox, Telisa Loveless, Dina Espinosa-Rivera, Martha Ramirez, and Rebecca Whetten. Thank you all for your patience and assistance. Also, thank you to Tan Tuong and the Dr. David Livingston lab for allowing access to growth chambers for the temperature experiments and the personnel in Dr. Kent Burkey's lab, notably Renee Tucker and Jeff Barton, for their help with planning, rating, and harvesting tissue for ozone experiments. These were a large effort, and thanks to everyone involved in extended.

Finally, I would like to thank the North Carolina Soybean Producers Association for providing funding for my research. Without the interest in increasing genetic diversity in soybean, this thesis would not have been possible.

TABLE of CONTENTS

LIST of TABLES.....	viii
LIST of FIGURES	xi
Chapter I: Review of the Literature	1
INTRODUCTION	2
GENETIC DIVERSITY IN SOYBEAN	2
Limited Diversity Exists in Soybean Breeding Programs	2
Use of <i>Glycine soja</i> in soybean breeding:.....	3
ABIOTIC STRESS	7
Temperature Effects on Soybean Germination.....	7
Increasing Ozone Levels Affect Plant Growth	8
GENE EXPRESSION METHODOLOGY	12
REFERENCES	16
Chapter II: Evaluation of Germination at Sub-Optimal Temperatures among Genetically Diverse <i>Glycine soja</i> Accessions	27
ABSTRACT.....	28
INTRODUCTION	28
MATERIALS AND METHODS.....	30
Germination of <i>Glycine soja</i> at Sub-Optimal Temperatures	30
Analysis of Germination Data	33
RESULTS	34
Evaluation of Percent Germination at Extreme Cold Temperatures	34
Evaluation of Average Root Length at Extreme Cold Temperatures	35
Evaluation of Percent Germination at Extreme Hot Temperatures	35
DISCUSSION	37
CONCLUSION.....	42
REFERENCES	43
TABLES and FIGURES	45
Chapter III: RNA-Seq Study Reveals Genetic Response of Diverse Wild Soybean Accessions to Increased Ozone Levels	62
ABSTRACT.....	63
INTRODUCTION	64

MATERIALS and METHODS.....	67
Identification of Phenotypic Response of <i>Glycine soja</i> to Increased Ozone.....	67
Collection of Ozone-Treated Samples.....	68
Isolation of RNA and Sequencing.....	69
Analysis of RNA-Seq Data.....	69
RESULTS.....	71
Sequencing of Ozone Sensitive and Resistant Wild Soybean Lines.....	71
Analysis of RNA-Seq Data.....	72
Ozone Response in Sensitive Genotypes PI 407179 and PI 424007.....	73
Ozone Response in Resistant Genotypes PI 424123 and PI 507656.....	76
Comparison of Resistant and Sensitive Genotype Response to Ozone.....	78
Ozone Response Varies in Genotype Based on Leaf Position.....	80
DISCUSSION.....	83
CONCLUSION.....	87
REFERENCES.....	88
TABLES and FIGURES.....	93
Chapter IV: Incorporation of Abiotic Stress Resistance from <i>Glycine soja</i> into Germplasm Releases for Soybean Breeding.....	117
ABSTRACT.....	118
INTRODUCTION.....	118
MATERIALS and METHODS.....	120
Genome Wide Association Study for Ozone Resistance.....	120
Development and Progression of Genetically Diverse Soybean Lines.....	121
RESULTS.....	122
Genome Wide Association Study for Ozone Resistance.....	122
Development and Progression of Genetically Diverse Soybean Lines.....	123
DISCUSSION.....	124
Genome Wide Association Study for Ozone Resistance.....	124
Development and Progression of Genetically Diverse Soybean Lines.....	126
CONCLUSION.....	128
REFERENCES.....	129
TABLES and FIGURES.....	131

APPENDIX.....	139
Appendix A. LSMEAN estimates of percent germination and average root length for each genotype tested at each of the temperatures used in the cold germination experiment. Summary statistics for each temperature are listed at the bottom of the table.....	140
Appendix B. LSMEAN estimates of percent germination and average root length for each genotype tested at each of the temperatures used in the hot germination experiment. Summary statistics for each temperature are listed at the bottom of the table.....	142
Appendix C. Correlation of means, screening Glycine soja genotypes at 8 temperatures for Percent Germination and Average Root Length.....	144
Appendix D. Geographical Information for USDA <i>Glycine soja</i> Diversity Core Collection used in characterization of abiotic stress response.....	145

LIST of TABLES

Table 2.1. Sample statistics of percent germination for cold germination temperatures (Temp). While average percent germination are similar at 17, 22, and 28 °C, there is a significant difference at 11 °C. Minimum and Maximum indicate the highest and lowest values recorded for each temperature.	45
Table 2.2. Percent germination ANOVA for evaluation of genotypes at cold germination temperatures. This model initially included the interaction term of genotype and temperatures. While this term was not left in the final model, it was included in this table to show the lack of significance ($p = .9989$). It was excluded from later models following the same test.	45
Table 2.3. Least squares estimates of mean percent germination and average root length for genotypes tested at cold and hot temperatures arranged by genotype.	46
Table 2.4. Sample statistics of average root length for cold germination temperatures. Root length increased as temperature approached the theoretical ideal, but standard deviations did as well. A large number of samples were present at each temperature level. Minimum and Maximum indicate the highest and lowest values recorded for each temperature.	48
Table 2.5. Average root length ANOVA for evaluation of genotypes at cold germination temperatures. The two significant fixed effects, genotype and temperature, were used for the final model. These values were used in determining significant differences in genotype based on percent germination.	48
Table 2.6. Sample statistics of percent germination for hot germination temperatures. Minimum and Maximum indicate the highest and lowest values recorded for each temperature.	49
Table 2.7. Percent germination ANOVA for evaluation of genotypes at hot germination temperatures. While the interaction term was not left in the final model, it was included in this table to show the lack of significance ($p = 1.0000$).	49
Table 2.8. Sample statistics of average root length for hot germination temperatures. Number of observations was lower for 43 °C due to lack of measurable roots from seeds not germinating at this high temperature. Minimum and Maximum indicate the highest and lowest values recorded for each temperature.	49
Table 2.9. Average root length ANOVA for evaluation of genotypes at hot germination temperatures.	49

Table 3.1. Average ozone damage ratings for selected USDA *Glycine soja* diversity core collection. Average represents the averaged leaf injury score, while SE represents the associated standard error of the measurement for 3 replications. Lower values indicate resistance, while greater values indicate sensitivity. Accessions highlighted in green indicated those identified as sensitive or resistant and used to RNA-Seq sequenced in this study.....93

Table 3.2. Differential expression of selected *Glycine soja* genotypes following ozone exposure. The table lists the number of genes responding in both directions at least 2-fold between ozone treated and the control in the genotypes from each of the three leaf positions collected (T1, T2 and T3) as well as the total number in each genotype accounting for genes expressed in multiple leaves. 94

Table 3.3. Spearman correlation coefficients of log2-fold changes between Fragments Per Kilobase of transcript per Million mapped reads (FPKM)-values of CF and ozone-treated leaf samples both within genotype at each different leaf collection point as well as between the two sensitive and two resistant genotypes at each leaf position. Higher values indicate more similarity in gene response. Geno 1 & 2 represent the genotype for which the comparison is made..... 95

Table 3.4. Filtered gene ontology categories for down-regulated genes in each genotype following exposure to increased ozone levels. Gene numbers are displayed under the column for designated by the corresponding genotype name. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Absence of a number in a column does not necessarily refer to zero responding genes but rather lack of significant enrichment (FDR < 0.05)..... 96

Table 3.5. Filtered gene ontology categories for up-regulated genes in each genotype following exposure to increased ozone levels. Gene numbers are displayed under the column for designated by the corresponding genotype name. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Absence of a number in a column does not necessarily refer to zero responding genes but rather lack of significant enrichment (FDR < 0.05)..... 99

Table 3.6. List of specific differentially-expressed genes shared by both resistant genotypes (PIs 424123 and 507656) but by neither sensitive genotype (PIs 407179 and 424007). Genes in this list were all down-regulated; no unique genes were up-regulated and shared by both resistant genotypes. Genes without a known annotation were excluded from this table..... 101

Table 3.7. List of genes that were down regulated in both resistant genotypes and up regulated in both sensitive genotypes. 102

Table 3.8. Filtered gene ontology categories for differentially-expressed genes in sensitive accession PI 407179 following exposure to increased ozone levels. Category refers to the

Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Shared by all positions signifies genes differentially expressed in each of the 3 trifoliates, while trifoliolate subheadings refer to genes that were only expressed at that position. “DE” refers to significant genes pooled in either direction, while “Up” and “Down” refer to the direction of genes matching each category when run separately. 103

Table 3.9. Filtered gene ontology categories for differentially-expressed genes in sensitive accession PI 424007 following exposure to increased ozone levels. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Shared by all positions signifies genes differentially expressed in each of the 3 trifoliates, while trifoliolate subheadings refer to genes that were only expressed at that position. “DE” refers to significant genes pooled in either direction, while “Up” and “Down” refer to the direction of genes matching each category when run separately. 105

Table 3.10. Filtered gene ontology categories for differentially-expressed genes in resistant accession PI 424123 following exposure to increased ozone levels. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Shared by all positions signifies genes differentially expressed in each of the 3 trifoliates, while trifoliolate subheadings refer to genes that were only expressed at that position. “DE” refers to significant genes pooled in either direction, while “Up” and “Down” refer to the direction of genes matching each category when run separately. 107

Table 3.11. Filtered gene ontology categories for differentially-expressed genes in resistant accession PI 507656 following exposure to increased ozone levels. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Shared by all positions signifies genes differentially expressed in each of the 3 trifoliates, while trifoliolate subheadings refer to genes that were only expressed at that position. “DE” refers to significant genes pooled in either direction, while “Up” and “Down” refer to the direction of genes matching each category when run separately. 108

Table 4.1. Tests of significance for ozone association in Single Nucleotide Polymorphism markers (SNP) in wild soybean diversity core accessions following GLM procedure in TASSEL 5 sorted by permutation p-value. Top 20 markers are shown in this table. 131

Table 4.2. List of genes in 1 Mbp region around putative ozone SNP marker ss715636367 at position 8227790 on Chromosome 19. Gene models without annotations were excluded; protein length (AA) refers to the number of 3 base pair amino acids in each gene model... 132

Table 4.3. ANOVA on significance of sieve size on Average 100 Seed Weight for F_{4:5} *G. max* x *G. soja* progeny lines in Population 23. 133

LIST of FIGURES

- Figure 2.1.** Needle plot of genotypic LS means estimates of percent germination for maturity groups III and IV tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A..... 50
- Figure 2.2.** Needle plot of genotypic LS means estimates of percent germination for maturity group V tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A..... 51
- Figure 2.3.** Needle plot of genotypic LS means estimates of percent germination for maturity groups VI and VII tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A..... 52
- Figure 2.4.** Needle plot of genotypic LS means estimates of average root length for maturity groups III and IV tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048..... 53
- Figure 2.5.** Needle plot of genotypic LS means estimates of average root length for maturity group V tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048..... 54
- Figure 2.6.** Needle plot of genotypic LS means estimates of average root length for maturity groups VI and VII tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048..... 55
- Figure 2.7.** Needle plot of genotypic LS means estimates of percent germination for maturity groups III and IV tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A..... 56

Figure 2.8. Needle plot of genotypic LS means estimates of percent germination for maturity group V tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A. 57

Figure 2.9. Needle plot of least squares estimates of percent germination for each of the MG VI and VII tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A. 58

Figure 2.10. Needle plot of genotypic LS means estimates of average root length for maturity groups III and IV tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048. 59

Figure 2.11. Needle plot of genotypic LS means estimates of average root length for maturity group V tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048. 60

Figure 2.12. Needle plot of genotypic LS means estimates of average root length for maturity groups III and IV tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048. 61

Figure 3.1. Venn diagram showing comparison of number of genes differentially expressed between ozone-resistant genotypes 424123 and 507656. 110

Figure 3.2. Venn diagram showing comparison of number of genes differentially expressed between ozone-sensitive genotypes 407179 and 424007. 111

Figure 3.3. Venn diagram showing comparison of number of genes differentially expressed between all genotypes used in the *Glycine soja* Ozone Study. This diagram shows several interesting numbers such as 2,119 genes shown to be uniquely expressed by both sensitive genotypes but not by either of the resistant genotypes and that 29 genes are shared by both of the resistant sources but neither sensitive genotype. 112

Figure 3.4. Heatmap of FPKM values for selected oxidative stress genes in ozone sensitive genotypes with twelve differentially-expressed genes related to oxidation/reduction reactions in ozone sensitive genotypes PIs 407179 and 424123 at each trifoliolate (T1, T2, T3). CF indicates a sample that was controlled in charcoal-filtered conditions, while Trt signifies exposure to 75 ppb ozone. Level of expression (FPKM) is noted by the intensity of color in each cell with high expression being dark purple and low expression being white. 113

Figure 3.5. Heatmap of FPKM values for selected apoplastic genes in ozone sensitive genotypes of nine differentially-expressed genes in the apoplasts of ozone sensitive genotypes PIs 407179 and 424007 at each trifoliolate (T1, T2, T3). These genes correspond to genes thought to code for xyloglucan endotransglucosylases/hydrolases. CF indicates a sample that was controlled in charcoal-filtered conditions, while Trt signifies exposure to 75 ppb ozone. Level of expression (FPKM) is noted by the intensity of color in each cell with high expression being dark purple and low expression being closer to white. 114

Figure 3.6. Heatmap of FPKM Values for selected oxidative stress genes in ozone resistant genotypes with twelve differentially-expressed genes related to oxidation/reduction reactions in ozone resistant genotypes PIs 424123 and 507656 at each trifoliolate (T1, T2, T3). CF indicates a sample that was controlled in charcoal-filtered conditions, while Trt signifies exposure to 75 ppb ozone. Level of expression (FPKM) is noted by the intensity of color in each cell with high expression being dark red and low expression being light yellow. 115

Figure 3.7. Heatmap of FPKM values for selected apoplastic genes in ozone resistant genotypes with nine differentially-expressed genes in the apoplasts of ozone resistant genotypes PIs 424123 and 507656 at each trifoliolate (T1, T2, T3). CF indicates a sample that was controlled in charcoal-filtered conditions, while Trt signifies exposure to 75 ppb ozone. Level of expression (FPKM) is noted by the intensity of color in each cell with high expression being dark red and low expression being light yellow. 116

Figure 4.1. Selections are made on individual F₃ plants from a *G. max* X *G. soja* cross. On the left is an example of a viney, unfavorable plant while the plant on the right displays a cultivated soybean appearance and would be selected for advancement. 134

Figure 4.2. Breeding scheme for Population 23, a cross developed to increase genetic diversity in soybean now connected to response to elevated ozone. 135

Figure 4.3. Manhattan plot of markers associated with ozone response in *Glycine soja*. Markers circled in red are those that were potentially most significant, while the marker on Chromosome 19 proved significant after permutation testing. 136

Figure 4.4. Bar chart of the number of F_{4;5} selections from original seed sieve size of F₃ plants for Population 23 137

Figure 4.5. Box plots of 100 seed weights (wt100) of $F_{4:5}$ Progeny grouped by sieve class of each F_3 ancestor. The legend displays the associated F-Value and Prob > F (p-value) resulting from the ANOVA for association between 100 seed weight and sieve size. 138

Chapter I: Review of the Literature

INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) is an important crop worldwide for both its protein (~40%) and oil content (~20%). It is produced in many countries but is grown most extensively in the United States, Brazil, Argentina, China, and India (Masuda and Goldsmith, 2009). *Glycine max* is considered to be one of the earliest examples of plant domestication for crop purposes with domestication thought to have occurred around ~ 5,500 years ago (Lee et al., 2011). However, estimates of divergence between wild and cultivated soybean subspecies are estimated to have occurred much earlier (around 0.8 million years ago) based on genomic analysis, suggesting natural divergent selection (Li et al., 2014).

GENETIC DIVERSITY IN SOYBEAN

Limited Diversity Exists in Soybean Breeding Programs

Three main bottlenecks have occurred in the history of North American soybean: original domestication, selection of Asian landraces for introduction, and many years of selective breeding. It has been estimated that ~81% of rare alleles have been lost from wild sources, some of which might be useful in future crop improvement efforts (Hyten et al., 2006). Selection of domesticated soybean has included adaptation to different latitudes, disease resistance, and yields ten times greater than its closely related wild relative, *G. soja*. Soybean breeding procedures in North America have consisted mainly of biparental crosses between high-yielding cultivars, which have created the third genetic bottleneck (Carter et al., 2004). In addition, 80% of varieties now being produced from private germplasm has led to fewer sources of genetic variability being utilized (Mikel et al., 2010). The results are that

the base of North American soybean varieties comes almost entirely from 80 ancestors, and 84% of the genetic base can be traced back to 17 lines (Gizlice et al., 1994).

Despite the narrow genetic base, improvements are still being observed for yield and seed composition indicating that favorable diversity still exists in soybean (Hyten et al., 2006); however, it is feared that these gains will reach plateaus without the introduction of more genetic diversity (Egli, 2008). With additional disease and environmental stressors likely in the future, genetic improvement must be maintained.

Use of *Glycine soja* in soybean breeding:

Wild ancestors have proven to be a source of improvement among many plant species. For example, in 1941, tomatoes (*Solanum lycopersicum*) were released which had been deliberately crossed with a wild ancestor to provide *Fusarium* resistance. Tomatoes are now one of the best examples of utilizing exotic species for crop improvement (Rick and Chetelat, 1995). Other examples of crosses with wild species for both abiotic and biotic stresses as well as other important traits can be found in sugar cane (*Saccharum officinarum* L.), potatoes (*Solanum tuberosum*), wheat (*Triticum aestivum*), and many others (Stalker, 1980; Ross, 1986; Hodgkin and Hajjar, 2008).

Glycine soja Sieb & Zucc, an annual weedy plant found in various regions of Asia, is believed to be the ancestor of domesticated soybean (Hadley and Hymowitz, 1973). The seeds are rich in protein (31.1-52.4%) but poor in oil (9.1-11.9%) (Hymowitz et al., 1972). *G. max* and *G. soja* both contain $2 \times 20 = 40$ chromosomes and can be crossed to form fertile offspring (Acquaah, 2012; Singh and Hymowitz, 1988). While chromosomally similar, much greater DNA marker diversity exists in *G. soja* compared to cultivated *G. max* because *G*

soja has not undergone the bottlenecks of domestication and crop breeding imposed on domesticated soybean (Singh and Hymowitz, 1999). The variation within the wild ancestor was assessed using accessions from different areas of Japan, Korea, and China. All studies concluded that not only was there a great amount of diversity present in each geographic region, but the variation was also great between regions (Dong, 2001; Wang, 2007; Yu, 1993). A newer study using *de novo* pan-genomic assembly from 7 geographically separate accessions of *G. soja* indicated that assembly of any two of those genomes sequenced represent an average of 78.2% of coverage for all lines sequenced. Therefore, a single reference genome cannot capture the diversity present within the wild ancestor, and a larger amount of diversity exists than previously thought. This study also showed a range of genome sizes from 889.33 – 1,118.34 Mbp or 93.6-117.7% of the soybean reference genome, Williams82. However, the overlap of 94% of gene models annotated for the reference line indicate that this diversity is largely due to mutations such as single nucleotide polymorphisms (SNPs), presence/absence variants (PAVs) and copy number variants (CNVs) (Li et al., 2014). These variants could serve to improve soybean cultivars while maintaining essential soybean genetic architecture.

While great diversity exists within *Glycine soja*, it is also plagued by undesirable traits. The wild progenitor is often characterized by vining, lodging susceptibility, lack of leaf abscission at maturity, shattering and small, hard black seeds (Carpenter and Fehr, 1986). Qualitative traits such as disease resistance may be more easily incorporated using phenotypic selection, but quantitative traits present a greater challenge. Tanksley and McCouch (1997) describe the previous challenges and underutilization of incorporating

exotic germplasm into elite lines of multiple plant species as well as the possibilities that have arisen with new molecular techniques to identify exotic germplasm with favorable genes not obvious in the phenotype.

Some useful genes within *Glycine soja* have already been identified for potential use in soybean improvement. Diers et al. (1992) identified genes responsible for increased seed protein. Li et al. (2008) identified two quantitative trait loci (QTL) for yield from a *Glycine soja* accession from the USDA germplasm bank. Resistance traits to both abiotic stresses such as salt stress and biotic stresses like soybean cyst nematode (*Heterodera glycines* Ichinohe) have also been identified in the wild ancestor (Wang et al., 2001; Luo et al., 2005). With an estimated 8,500 accessions of *Glycine soja* in world germplasm banks and ~1,100 maintained by the USDA, many other useful genes might exist as well (Carter et al., 2004).

As part of a grant aimed at increasing soybean genetic diversity funded by the United Soybean Board, 806 *G. soja* accessions from the USDA Collection were identified as being >.01% genetically different than any other accession in the collection. Among these accessions, genetic clusters were created from which a 10% maximized genetic diversity Core Collection (81 accessions) was chosen by Dr. Qiagin Song (personal comm). This “diversity core” has been used to make crosses with a limited number of different *G. max* cultivars to introgress valuable genetic variation into new germplasm resources.

Previous attempts have been made at incorporating *Glycine soja* into the soybean gene pool. Ertl and Fehr (1985) performed a backcross method using two separate *G. soja* accessions as the non-recurrent parents with one of two high-yielding *G. max* recurrent parents. It was determined that a minimum of two backcrosses was needed in order to obtain

a plant containing most of the desirable agronomic characteristics of a soybean variety. It was also discovered that the plant continued to improve through the BC₄ generation, but these plants contained very little of the *G. soja* genome. The study concluded that by using this method up to 12.5% of the wild genome could be incorporated into a favorable soybean variety. This method has also been used in work with introgression of seed size for use in obtaining small seeded varieties for food use (Leroy et al., 1991; Cui et al., 2004).

While backcross methods seem to be the most effective way to obtain a desirable phenotype, it can limit the passing of quantitative traits from donor parents as well as reduce the amount of diversity retained. The advanced QTL-backcross method developed by Tanksley and Nelson (1996) is not effective at retaining traits with multiple loci that interact epistatically. Therefore, Carter and Delheimer (2012) developed alternative methods. These efforts have relied mainly upon mass selection of “mega-populations” of progeny developed from biparental crosses at the F₃ and later generations based on desirable phenotypic characteristics. Selection is based mainly on erectness and shattering as well other *G. max*-like characteristics. It was found that up to 40% of alleles from *G. soja* could be retained in an individual while still obtaining a plant with favorable traits and up to 86% of the yield from the recurrent *G. max* parent. Among as few as 16 individuals selected from these large populations, 99% of the wild soybean SNPs could be recovered in agronomically “adapted” plants while still maintaining 70% of the yield of *G. max* checks. While space intensive, this method combined with modern genomic tools allows for successful introduction of large amounts of genetic variability and capture of most of the *G. soja* genome in plants suitable to

be added to a public or private breeding program to address challenges in soybean production such as emerging diseases and pests, yield plateaus, and climate change.

ABIOTIC STRESS

Temperature Effects on Soybean Germination

The vast majority of the scientific community has come to a consensus that global temperature patterns are continually increasing (Cook et al., 2013). While different models of climate change and their effects on agriculture currently exist (Mendelsohn, 1994; Schlenker, 2006), it is estimated that areas such as the Midwestern corn belt and areas of California that depend largely on agriculture will be the most heavily impacted by temperature increases. Recent suggestions by Schlenker et al., (2009) indicate that corn, soybean, and cotton yields increase at a linear rate from colder temperatures until reaching an optimum growing temperature. However, the rate of yield damage increases more rapidly in a nonlinear trend at temperatures above the optimum, resulting in greater crop losses. Soil temperature is also a factor in early planting schemes common in the Mid-Southern US where soybean is planted in cool soils in a process known as Early Soybean Production System (ESPS) (Heatherly, 1999).

Germination experiments on *Glycine max* have shown the optimal germination temperature to be between 28-30 °C, while germination stops around 10 °C with moisture being a major contributing factor (Heath, 1987; Hatfield and Egli, 1974). Wuebker et al. used this temperature range to investigate temperature effect in combination with flooding and confirmed the estimates (2001). It was also shown by Alm et al. (1993) that seedling

elongation rate increased steadily in both corn and soybean as temperatures increased from 10 - 25 °C.

Koti (2005) conducted physiological stress studies to characterize the response of soybean when exposed to elevated temperatures at both flowering and pod development. This work revealed lower levels of productivity at multiple reproductive points when exposed to a temperature of 38 °C. Similar studies of heat stress on soybean at flowering and seed development showed decreased vigor as well. In a heat-sensitive soybean variety developed for ESPS, germination of seed developed under a daytime temperature of 36 °C showed an average 50% reduction in percent germination, while a heat-resistant variety showed no reduction. However at 42 °C, heat-resistant soybean was severely stunted, while heat-sensitive showed no germination at all (Chebrolu et al., 2016). While studies in cultivated soybean demonstrate effects caused by suboptimal temperatures, characterization of germination in wild soybean at these temperatures is not well documented.

Increasing Ozone Levels Affect Plant Growth

Ozone (O₃) is a secondary air pollutant that has been of increasing concern with rises in greenhouse gases. The majority of tropospheric ozone comes from photochemical reactions of methane (CH₄), Volatile Organic Compounds (VOCs) and mono-nitrogen oxide species (NO_x) from human pollution (Stevenson et al., 2006). O₃ levels have risen from < 10 parts per billion (ppb) during the late 1800s (Volz and Kley, 1988) to daytime levels of nearly 40 ppb in many areas of the Northern Hemisphere (Stevenson et al., 2006). These levels are expected to continue to rise along with human population levels and technological advances (Ainsworth et al., 2012).

Effects of these increased levels are seen not only directly on human health but also plant productivity. The damaging effects on carbon assimilation, stomatal conductance, and plant growth result in losses to crop yield (Ainsworth et al., 2008; Ashmore et al., 2005; Feng et al., 2008; Fiscus et al. 2005). Losses in total crop global production due to ozone damage have been estimated to be from \$14 billion to \$26 billion (Van Dingenen, 2009).

The rate of ozone uptake by leaf tissue and leaf tolerance to reactive oxygen species (ROS) created by ozone are control points for damage caused by ozone (Eller and Sparks, 2006). Movement of O₃ into intercellular membrane space is controlled by the aperture of stomata. Inside leaves, ozone reacts in the apoplast with many different molecules to produce different harmful products like OH[•] radicals and hydrogen peroxide (Grimes et al., 1983; Heath, 1987), making the capacity of the apoplast to handle these products the primary defense mechanism of the plant (Moldau, 1998). After exposure to levels of >150 ppb, plants respond by release of ROS, hormones, calcium ions, and mitogen-activated protein kinase (MAPK) signaling cascades. The ozone response pathway shares many similarities to that of programmed cell death often seen in response to pathogens. Both processes involve rise in ROS production, which further activates the ethylene, salicylic acid, and jasmonic acid pathways for expression of defense genes (Ainsworth, 2012).

Chronic exposure to ozone has been shown to cause decreases in photosynthesis and plant mass as well as early leaf senescence (Fiscus et al., 2005). Exposure to acute treatments with ozone has been observed for different species. *Arabidopsis* displayed a rapid temporary decrease in stomatal conductance with a rise in ROS followed by a slower recovery to normal levels of conductance; however, this was not thought to affect overall photosynthetic

activity (Kollist et al., 2007). Chronic exposure causes irreversible lower stomatal conductance, which can directly damage plant productivity (Witting et al., 2007). Stomatal sensitivity to abscisic acid may also be affected by O₃ stress causing transpiration levels to increase when exposed to drought (Mills et al. 2009).

It is well known that chronic O₃ also decreases rates of CO₂ assimilation in leaves (Fiscus, 2005). Meta-analyses of crop species show that light-saturated photosynthesis significantly decreased for soybean (*Glycine max*), wheat (*Triticum aestivum*), and rice (*Oryza sativa*) (Morgan et al. 2003, Feng, 2008; Ainsworth, 2008) in response to elevated ozone. Consistent with this ozone response were decreased levels of nonstructural carbohydrates for growth such as sucrose and starch. In addition to less fixation of CO₂, plants in elevated O₃ show higher rates of cellular respiration causing further reduction in productivity (Gillespie et al., 2011).

Current estimates of ozone damage show a reduction in yield of around 6-16% for soybean (Van Dingenen et al., 2009). Studies have been done on *Glycine max* in recent years to assess the damage caused by both current ambient ozone (50-62 ppb) and expected elevated ozone by 2050 (63-75 ppb) using O₃ Free Air CO₂ Enrichment (FACE) technology. These studies showed a reduction of 15-25% in yield largely due to smaller seed size and a reduction of 17% in the number of pods/plant (Morgan et al., 2006; Morgan et al., 2004; Dermody et al., 2008; Betzelberger et al., 2010).

Burkey and Carter (2009) conducted a screen of 30 North American ancestral lines of domesticated soybean to determine if resistance exists in the background of present cultivars. A ten-fold difference in response to elevated ozone was observed among these lines

indicating development of resistant soybean varieties is possible. Phenotypic ratings were assigned according to the system developed by Heagle (1979). Because it has been noted that mature trifoliates lower on the plant tend to display more damage, each exposed trifoliolate on *G. max* checks was given a percentage designation for area that appears to show stippling (maroon-colored necrotic spotting) characteristic of ozone damage. An overall score from 1-4 was assigned to each plant (*G. max*) with 1 indicating 0-5% damage, 2 indicating 5-25% damage, 3 indicating 25-50%, and 4 indicating 50-100% damage. That is, plants assigned a score of 1 are more resistant, while plants with a score of 4 are sensitive.

More in-depth studies on soybean response to ozone damage have been done in recent years. Whaley et al. (2015) conducted a gene expression analysis comparing a low (27 ppb) and high (75 ppb) ozone treatment of a resistant *G. max* cultivar, Fiskeby III, and a sensitive cultivar, Mandarin (Ottawa). This study shows that cultivated soybean exhibits differential expression among expected ROS and defense pathways, but there are cultivar-specific genes between resistant and sensitive cultivars. In particular, it seems that metabolism genes were up regulated in resistant soybean, while expression of many genes varied greatly at different time points for the sensitive cultivar. This study also found a possible link between leaf architecture and ozone response as demonstrated by the regulation in expression of wax and cutin biosynthesis genes of the sensitive variety.

Another study by Burton et al. (2016) created a mapping population between the same two soybean accessions. Two key findings resulted from this study. Firstly, it was determined that ozone damage appears to be more visible in older leaf tissue, and that damage decreases corresponding to younger tissue. It also determined that ozone damage has

a multigenic inheritance pattern, and significant quantitative trait loci (QTL) were identified on Chromosome 17 for the oldest leaf positions, Chromosome 4 for the next oldest, and additional loci on Chromosomes 6, 18, 19, and 20. Of these loci, only one resistance allele was donated from Mandarin (Ottawa), the sensitive parent.

GENE EXPRESSION METHODOLOGY

There have been several methods developed in recent years to quantify transcripts in a cell in response to different physiological states. These methods include quantitative reverse transcription followed by polymerase chain reaction (qRT-PCR) (Kubista et al., 2006), microarray analysis for specific isoforms (Clark, 2002), and high-density microarrays (Cheng, 2005). These methods are relatively inexpensive and high throughput but generally rely on a fluorescently-labeled cDNA with some form of prior knowledge of the genes for which expression is expected and often show large levels of background noise from nearby hybridizations. In addition, comparison across experiments can also be difficult and requires complex normalization strategies (Wang et al., 2009).

Sequencing-based approaches have also been developed that directly identify and quantify cDNA sequences derived from mRNA. Low throughput methods like Sanger sequencing of expressed sequence tag libraries were early approaches that were expensive and not quantitative in nature (Boguski et al., 1994). Other methods like serial analysis of gene expression (SAGE) (Velculescu, 1995), cap analysis of gene expression (CAGE) (Kodzius 2006), and massively parallel signature sequencing (MPSS) (Brenner, 2000) utilize a tag-based expression approach to sequencing using traditional Sanger methods. Because of the reliance on this expensive method of sequencing, short sequences are produced and can

often not be uniquely mapped to a reference genome. These methods showed the need for an improved method of sequencing for this approach to be effective (Wang et al., 2009).

RNA-Seq is a newer technology that seeks to overcome challenges in profiling gene expression presented by older technologies by using next generation sequencing technologies to deeply sequence the transcriptome. It has been refined in several species with reference genomes such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, and *Glycine max* (Nagalakshmi, 2008; Wilhelm, B, 2008; Lister, 2008; Schmutz et al., 2010). RNA-Seq methods have also been developed for organisms without reference genomes though the analyses are more difficult.

RNA-Seq utilizes recently developed deep-sequencing chemistries. First, a sample of mRNA is reverse-transcribed into a library of more stable cDNA copies containing adapters and sample identifiers to either one of both ends of the strand. The molecules are then sequenced in a high-throughput system, sometimes amplification based, and sometimes not; reads obtained are generally 30-400 bp. Systems for this sequencing come from companies such as Illumina, Applied Biosystems, and Roche, while new competitors are entering the field. Different platforms offer unique advantages and may be useful depending on the purpose of the experiment (Wang et al., 2009).

Advantages of RNA-Seq, though still being discovered, are independence from *a priori* sequence information needed in hybrid-based approaches, capability of producing longer sequence reads that may be uniquely mapped to a reference genome in contrast to the short reads produced from dideoxynucleotide termination methods (traditional sequencing), as well as the ability to determine start and stop points of transcription leading to isoform

discovery. There is also very little background noise as compared to DNA microarrays because of the unique mapping ability of these regions providing more reliable quantification data (Wang et al., 2009).

Though a robust technology, RNA-Seq does offer challenges. Many of these pitfalls occur during library preparation, library amplification or the sequence processing. Small RNAs can be sequenced directly, but larger molecules (> 500 bp) must be shortened in a couple of different ways – either as RNA or cDNA. Each of these approaches introduces their own set of biases. RNA methods (most commonly nebulization or hydrolysis) create a good representation of inner molecules but tend to lose the ends of molecules (Mortazavi, 2008), while cDNA (sonication or DNase I treatment) tends to have good representation of 3' ends but may underrepresent internal molecules (Nagalakshmi, 2008). In addition, these newer sequencing methods often rely on amplification also may create bias from PCR artifacts rather than true greater gene expression (Wang et al., 2009). PCR biases can be reduced by identifying sequence reads with exactly the same ends and statistically correcting for the PCR bias.

As with all forms of transcriptome analysis, RNA-Seq brings with it its own set of informatics concerns. The most obvious of these is the amount of raw data that must be processed, which necessitates larger and faster computing systems and more memory-efficient algorithms to parse and analyze the data. Secondly, the splicing of introns and other RNA processing complicates analyses of expression in eukaryotes. Programs exist that can adapt to the uniqueness of multiple experimental and organismal concerns. Further complicating this is the tendency of larger genomes to contain repeated regions on which

short reads may align at multiple positions. Many open source programs have been created that deal with the complexity of aligning transcript sequences to eukaryotic genomes (Wang et al., 2009).

While there are considerations that must be taken into account, RNA-Seq offers the first high-throughput, relatively inexpensive method to quantify overall transcriptome expression of an organism. Depending on the aim of each study, several procedures exist to enrich for certain types of molecules and an even greater number of computing processes exist for the analysis of such datasets.

REFERENCES

- Acquaah, G. (2012). Breeding soybean. *Principles of Plant Genetics and Breeding* (pp. 629-638). John Wiley & Sons, Ltd.
- Ainsworth, E. A., Yendrek, C. R., Sitch, S., Collins, W. J., & Emberson, L. D. (2012). The effects of tropospheric ozone on net primary productivity and implications for climate change. *Annual Review of Plant Biology*, *63*, 637.
- Ainsworth, E. A. (2008). Rice production in a changing climate: A meta-analysis of responses to elevated carbon dioxide and elevated ozone concentration. *Global Change Biology*, *14*(7), 1642-1650.
- Alm, D. M., Stoller, E. W., & Wax, L. M. (1993). An index model for predicting seed germination and emergence rates. *Weed Technology*, *7*(3), 560-569.
- Ashmore, M. R. (2005). Assessing the future global impacts of ozone on vegetation. *Plant, Cell & Environment*, *28*(8), 949-964.
- Betzberger, A. M., Gillespie, K. M., McGrath, J. M., Koester, R. P., Nelson, R. L., & Ainsworth, E. A. (2010). Effects of chronic elevated ozone concentration on antioxidant capacity, photosynthesis and seed yield of 10 soybean cultivars. *Plant, Cell & Environment*, *33*(9), 1569-1581.
- Boguski, M. S., Tolstoshev, C. M., & Bassett, D. E., Jr. (1994). Gene discovery in dbEST. *Science (New York, N.Y.)*, *265*(5181), 1993-1994.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., et al. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*, *18*(6), 630-634.

- Burkey, K. O., & Carter Jr., T. E. (2009). Foliar resistance to ozone injury in the genetic base of U.S. and Canadian soybean and prediction of resistance in descendent cultivars using coefficient of parentage. *Field Crops Research*, 111(3), 207-217.
- Burton, A. L., Burkey, K. O., Carter, T. E., Orf, J., & Cregan, P. B. (2016). Phenotypic variation and identification of quantitative trait loci for ozone tolerance in a Fiskeby III x Mandarin (Ottawa) soybean population. *Theoretical and Applied Genetics*. 1-13.
- Carpenter, J., & Fehr, W. (1986). Genetic-variability for desirable agronomic traits in populations containing *Glycine soja* germplasm. *Crop Science*, 26(4), 681-686.
- Carter, T. E., Nelson, R. L., Sneller, C. H., & Lee, Z. L. (2004). Genetic diversity in soybean. In H. R. Boerma, & J. E. Specht (Eds.), *Soybeans: Improvement, production, and use* (3rd ed., pp. 303-304-416). Madison, WI, USA: ASA - CSSA - SSSA.
- Chebrolu, K. K., Fritschi, F. B., Ye, S., Krishnan, H. B., Smith, J. R., & Gillman, J. D. (2016). Impact of heat stress during seed development on soybean seed metabolome. *Metabolomics*, 12(2), 1-14.
- Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science (New York, N.Y.)*, 308(5725), 1149-1154.
- Clark, T. A., Sugnet, C. W., & Ares, M., Jr. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science (New York, N.Y.)*, 296(5569), 907-910.

- Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., et al. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, 8(2), 024024.
- Cui, Z., Wilson, R. F., Carter, T. E., James, A. T., & Miyazaki, S. (2004). Breeding specialty soybeans for traditional and new soyfoods. *Soybeans as functional foods and ingredients* (pp. 264-265-322). Champaign, IL: AOCS Publishing.
- Delheimer, J. C. (2012). *Agronomic performance of breeding lines developed from a single cross of Glycine max X Glycine soja*. Unpublished Doctoral dissertation, North Carolina State University.
- Dermody, O., Long, S. P., McConnaughay, K., & DeLucia, E. H. (2008). How do elevated CO₂ and O₃ affect the interception and utilization of radiation by a soybean canopy? *Global Change Biology*, 14(3), 556-564.
- Diers, B., Keim, P., Fehr, W., & Shoemaker, R. (1992). RFLP analysis of soybean seed protein and oil content. *Theoretical and Applied Genetics*, 83(5), 608-612.
- Dong, Y., Zhuang, B., Zhao, L., Sun, H., & He, M. (2001). The genetic diversity of annual wild soybeans grown in China. *Theoretical and Applied Genetics*, 103(1), 98-103.
- Egli, D. B. (2008). Comparison of corn and soybean yields in the United States: Historical trends and future prospects. *Agronomy Journal*, 100(3), S79-S88.
- Eller, A. S. D., & Sparks, J. P. (2006). Predicting leaf-level fluxes of O₃ and NO₂: The relative roles of diffusion and biochemical processes. *Plant, Cell & Environment*, 29(9), 1742-1750.

- Ertl, D. S., & Fehr, W. R. (1985). Agronomic performance of soybean genotypes from *Glycine max* X *Glycine soja* crosses. *Crop Science*, 25(4), 589-592.
- Feng, Z., Kobayashi, K., & Ainsworth, E. A. (2008). Impact of elevated ozone concentration on growth, physiology, and yield of wheat (*Triticum aestivum* L.): A meta-analysis. *Global Change Biology*, 14(11), 2696-2708.
- Fiscus, E. L., Booker, F. L., & Burkey, K. O. (2005). Crop responses to ozone: Uptake, modes of action, carbon assimilation and partitioning. *Plant, Cell & Environment*, 28(8), 997-1011.
- Gillespie, K. M., Xu, F., Richter, K. T., McGrath, J. M., Markelz, R. J. C., Ort, D. R., et al. (2012). Greater antioxidant and respiratory metabolism in field-grown soybean exposed to elevated O₃ under both ambient and elevated CO₂. *Plant, Cell & Environment*, 35(1), 169-184.
- Gizlice, Z., Carter, T., & Burton, J. (1994). Genetic base for North-American public soybean cultivars released between 1947 and 1988. *Crop Science*, 34(5), 1143-1151.
- Grimes, H. D., Perkins, K. K., & Boss, W. F. (1983). Ozone degrades into hydroxyl radical under physiological conditions: A spin trapping study. *Plant Physiology*, 72(4), 1016-1020.
- Hadley, H. H., & Hymowitz, T. (1973). *Speciation and cytogenetics*. In: Cald Howell RW, Judd RW, and Johnson HW (eds.) *Soybeans: Improvement, Production, and Uses* (American Society of Agronomy: Madison, WI, USA), pp. 97- 116.
- Hatfield, J. L., and D. B. Egli. (1974). Effect of temperature on the rate of soybean hypocotyl elongation and field emergence¹. *Crop Science*. 14:423-426.

- Heagle, A. S. (1979). Ranking of soybean cultivars for resistance to ozone using different ozone doses and response measures. *Environmental Pollution (1970)*, 19(1), 1-10.
- Heath, R. (1987). The biochemistry of ozone attack on the plasma membrane of plant cells. In J. Saunders, L. Kosak-Channing & E. Conn (Eds.), (pp. 29-54) Springer US.
- Heatherly, L. G., & Spurlock, S. R. (1999). Yield and economics of traditional and early soybean production system (ESPS) seedlings in the midsouthern United States. *Field Crops Research*, 63(1), 35-45.
- Hodgkin, T., & Hajjar, R. (2008). Using crop wild relatives for crop improvement: Trends and perspectives. In Maxted, N, Ford Lloyd, BV Kell, SP Iriondo, JM Dulloo, E Turok, J. (Ed.), (pp. 535-548). Cabi Publishing, Walingford, Oxon, UK.
- Hymowitz, T., Hadley, H., & Palmer, R. (1972). Seed weight, protein, oil and fatty-acid relationships within genus *Glycine*. *Tropical Agriculture*, 49(3), 245-&.
- Hyten, D. L., Song, Q., Zhu, Y., Ik-Young Choi, Nelson, R. L., Costa, J. M., et al. (2006). Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(45), 16656-16671.
- Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., et al. (2006). CAGE: Cap analysis of gene expression. *Nature Methods*, 3(3), 211-222.
- Kollist, T., Moldau, H., Rasulov, B., Oja, V., Rämme, H., Hüve, K., et al. (2007). A novel device detects a rapid ozone-induced transient stomatal closure in intact *Arabidopsis* and its absence in *Abi2* mutant. *Physiologia Plantarum*, 129(4), 796-803.
- Koti, S., Reddy, K. R., Reddy, V. R., Kakani, V. G., & Zhao, D. (2005). Interactive effects of carbon dioxide, temperature, and ultraviolet-B radiation on soybean (*Glycine max* L.)

flower and pollen morphology, pollen production, germination, and tube lengths.

Journal of Experimental Botany, 56(412), 725-736.

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., et al. (2006).

The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, 27(2–3), 95-125.

Kusumi, K., Hirotsuka, S., Kumamaru, T., & Iba, K. (2012). Increased leaf photosynthesis

caused by elevated stomatal conductance in a rice mutant deficient in SLAC1, a guard cell anion channel protein. *Journal of Experimental Botany*.

Lee, G., Crawford, G. W., Liu, L., Sasaki, Y., & Chen, X. (2011). Archaeological soybean

(*Glycine max*) in East Asia: Does size matter? *PloS One*, 6(11), e26720.

Leroy, A., Fehr, W., & Cianzio, S. (1991). Introgression of genes for small seed size from

Glycine soja into *G. max*. *Crop Science*, 31(3), 693-697.

Li, D., Pfeiffer, T. W., & Cornelius, P. L. (2008). Soybean QTL for yield and yield

components associated with *Glycine soja* alleles. *Crop Science*, 48(2), 571-581.

Li, Y., Zhou, G., Ma, J., Jiang, W., Jin, L., Zhang, Z., et al. (2014). *De novo* assembly of

soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat Biotech*, 32(10), 1045-1052.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., et

al. (2008). Highly integrated single-base resolution maps of the epigenome in

Arabidopsis. *Cell*, 133(3), 523-536.

- Luo, Q., Yu, B., & Liu, Y. (2005). Differential sensitivity to chloride and sodium ions in seedlings of *Glycine max* and *G. soja* under NaCl stress. *Journal of Plant Physiology*, *162*(9), 1003-1012.
- Masuda, T., & Goldsmith, P. D. (2009). World soybean production: Area harvested, yield, and long-term projections. *International Food and Agribusiness Management Review*, *12*(4), 143-162.
- Mendelsohn, R., Nordhaus, W. D., & Shaw, D. (1994). The impact of global warming on agriculture: A Ricardian analysis. *The American Economic Review*, *84*(4), 753-771.
- Mikel, M. A., Diers, B. W., Nelson, R. L., & Smith, H. H. (2010). Genetic diversity and agronomic improvement of North American soybean germplasm. *Crop Science*, *50*(4), 1219-1229.
- Mills, G., Hayes, F., Wilkinson, S., & Davies, W. J. (2009). Chronic exposure to increasing background ozone impairs stomatal functioning in grassland species. *Global Change Biology*, *15*(6), 1522-1533.
- Moldau, H. (1998). Hierarchy of ozone scavenging reactions in the plant cell wall. *Physiologia Plantarum*, *104*(4), 617-622.
- Morgan, P. B., Ainsworth, E. A., & Long, S. P. (2003). How does elevated ozone impact soybean? A meta-analysis of photosynthesis, growth and yield. *Plant, Cell & Environment*, *26*(8), 1317-1328.
- Morgan, P. B., Bernacchi, C. J., Ort, D. R., & Long, S. P. (2004). An *in vivo* analysis of the effect of season-long open-air elevation of ozone to anticipated 2050 levels on photosynthesis in soybean. *Plant Physiology*, *135*(4), 2348-2357.

- Morgan, P. B., Mies, T. A., Bollero, G. A., Nelson, R. L., & Long, S. P. (2006). Season-long elevation of ozone concentration to projected 2050 levels under fully open-air conditions substantially decreases the growth and production of soybean. *New Phytologist*, *170*(2), 333-343.
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Meth*, *5*(7), 621-628.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., et al. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science (New York, N.Y.)*, *320*(5881), 1344-1349.
- Rick, C., & Chetelat, R. (1995). Utilization of related wild species for tomato improvement. *First International Symposium on Solanacea for Fresh Market*, (412), 21-38.
- Ross, H. (1986). Advances in plant breeding 13: Potato breeding problems and perspectives. *Verlag Paul Parey: Berlin, West Germany.Illus.Paper*, , 132P.
- Schlenker, W., Hanemann, W. M., & Fisher, A. C. (2006). The impact of global warming on U.S. agriculture: An econometric analysis of optimal growing conditions. *The Review of Economics and Statistics*, *88*(1), 113-125.
- Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, *106*(37), 15594-15598.
- Singh, R. J., & Hymowitz, T. (1999). Soybean genetic resources and crop improvement. *Genome*, *42*(4), 605-616.

- Singh, R., & Hymowitz, T. (1988). The genomic relationship between *Glycine max* (L) Merr and *Glycine soja* Sieb and Zucc as revealed by pachytene chromosome analysis. *Theoretical and Applied Genetics*, 76(5), 705-711.
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, 463(7278), 178-183.
- Stalker, H. T. (1980). Utilization of wild species for crop improvement. *Brady, N.C. (Ed.). Advances in Agronomy, Vol.33.Xiii+374p.Academic Press, Inc.: New York, N.Y., USA; London, England.Illus.Maps, , P111-148.*
- Stevenson, D. S., Dentener, F. J., Schultz, M. G., Ellingsen, K., Van Noije, T., Wild, O., et al. (2006). Multimodel ensemble simulations of present-day and near-future tropospheric ozone. *Journal of Geophysical Research.D.Atmospheres*, 111(D8).
- Tanksley, S. D., & Nelson, J. C. (1996). Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics*, 92(2), 191-203.
- Tanksley, S. D., & McCouch, S. R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science*, 277(5329), 1063-1066.
- Van Dingenen, R., Dentener, F. J., Raes, F., Krol, M. C., Emberson, L., & Cofala, J. (2009). The global impact of ozone on agricultural crop yields under current and future air quality legislation. *Atmospheric Environment*, 43(3), 604-618.
- Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science (New York, N.Y.)*, 270(5235), 484-487.

- Volz, A., & Kley, D. (1988). Evaluation of the Montsouris series of ozone measurements made in the nineteenth century. *Nature*, 332(6161), 240.
- Wang, K., & Takahata, Y. (2007). A preliminary comparative evaluation of genetic diversity between Chinese and Japanese wild soybean (*Glycine soja*) germplasm pools using SSR markers. *Genetic Resources and Crop Evolution*, 54(1), 157-165.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-seq: A revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, 10(1), 57-63.
- Whaley, A., Sheridan, J., Safari, S., Burton, A., Burkey, K., & Schlueter, J. (2015). RNA-seq analysis reveals genetic response and tolerance mechanisms to ozone exposure in soybean. *BMC Genomics*, 16, 426-015-1637-7.
- Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., et al. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, 453(7199), 1239-1243.
- Wittig, V. E., Ainsworth, E. A., & Long, S. P. (2007). To what extent do current and projected increases in surface ozone affect photosynthesis and stomatal conductance of trees? A meta-analytic review of the last 3 decades of experiments. *Plant, Cell & Environment*, 30(9), 1150-1162.
- Wuebker, E. F., Mullen, R. E., & Koehler, K. (2001). Flooding and temperature effects on soybean germination journal paper no. J-18452 of the Iowa Agric. and Home Econ. Exp. Stn., Ames, IA, project no. 3244, and supported by Hatch Act and State of Iowa funds. *41*(6), 1857-1861.

Yu, H., & Kiang, Y. (1993). Genetic-variation in South-Korean natural-populations of wild soybean (*Glycine soja*). *Euphytica*, 68(3), 213-221.

**Chapter II: Evaluation of Germination at Sub-Optimal Temperatures among
Genetically Diverse *Glycine soja* Accessions**

ABSTRACT

To adapt to expected rising global temperatures or to produce a soybean variety suitable for early-season planting, producing breeding materials tolerant of germination temperatures above or below the ideal 28 – 30 °C is of interest. In this study, a diverse group of sixty-nine wild soybean accessions are characterized for their responses to germination at suboptimal temperatures. Measures of percent germination at both hot and cold temperatures indicated a range of responses in the accessions with PI 522211A having the highest least squares mean (LSMEAN) estimates at both hot and cold temperatures. Similarly, accessions differed for early root growth with PI 549048 having the longest estimates for root growth at both temperature extremes. Several accessions showed promise in both of these categories, and are therefore candidates as superior germinating wild soybean resources. This research may be of interest in producing genetically diverse soybean germplasm releases with improved early growth in response to suboptimal temperatures.

INTRODUCTION

The majority of the scientific community has reached a consensus that global temperatures are continuing to rise (Cook et al., 2013) and that the effects of this temperature change on agriculture are not fully known. Though different predictive models of climate change and their impacts on major crops currently exist (Mendelsohn, 1994; Schlenker, 2006), it is estimated that areas such as the Midwestern corn belt and areas of California that depend largely on agriculture will be the most heavily hit by temperature increases. It has been recently suggested by Schlenker et al., (2009) that corn, soybean, and cotton yields increase at a linear rate from colder temperatures until reaching an optimum growing

temperature. However, the rate of yield damage increases more rapidly at temperatures above the optimum, resulting in greater crop losses. Soil temperature may also be a factor in early planting schemes common in the Mississippi delta where soybean is bred to be planted in cool soils (Heatherly, 1999). Early planting may spread to other areas of the United States as the temperatures are expected to rise and might expose early-planted crops to cold snaps. Identifying sources of genetic resistance to temperature extremes to incorporate into economically important crops could help secure the food supply. With climate change becoming a concern as well as the interest in utilizing earlier planting dates to avoid harsh conditions expected later in the growing season, the ability of crop varieties to germinate in suboptimal temperature has value for incorporation into breeding stock.

Experiments on *Glycine max* have shown the optimal germination temperature to be somewhere between 28-30 °C, while germination stops around 10 °C as long as adequate moisture is present (Hatfield and Egli, 1974). Wuebker et al. used this temperature range to investigate temperature effect in combination with flooding and confirmed the estimates (2001). It was also shown by Alm et al. (1993) that seedling elongation rate increased steadily in both corn and soybean as temperatures increased from 10 – 25 °C.

Physiological studies at both flower and pod development of soybean at elevated temperatures by Koti (2005) revealed reduced floral development, pollen production, and pollen tube length at 38 °C. Further phenotypic studies of heat stress on soybean at flowering and seed development showed decreased vigor as well. In a heat-sensitive soybean variety developed for ESPS, germination of seed developed under a daytime temperature of 36 °C showed an average 50% reduction in percent germination, while a heat-resistant variety

showed no reduction. However at 42 °C, heat-resistant soybean germination was severely limited, and heat-sensitive showed no germination at all (Chebrolu et al., 2016). While studies in cultivated soybean demonstrate effects caused by suboptimal temperatures, characterization of germination in wild soybean at these temperatures is not well documented.

In this study, we identify sources of early germination vigor and tolerance to temperatures both colder and hotter than those expected to be ideal for germination in soybean. We utilize a core set of wild soybean (*Glycine soja* Siebold & Zuccarini) from the USDA germplasm bank representing maximum genetic diversity to screen for improved temperature tolerance for germination and seed vigor. Identification of valuable sources of germination potential could provide sources for improved breeding of temperature-tolerant soybean breeding lines.

MATERIALS AND METHODS

Germination of *Glycine soja* at Sub-Optimal Temperatures

Seed from increases of Maturity Groups III-VII from the Core Collection grown in Clayton, NC in 2013 and 2014 were available to study the effects of temperature on seed germination. The increases were planted as randomized complete blocks by maturity groups (MG): MG III & IV were in one block (Source 1); MG V contained the most accessions and was in one block (Source 2); and MGs VI and VII (Source 3) constituted the final block. Each of these increases contained three randomized block replicates from two locations to account for environmental effects. This blocking information was furthered into the experimental design of the germination and is referred to as “Seed Source.”

Seed used for the experiment was sterilized by one of two methods. Seed tested at colder temperatures was immersed for 5 minutes in a 10% bleach solution followed by a thorough rinse with sterile water and a 30 second dip in 100% ethanol and dried. Seed tested at hotter temperatures was treated under a fume hood with 4 hour chlorine gas exposure by placing the seed in a desiccator adjacent to a beaker containing 100 mL of bleach and an added 3 mL of concentrated HCl. Treated seed was put into individual 15 mL tubes by genotype and replication within each experiment along with 2 ceramic beads and ~1.25 g Garnet Matrix A (MP Biomedicals, Germany). The tubes were shaken in a 2010 Geno/Grinder (SPEX SamplePrep, NJ) at 1640 RPM for increments of 10 minutes to allow for scarification. Scarified seeds were then tested for imbibition by placing 10 seeds in a small amount of water for 3 hours. If each sample demonstrated imbibition of at least 80% of seeds, the sample was deemed adequately scarified. If a lower percentage was seen for a seed sample, it was deemed as likely hard-seeded and was shaken for another 10 minutes and retested. This was repeated until all samples reached at least 80% imbibition. Remaining hard seed in rag dolls were discarded prior to germination counts.

The seed were then ready for assembly into “Rag-doll” germination set-ups. This was accomplished by cutting a sheet of large germination paper in half and spreading ~25 scarified seeds of each replication of a genotype evenly across at approximately 4 cm from the top of the paper. The other half of the cut paper was then placed on top of the seed and wetted with water until saturated. The papers were rolled and secured with a large rubber band. This was repeated four times for exposure to 4 testing temperatures for each genotype. Following assembly, each Rag doll was placed in a clear plastic germination box

corresponding to one of 12 boxes for each of 3 maturity group blocks of seed source to be placed in 4 different temperatures.

Based on the optimal germination temperature of *Glycine max* (28 °C) and the basal minimum temperature recorded for germination (10 °C), the effects of 4 temperatures at either extreme on germination of wild soybean were tested. The colder temperatures were chosen in increments of 6-7 °C below 28 °C (11°C, 17 °C, 22 °C, 28 °C); the hotter temperatures were chosen in 5 °C increments based on pilot experiments to determine the appropriate range (28 °C, 33 °C, 38 °C, 43 °C). These temperatures were applied to four different growth chambers. One box containing Rag-dolls blocked by seed source, comprised of maturity groupings as defined previously in the production of seed for the experiment, was placed in each chamber and allowed to germinate in a controlled dark environment for 7 days for cold trials and 4 days for heat trials while maintaining constant moisture. The different durations were chosen based on pilot experiments to determine when root differentiation began to occur but before seed experienced significant fungal pressure. Three boxes, each containing a different seed source of maturity groups, were placed in each chamber. Cold treatments were run in September and October 2015. Hot treatments were run in January and February 2016.

Following this period, the number of seeds germinated out of the total planted was calculated as a percentage. For those samples containing seeds with measurable root length, a maximum of 10 length measurements were recorded for a sampling of roots starting from the point of cellular differentiation. In an effort to control for measurement error, samples were

randomly assigned to each of 3 people measuring root length prior to each measurement period.

Analysis of Germination Data

The samples were arranged in a split plot design with samples blocked by seed source in each complete plot (germination box). The Data from 2 replications of each year (2013 and 2014) were recorded for both heat and cold stress run at separate time periods. The data was analyzed using SAS 9.4 (SAS Institute, NC) with the MIXED procedure in order to determine differences in means between genotypes, to identify those that perform better under suboptimal germination temperatures, and to test for genotype and treatment interactions. Two separate analyses for each set of trials were performed using percent germination and average root length as response variables.

Models were fit treating temperature and genotype as fixed effects, while seed source, year, rep, and corresponding interaction terms were treated as random effects. To deal with a degree of missing data points, Kenward-Roger denominator degrees of freedom were used to approximate fixed effects' inference (Kenward and Roger, 1997). A Tukey-Kramer adjustment was also applied to control the familywise error rate of the many comparisons being made (Tukey, 1953; Kramer, 1956). The SGPLOT procedure of SAS 9.4 (SAS Institute, NC) was used to create needle plots of the genotypic least squares estimates of each genotype grouped by seed source.

RESULTS

Evaluation of Percent Germination at Extreme Cold Temperatures

The effect of cold temperature on germination of seed from the *Glycine soja* diversity core collection was assessed. Selected lines were germinated in ragdolls at 11, 17, 22, and 28 °C to represent a range between the proposed minimum and ideal germination temperatures for soybean. In the evaluation of the genetic materials from 2 replicates of 2 years of seed germinated at lower temperatures, mean percent germination over all samples was obtained for each temperature (11 °C to 28 °C) (**Table 2.1**). In the mixed model analysis, genotype and temperature were significant ($p < 0.05$), but not the temperature and genotype interaction (**Table 2.2**). Due to similarity in percent germination at 17, 22, and 28 °C vs. 11 °C, these treatments groups were partitioned into an orthogonal contrast for analysis. The 11 °C treatment had 20 percentage units lower germination than the mean of the 17, 22, and 28 °C treatments ($p < 0.05$). In addition, the genotype by temperature interaction as 11°C run as a contrast against the other three temperatures showed significance ($p < 0.05$) (**Appendix A**). PI 522211A (MG III) had the highest percent germination least squares estimate at 107% with a standard error of 13%, while PI 479751 (MG III) had the lowest least squares estimate at 39% and a standard error of 13.07% (**Table 2.3**). PI 522211A had a germination that was statistically superior ($p < 0.05$) to 21 of the other 68 genotypes tested. Of these, 14 were MG III or IV accessions (**Figure 2.1**), 2 were MG V (**Figure 2.2**), and 5 were MG VI or VII (**Figure 2.3**).

Evaluation of Average Root Length at Extreme Cold Temperatures

The effect of cold temperature on early root elongation of the *Glycine soja* diversity core collection was assessed. Selected accessions were germinated in ragdolls at 11, 17, 22, and 28 °C, and roots were measured to the nearest 0.1 cm. In evaluation of the genetic materials from 2 replicates of 2 years of seed source germinated at lower temperatures, basic statistics were obtained on the lengths of roots for the observed samples (**Table 2.4**). Subsequently, the mixed effects model showed that both genotype and temperature were significant ($p < 0.05$) for root length; however, the interaction was insignificant and was dropped from the model to improve the fit (**Table 2.5**). PI 549048 (MG III) was estimated to have the longest average root length at 5.3 cm with a standard error of 0.85 cm. In contrast, PI 407042 (MG V) had the shortest average root length estimate at 2.4 cm with a standard error of 0.85 cm (**Table 2.3**). PI 549048 had an average root length that was statistically superior ($p < 0.05$) to 42 of the other 68 genotypes tested. Of these, 11 were early maturity group (MG III or IV) accessions (**Figure 2.4**), 22 were MG V (**Figure 2.5**) and 9 were MG VI or VII (**Figure 2.6**).

Evaluation of Percent Germination at Extreme Hot Temperatures

The effect of elevated temperature on germination of the *Glycine soja* diversity core collection was assessed. Selected lines were germinated in ragdolls at 28, 33, 38, and 43 °C to represent a range between proposed ideal germination temperature for soybean and increments of 5 °C above the ideal. Four temperatures were used, as 4 growth chambers were available for testing. In evaluation of the genetic materials from 2 replicates from 2 years of seed source germinated at elevated temperatures, basic statistics were run on observed

samples (**Table 2.7**). The mixed model showed that genotype and temperature were significant, but the interaction was not significant (**Table 2.8**). Similar to the studies on colder temperatures, PI 522211A (MG III) had the highest estimated percent germination at 89% with a standard error of 10%. PI 407029 (MG IV) had the lowest estimated percent germination at 27% with a standard error of 10% (**Table 2.3**). The germination of PI 522211A was statistically superior to 25 of the other 67 genotypes tested. Of these 25, 8 were MG III and MG IV accessions (**Figure 2.7**), 9 were MG V (**Figure 2.8**), and 8 were later-maturing (MG VI and MG VII) accessions (**Figure 2.9**).

Evaluation of Average Root Length at Extreme Hot Temperatures

The effect of elevated temperature on early seedling vigor of the *Glycine soja* diversity core collection was assessed. Selected lines were germinated in ragdolls at 28, 33, 38, and 43 °C to represent a range between proposed ideal germination temperature for soybean and increments of 5 °C above the ideal. In evaluation of the genetic materials from 2 replicates of 2 years of seed source germinated at elevated temperatures, basic statistics were run on observed samples (**Table 2.10**). The mixed model showed that genotype and temperature were significant; however, the interaction was not and was dropped from the model to improve the fit (**Table 2.11**). Similar to the extreme cold results, PI 549048 (MG III) demonstrated the longest average root length with an estimate of 6.9 cm and a standard error of 0.8386 cm. PI 507656 (MG VII) produced an estimate of only 0.8 cm with a standard error of 1.2462 cm (**Table 2.3**). PI 549048 was statistically superior to 54 of the 68 other genotypes tested. Of the 54, 13 were early maturity groups (MG III or MG IV) accessions

(Figure 2.10), 28 were MG V (Figure 2.11), and 13 were MG VI or VII accessions (Figure 2.12).

DISCUSSION

The germination of accessions from the USDA *Glycine soja* diversity core collection at sub-optimal temperatures revealed differences in the germination and early root growth of wild soybean genotypes. The genotypes were selected to represent maximum genetic variance and thus a variety of responses were expected from this study. While there was not a genotype that outperformed all others under either cold or heat stress, clear groupings can be identified for lines exhibiting superior germination. Under both colder and hotter temperatures, PI 522211A had the highest estimates for percent germination, while PI 549048 had the highest estimates for average root length.

Multiple statistical models were constructed in order to fit the data properly. A generalized linear structure was first attempted in analysis of the percent germination data due to the binary nature of a seed's ability to either germinate or not germinate; however, convergence could not be met using this method, so a mixed effects model with restricted maximum likelihood estimates for genotypic means was utilized. Both percent germination and average root length had a high number of observations ($n = 898$ for cold study, $n = 846$ for heat study) and normality assumptions were met following model diagnostics. The final models were chosen based on lowest values of Akaike information criterion (AIC) and Bayesian information criterion (BIC) for each model fit.

Following assessment of the type of model to be run, we assessed terms to be included in the models. Use of 4 genotypic replicates coming from two field replicates of

maturity group source from each of two years of seed production helped account for expected variance present due to seed source. Replicate within source and year, source within year, and year were treated as random effects representing only a sampling of different levels that these factors could assume. In this way, the effects due to genotype and temperature as well as their interaction could be more accurately assessed.

Maturity group was also included in earlier models to see if groupings could be made based on growth habit; however, there was not a significant association, so this was ultimately excluded from the resulting models ($p = 0.62$). Lack of significance in this case means, for the diverse genotypes utilized in this study, a stronger germination trend could not be assigned based on maturity group. This result was unexpected as early maturity group accessions are planted in shorter growing season areas that may tend to have lower soil temperatures on average. We hypothesized that MG III and IV accessions may have performed better in the cold temperatures and conversely that MG VI and VII accessions may have withstood heat stress; however, this association was not seen in this study. This could expand the breeding capabilities of these diverse lines to be used by soybean breeders in different areas that may not normally use a range of maturity groups in order to capture increased germination vigor for breeding lines.

The lack of a significant interaction term in the statistical model suggests that no certain combination of genotype and temperature causes an effect that is unsuspected from the trend created by the individual components of temperature and genotype. This means that if a genotype performed better in a colder environment, it generally performed better under all colder temperatures; that is, the trend appears to be linear. This trend held for both

experiments including all temperatures tested in both directions; however, the significant interaction term found in the 11 °C suggests a possible genotype by environment effect at the lowest temperature. Generally, a decreasing trend was seen in genotypes at this lower temperature; however, PIs 522211A (MG III), 549048 (MG III), 407042 (MG V), 407191 (MG V), 407195 (MG IV), 424045 (MG V), 562561 (MG V), 378686B (MG VI), 597458C (MG V), and 597461 B (MG V) did not follow this trend. The germination of these PIs appeared to be robust even at 11 °C. Further, PI 424052 appears to show a similar trend when exposed to the highest temperature (43 °C), but the estimate is only based on one replication so confirmation would be needed on this observation. This might be useful in the practice of breeding for a soybean variety with improved ability to adapt to changing temperatures, as it is unlikely to have a constant temperature in a field setting but rather a range of generally cooler or hotter temperatures. If climate changes causes farmers to plant soybean earlier, lines that remain vigorous during episodes of cool temperatures should perform well. Therefore, the lack of an overall interaction between genotype and temperature might suggest an improvement in the breeding value of the trends noted from this study.

A recommendation for choosing a line performing better under temperature stress could be made in a number of ways depending on the goal of the breeding project; however, it is reasonable from the results of this study to identify accessions that appeared in the superior groupings from both the percent germination and average root length studies that might be of interest. For the cold germination trials this includes: early maturing accessions: PI 549048, PI 522211A, PI 424116, and PI 549046, MG V accessions: PI 562551, PI 407287, PI 424025B, PI 562561, PI 407191, PI 339871A, PI 424045, PI 407206, PI 407248,

PI 597458C, PI 407314, PI 424035, PI 407214, and later maturing accessions: PI 407096, PI 378686B, PI 407059, and PI 507624.

There appeared to be more separation among the accessions tested at hotter temperatures; however, the combination of percent germination and average root length still seems to prove a valid approach. Plant introductions satisfying these criteria were: early maturing accessions: PI 549048, PI 424116, PI 424052, PI 522211A, PI 532453A, PI 549032, MG V accessions: PI 424025B, PI 562561, PI 407228, PI 339871A, PI 597458C, PI 597461B, PI 407191, and one representative from the later maturity groups: PI 407096. In general, more early maturing accessions were represented in the upper grouping for heat stress than later maturity groups. Again heat-stress tolerance among earlier maturity groups was unexpected given that earlier maturity groups are from cooler climates and late maturity group are generally from warmer climates.

Upon beginning this study, it was obvious that extensive research on the germination patterns of wild soybean was not available. As such, there were some problems of optimization in this study that might be improved in further experimenting. For example, though it was suggested in the literature that 28-30 °C is the ideal germination temperature for cultivated soybean, it is possible that this figure differs for wild soybean as is suggested by the longer average root lengths shown in the sample means at 33 C compared to 28 C (**Table 2.10**), though further study would need to be done to validate this claim.

Further, the harder seed coats of wild soybean create a challenge for imbibition and subsequently germination. A method for scarification was developed by our group; however, finding a standard procedure that was optimized for each genotype proved difficult due to the

differences in seed coat even amongst the genotypes surveyed. The problem of the hard seed coat was dealt with by incrementally scarifying seeds to avoid damaging seeds with softer seed coats. However, some issues could have been created by either under- or over-scarifying samples. This is also important as it could bias the average root length results to favor those that did germinate. Therefore, a very vigorous root may not have developed simply because it failed to break the seed coat.

The primary utility of this data is to identify wild soybean parents that can improve vigor in progeny derived from crosses with domesticate soybean. It is possible that the improved vigor is completely dwarfed by all cultivated soybean backgrounds and that the contribution of a wild source would be insignificant. Though there is believed to be greater seedling vigor by cultivated soybean as compared to wild soybean, additional study with more *Glycine max* lines or progeny of crosses between wild soybean and domesticated would be needed to validate this claim. However the *G. soja* Diversity Core Collection was selected to be genetically diverse; therefore, these wild soybean selections are likely to include loci that improve these targeted traits. Studies identifying QTL or the nature of the trait heritability would determine the usefulness of this data and a difference in alleles common to wild versus cultivated soybean. This study was conducted to identify phenotypic variation in genetic materials being utilized for increasing genetic diversity in the soybean gene pool. Relevant progeny from many of the wild soybean accessions used in this study will be available to determine if the ability to germinate at extreme temperatures observed in this study is heritable and if it substantially invigorates growth of progeny at extreme temperatures,

CONCLUSION

Differences in germination at suboptimal temperatures were found amongst the USDA *Glycine soja* diversity core collection. A significant interaction between genotype and temperature was found for select wild soybean accessions that suggests a robust germination even at the coldest temperatures tested. The data could be used in identification of more vigorous parental materials for making genetically diverse soybean germplasm releases. In addition, the collection of this phenotypic data provides characterization of historically underutilized genetic resources for use in soybean improvement.

REFERENCES

- Alm, D. M., Stoller, E. W., & Wax, L. M. (1993). An index model for predicting seed germination and emergence rates. *Weed Technology*, 7(3), 560-569.
- Chebrolu, K. K., Fritschi, F. B., Ye, S., Krishnan, H. B., Smith, J. R., & Gillman, J. D. (2016). Impact of heat stress during seed development on soybean seed metabolome. *Metabolomics*, 12(2), 1-14.
- Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., et al. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, 8(2), 024024.
- Hatfield, J. L., and D. B. Egli. (1974). Effect of temperature on the rate of soybean hypocotyl elongation and field emergence¹. *Crop Science*. 14:423-426.
- Heatherly, L. G., & Spurlock, S. R. (1999). Yield and economics of traditional and early soybean production system (ESPS) seedlings in the midsouthern United States. *Field Crops Research*, 63(1), 35-45.
- Kenward, M., & Roger, J. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, 53(3), 983-997.
- Kramer, C. (1956). Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications. *Biometrics*, 12(3), 307-310.
- Koti, S., Reddy, K. R., Reddy, V. R., Kakani, V. G., & Zhao, D. (2005). Interactive effects of carbon dioxide, temperature, and ultraviolet-B radiation on soybean (*Glycine max* L.) flower and pollen morphology, pollen production, germination, and tube lengths. *Journal of Experimental Botany*, 56(412), 725-736.

- Mendelsohn, R., Nordhaus, W. D., & Shaw, D. (1994). The impact of global warming on agriculture: A Ricardian analysis. *The American Economic Review*, 84(4), 753-771.
- Schlenker, W., Hanemann, W. M., & Fisher, A. C. (2006). The impact of global warming on U.S. agriculture: An econometric analysis of optimal growing conditions. *The Review of Economics and Statistics*, 88(1), 113-125.
- Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106(37), 15594-15598.
- Tukey, J W (1953). "Some selected quick and easy methods of statistical analysis". Transactions of the New York Academy of Sciences (0028-7113), 16 (2), p. 88.
- Wuebker, E. F., Mullen, R. E., & Koehler, K. (2001). Flooding and temperature effects on soybean germination journal paper no. J-18452 of the Iowa Agric. and Home Econ. Exp. Stn., Ames, IA, project no. 3244, and supported by Hatch Act and State of Iowa funds. *41*(6), 1857-1861.

TABLES and FIGURES

Table 2.1. Sample statistics of percent germination for cold germination temperatures (Temp). While average percent germination are similar at 17, 22, and 28 °C, there is a significant difference at 11 °C. Minimum and Maximum indicate the highest and lowest values recorded for each temperature.

Analysis Variable : Percent Germination					
Temp	Observations	Average	Std Dev	Minimum	Maximum
11	227	67.14	29.74	0.00	100.00
17	223	87.51	18.84	0.00	100.00
22	224	88.65	18.39	0.00	100.00
28	224	86.46	18.55	4.00	100.00

Table 2.2. Percent germination ANOVA for evaluation of genotypes at cold germination temperatures. This model initially included the interaction term of genotype and temperatures. While this term was not left in the final model, it was included in this table to show the lack of significance ($p = .9989$). It was excluded from later models following the same test.

Type 3 Tests of Fixed Effects				
Source	Numerator DF*	Denominator DF*	F Value	p-value
Genotype	68	528	3.37	< .0001
Temperature	3	613	67.02	< .0001
Genotype*Temperature	204	613	0.70	0.9989

* Degrees of Freedom

Table 2.3. Least squares estimates of mean percent germination and average root length for genotypes tested at cold and hot temperatures arranged by genotype.

Genotype	Cold Temperatures		Hot Temperatures	
	Percent Germination	Average Root Length	Percent Germination	Average Root Length
339871A	87.54	4.1291	64.63	5.2297
366122	79.05	4.8126	40.44	3.9376
378683	58.95	2.7910	43.99	2.2177
378684B	87.34	3.6700	64.95	3.9928
378686B	87.60	4.1879	41.22	2.3996
378690	74.11	1.8175	35.56	2.3641
378696B	82.56	3.7088	51.33	3.7448
378697A	86.34	3.0776	70.53	4.1895
407020	91.50	2.5475	39.84	1.7229
407029	90.40	3.8173	27.20	3.0635
407038	86.70	3.3544	46.65	2.2503
407042	90.35	2.4350	67.41	3.4552
407052	73.95	2.7888	46.48	2.4401
407059	83.07	3.9835	44.90	1.7304
407060	68.52	3.5677	52.34	3.9890
407085	79.93	3.6650	62.98	3.3380
407096	87.18	4.2050	69.49	5.1797
407156	64.32	2.9629	52.77	2.4168
407157	84.55	3.1176	59.21	3.6595
407171	81.56	4.0130	73.01	4.6279
407179	91.12	3.4725	66.02	3.9390
407191	84.62	4.1923	58.00	4.8673
407195	81.81	3.8594	63.26	4.0700
407206	81.88	4.1478	58.23	4.5555
407214	87.29	3.8094	50.79	3.9201
407228	89.72	3.6875	60.04	5.3041
407231	85.27	3.1609	50.02	4.1664
407240	82.17	2.8897	60.23	4.0183
407248	83.25	4.0684	47.13	3.4262
407287	90.08	4.7134	49.00	3.4990
407300	82.63	3.6631	54.07	4.0956
407314	90.73	4.0163	67.18	4.5644
424007	75.68	3.5524	48.03	2.3857
424025B	87.16	4.4994	66.14	5.6233
424035	90.40	3.9125	61.69	3.9007
424045	91.58	4.1088	67.50	3.2319

Table 2.3 Continued.

Genotype	Cold Temperatures		Hot Temperatures	
	Percent Germination	Average Root Length	Percent Germination	Average Root Length
424070B	88.54	3.1163	55.83	4.1497
424082	82.68	3.0434	61.21	3.1724
424083A	80.99	3.6135	54.89	4.4009
424102A	83.72	3.2944	41.02	2.1415
424116	92.18	4.9475	64.34	5.7635
424123	85.99	3.3500	54.63	3.4074
468916	80.74	3.6194	77.82	4.6721
479751	39.99	2.4072	-	-
479752	99.75	3.8198	78.27	3.6776
483466	91.18	3.1369	69.45	4.0216
483467	80.65	4.6477	55.15	4.2974
507618	83.63	3.4194	68.06	3.1201
507624	82.64	3.8138	63.43	3.0713
507630	80.73	3.3900	43.71	1.8690
507641	86.03	3.7231	58.17	3.7175
507656	73.02	3.9166	35.93	0.7514
522211A	107.38	4.0973	88.86	5.7024
532453A	60.68	4.4688	60.28	5.0142
549032	57.91	2.5182	63.91	4.8015
549046	85.04	4.4980	47.87	3.9345
549048	89.69	5.3446	73.11	6.9163
562547	83.29	3.2347	68.81	3.7780
562551	86.88	4.7425	60.24	3.9509
562553	87.30	3.4394	55.79	3.5515
562561	91.62	4.2000	84.54	5.3079
562564	58.27	1.9530	48.44	0.9895
562565	78.05	4.5844	44.58	3.9596
593983	56.86	3.3150	48.61	4.6874
597458C	89.35	4.0481	76.46	5.1596
597460A	65.68	3.3327	40.23	2.4646
597461B	84.80	3.7452	69.91	5.0402
597462B	58.01	3.5613	33.41	3.4457

Table 2.4. Sample statistics of average root length for cold germination temperatures. Root length increased as temperature approached the theoretical ideal, but standard deviations did as well. A large number of samples were present at each temperature level. Minimum and Maximum indicate the highest and lowest values recorded for each temperature.

Analysis Variable: Average Root Length					
Temp	Observations	Average	Std Dev	Minimum	Maximum
11	227	0.19	0.36	0	1.50
17	223	1.89	0.92	0	5.72
22	224	4.90	2.25	0	10.54
28	224	7.67	3.06	0.07	13.79

Table 2.5. Average root length ANOVA for evaluation of genotypes at cold germination temperatures. The two significant fixed effects, genotype and temperature, were used for the final model. These values were used in determining significant differences in genotype based on percent germination.

Type 3 Tests of Fixed Effects				
Source	Numerator DF*	Denominator DF*	F Value	p-value
Genotype	68	687	1.79	0.0002
Temperature	3	817	857.29	<.0001

* Degrees of Freedom

Table 2.6. Sample statistics of percent germination for hot germination temperatures. Minimum and Maximum indicate the highest and lowest values recorded for each temperature.

Analysis Variable: Percent Germination					
Temp	Observations	Average	Std Dev	Minimum	Maximum
28	213	78.50	23.15	0.00	100.00
33	211	67.78	24.27	0.00	100.00
38	210	58.79	25.63	0.00	100.00
43	212	25.39	29.00	0.00	96.15

Table 2.7. Percent germination ANOVA for evaluation of genotypes at hot germination temperatures. While the interaction term was not left in the final model, it was included in this table to show the lack of significance ($p = 1.0000$).

Type 3 Tests of Fixed Effects				
Effect	Numerator DF*	Denominator DF*	F Value	Pr > F
Genotype	67	489	3.09	<.0001
Temp	3	565	164.65	<.0001
Genotype*Temp	201	565	0.49	1.0000

* Degrees of Freedom

Table 2.8. Sample statistics of average root length for hot germination temperatures. Number of observations was lower for 43 °C due to lack of measurable roots from seeds not germinating at this high temperature. Minimum and Maximum indicate the highest and lowest values recorded for each temperature.

Analysis Variable: Average Root Length					
Temp	Observations	Average	Std Dev	Minimum	Maximum
28	212	4.43	2.09	0.13	12.09
33	209	5.79	2.95	0.00	14.32
38	209	3.37	2.52	0.00	11.00
43	142	1.71	2.24	0.00	8.85

Table 2.9. Average root length ANOVA for evaluation of genotypes at hot germination temperatures.

Type 3 Tests of Fixed Effects				
Effect	Numerator DF*	Denominator DF*	F Value	p-value
Genotype	67	596	2.28	<.0001
Temp	3	696	95.33	<.0001

* Degrees of Freedom

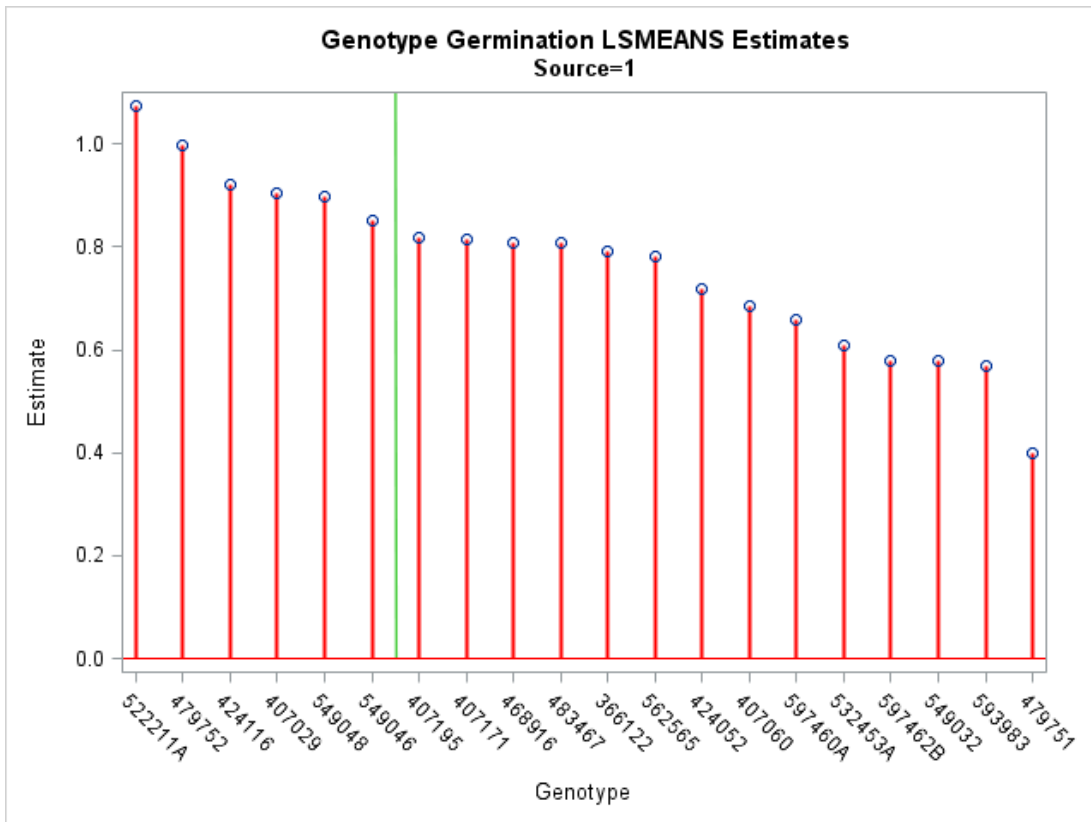


Figure 2.1. Needle plot of genotypic LS means estimates of percent germination for maturity groups III and IV (Source 1) tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A.

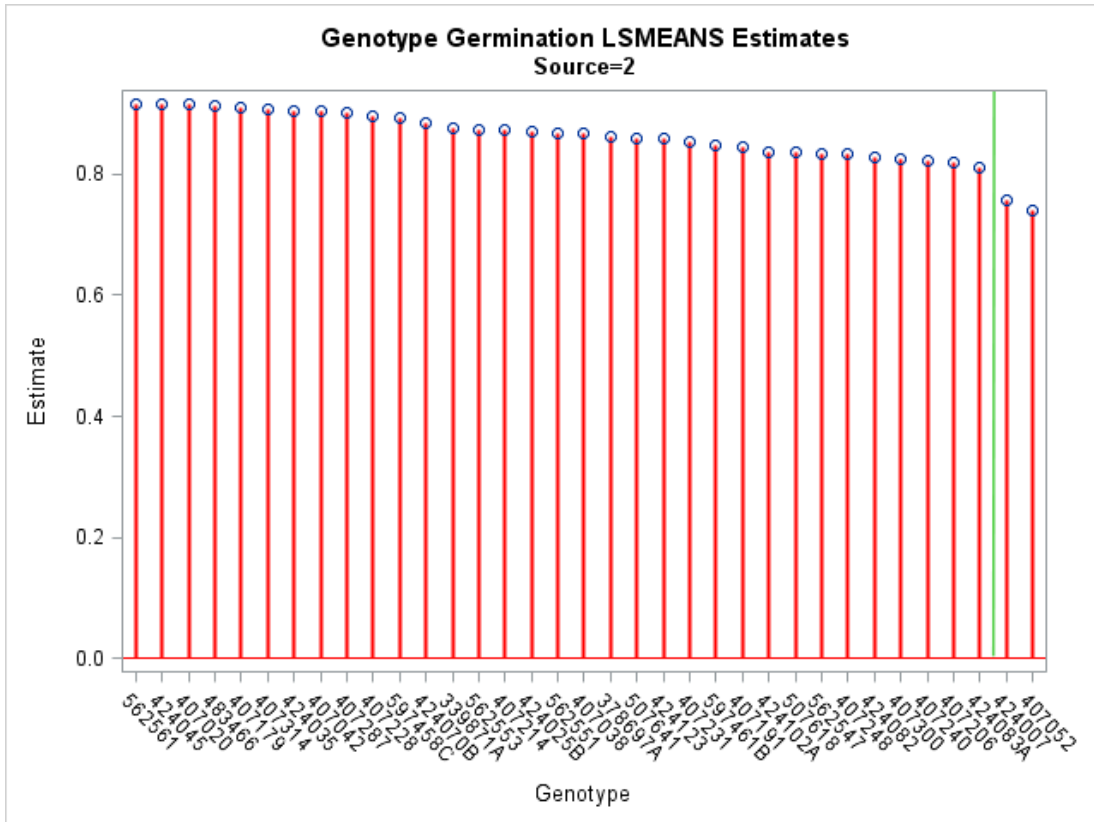


Figure 2.2. Needle plot of genotypic LS means estimates of percent germination for maturity group V (Source 2) tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A.

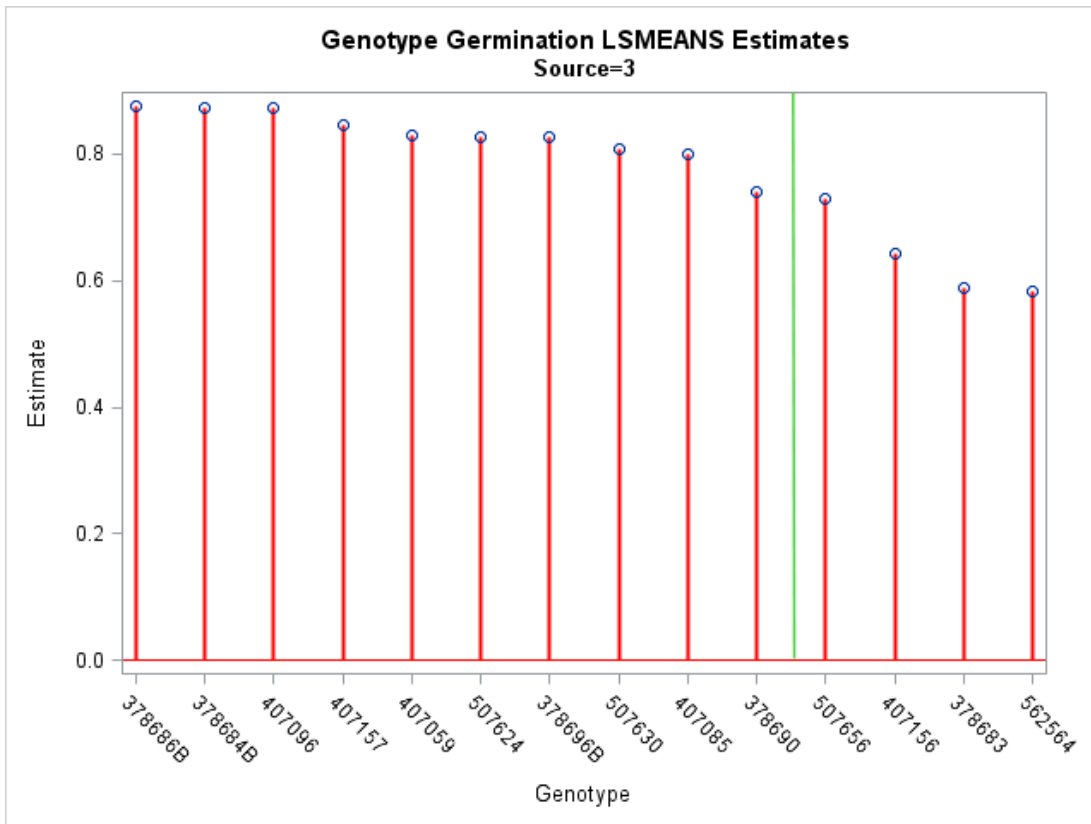


Figure 2.3. Needle plot of genotypic LS means estimates of percent germination for maturity groups VI and VII (Source 3) tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A.

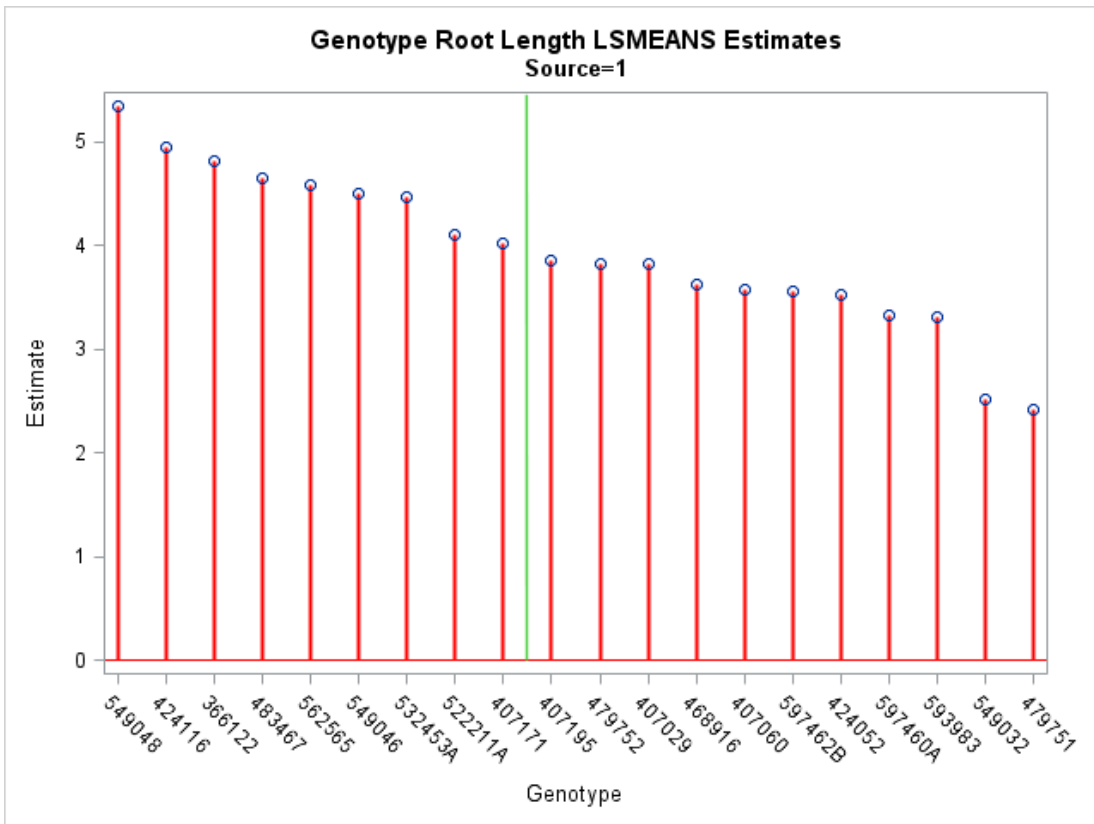


Figure 2.4. Needle plot of genotypic LS means estimates of average root length for maturity groups III and IV (Source 1) tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048.

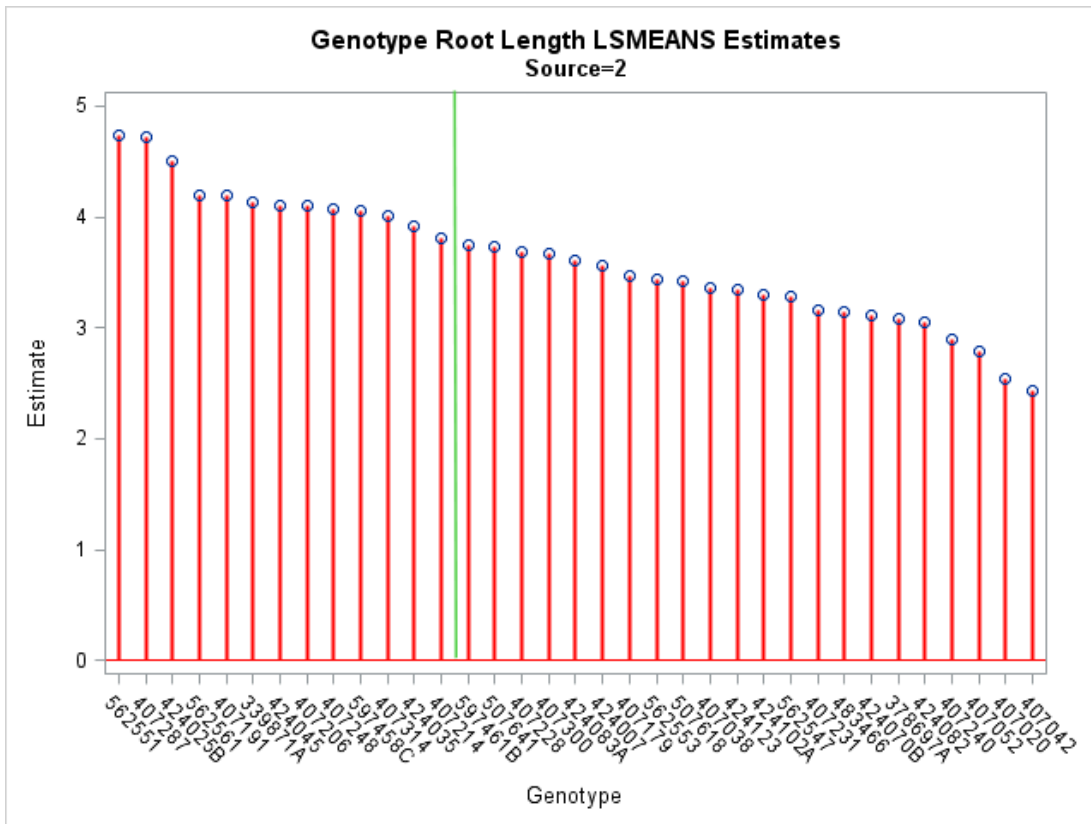


Figure 2.5. Needle plot of genotypic LS means estimates of average root length for maturity group V (Source 2) tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048.

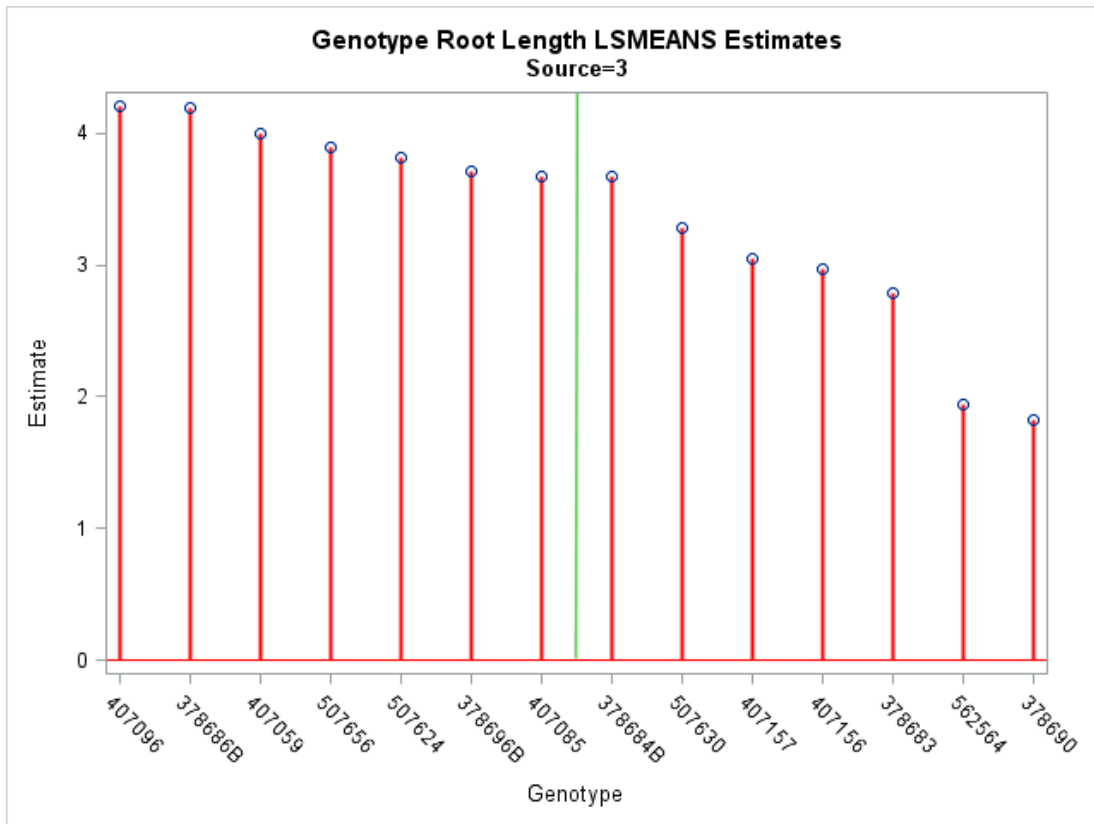


Figure 2.6. Needle plot of genotypic LS means estimates of average root length for maturity groups VI and VII (Source 3) tested at extreme cold temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048.

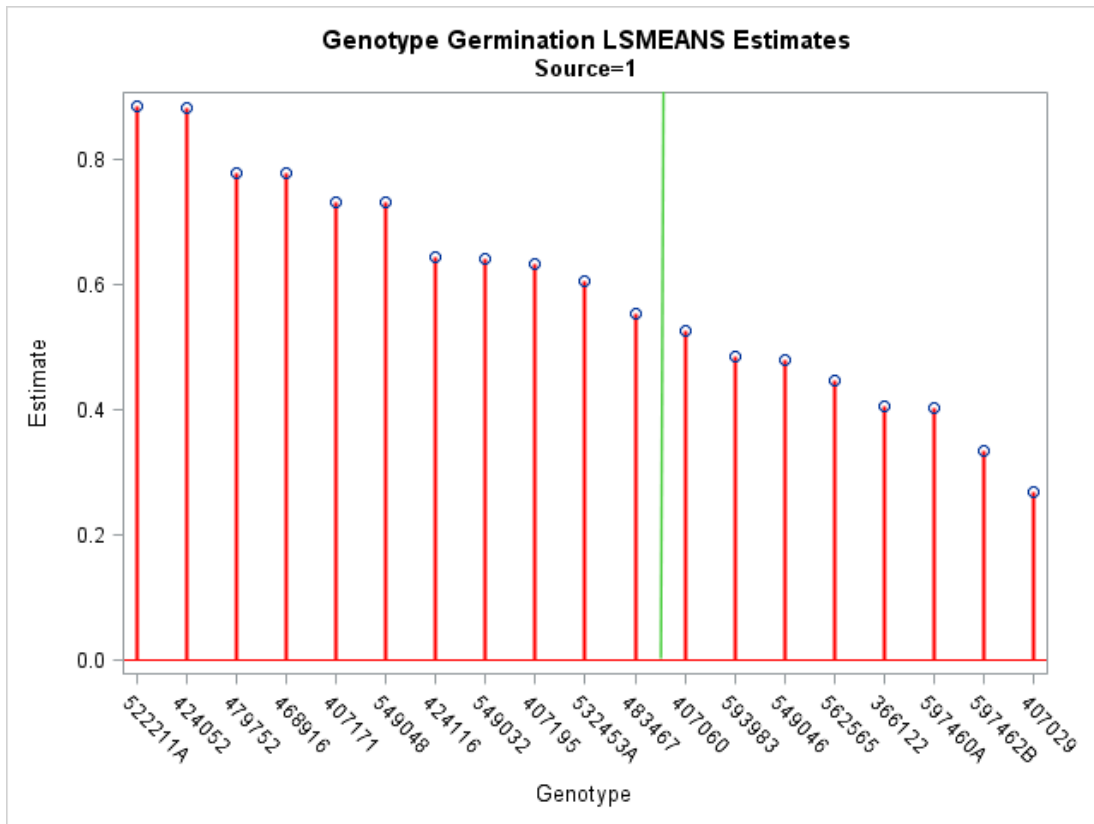
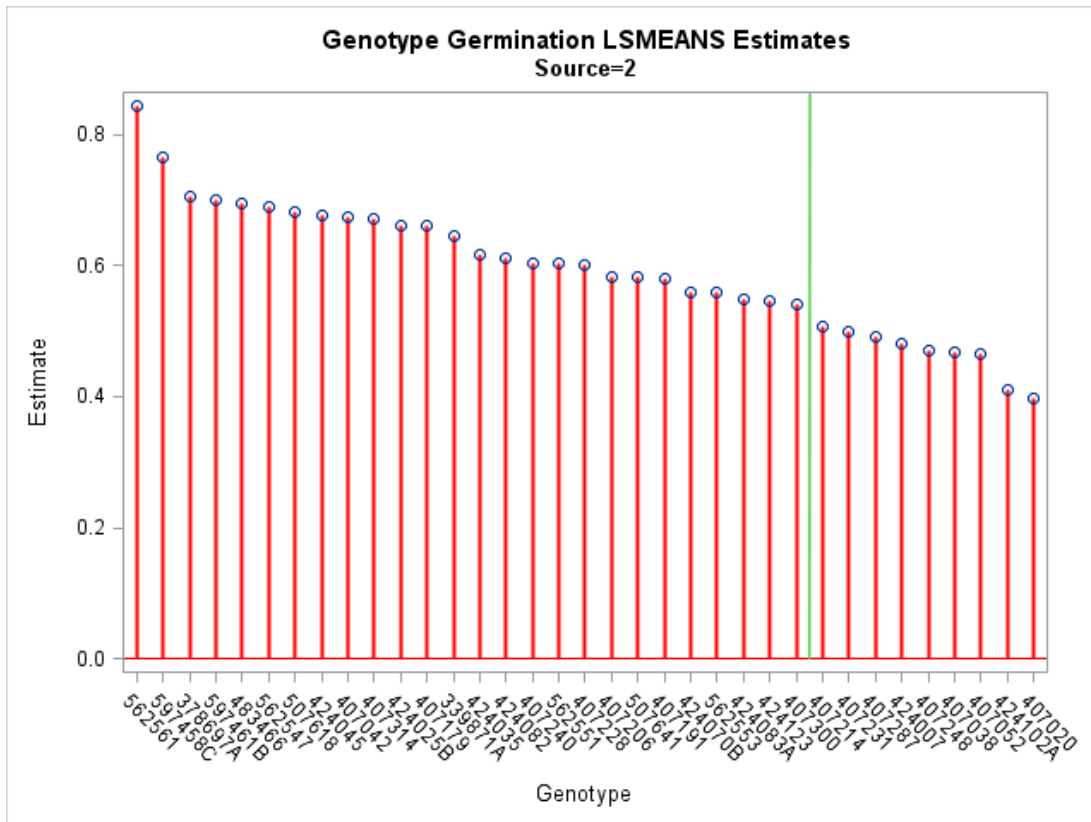


Figure 2.7. Needle plot of genotypic LS means estimates of percent germination for maturity groups III and IV (Source 1) tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A.



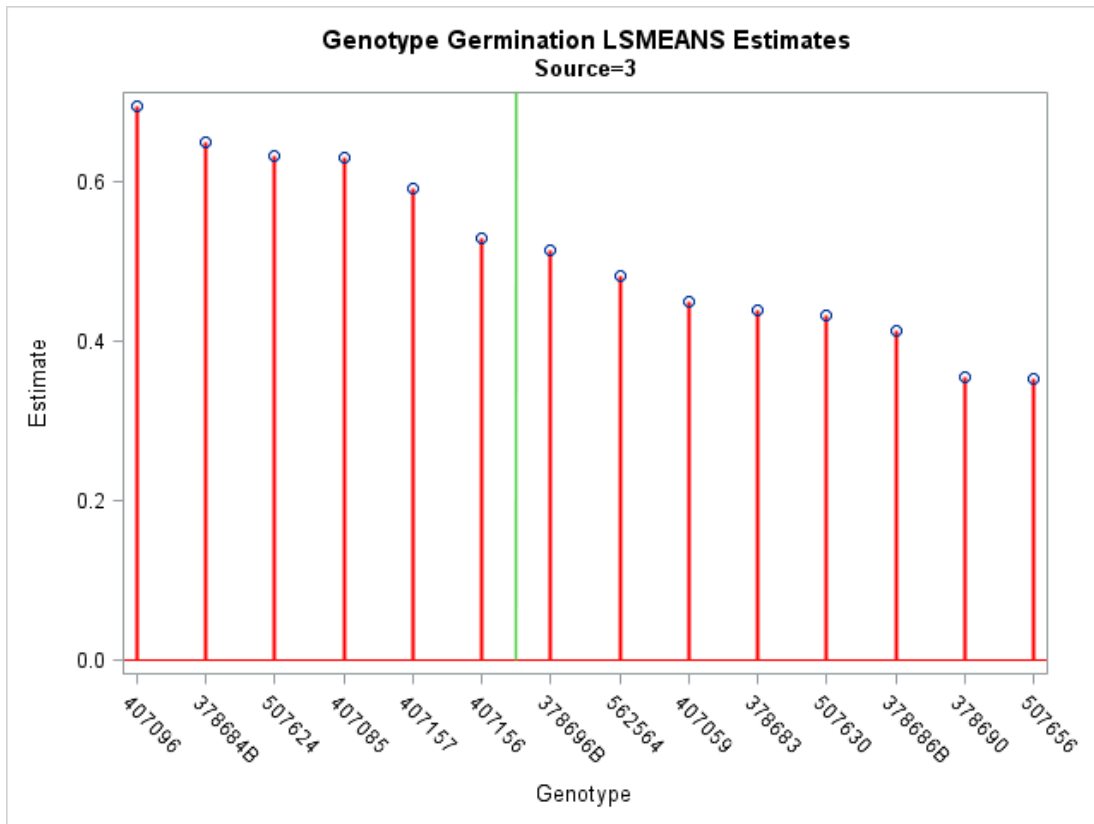


Figure 2.9. Needle plot of least squares estimates of percent germination for each of the MG VI and VII (Source 3) tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of percent germination, PI 522211A, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 522211A.

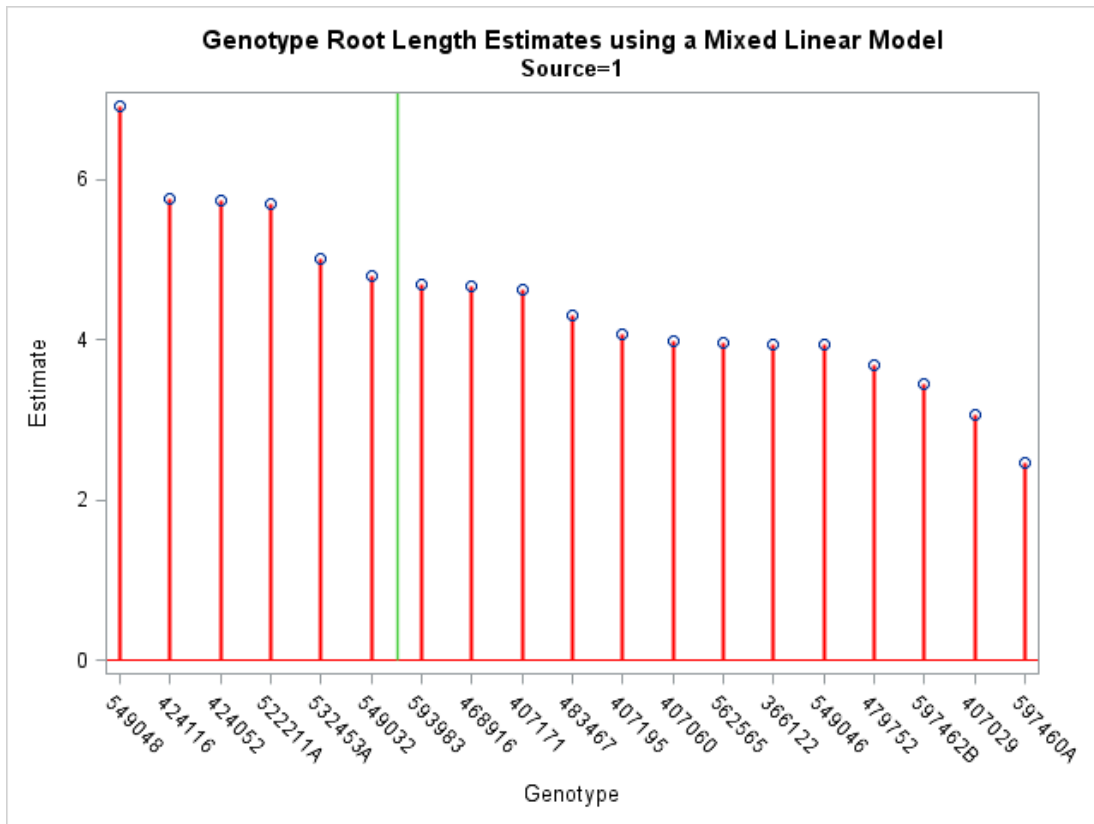


Figure 2.10. Needle plot of genotypic LS means estimates of average root length for maturity groups III and IV (Source 1) tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048.

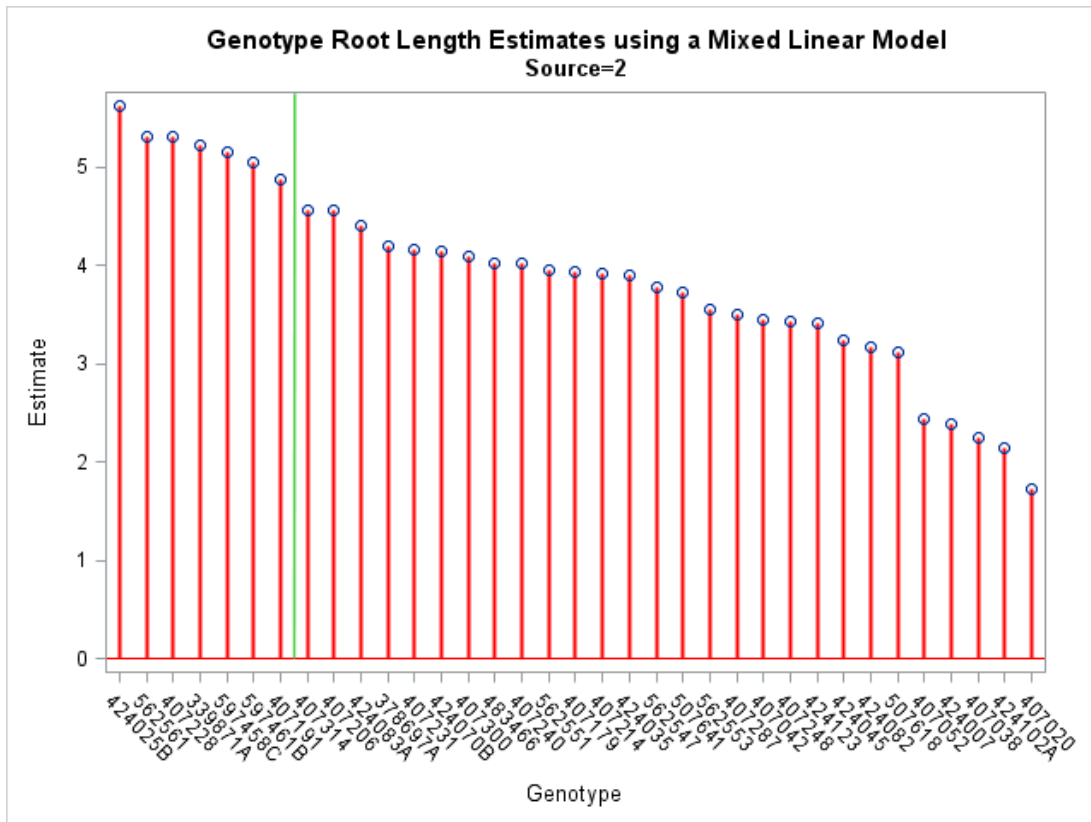


Figure 2.11. Needle plot of genotypic LS means estimates of average root length for maturity group V (Source 2) tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048.

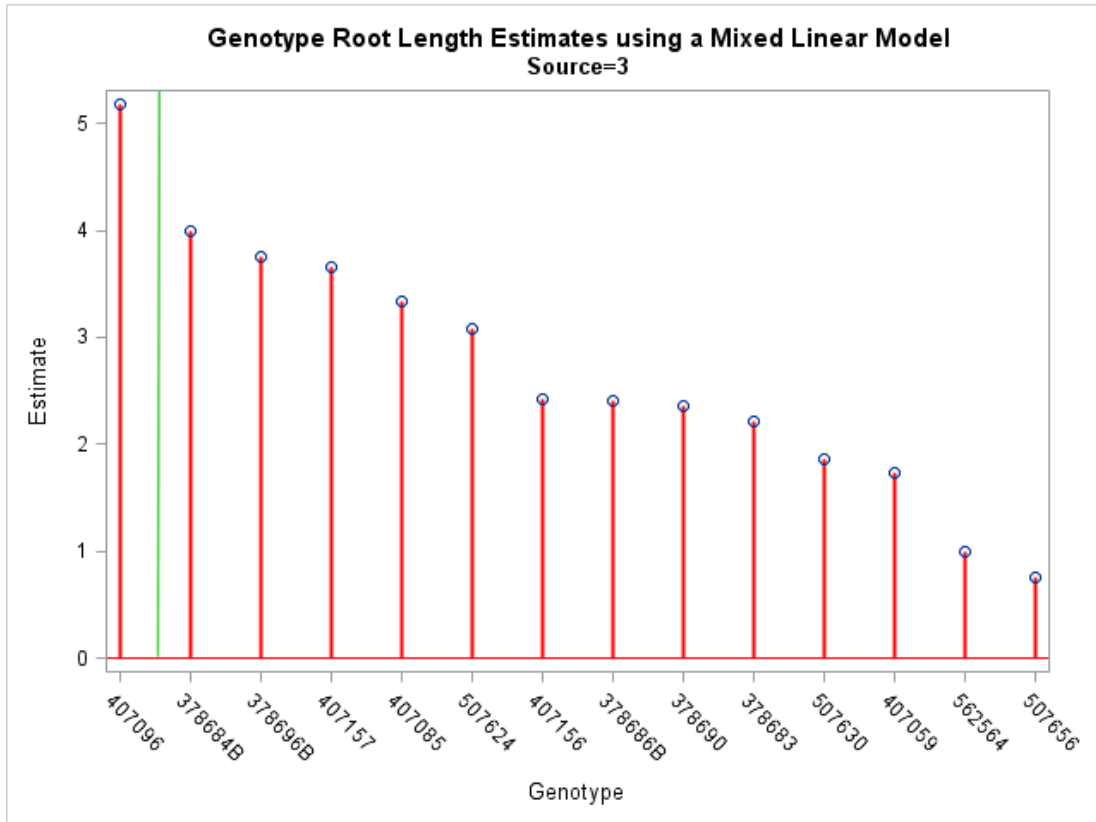


Figure 2.12. Needle plot of genotypic LS means estimates of average root length for maturity groups VI and VII (Source 3) tested at extreme hot temperatures. The green bar indicates the point at which the best overall genotype in terms of average root length, PI 549048, maintains a significant advantage over the genotypes listed on this plot. Those to the right of the bar were statistically inferior to PI 549048.

**Chapter III: RNA-Seq Study Reveals Genetic Response of Diverse Wild Soybean
Accessions to Increased Ozone Levels**

ABSTRACT

Ozone is a pollutant widely known to cause decreases in productivity in many plant species, including soybean. With ozone levels expected to rise significantly by the year 2050, characterizing plant response to ozone and developing varieties capable of tolerating increased levels of the gas have become of interest. While cultivated soybean response to ozone has been studied, less work has been done to identify sources of resistance from wild relatives. This study characterizes the response to ozone of a genetically diverse group of wild soybean accessions and uses RNA-Seq to identify changes in gene expression following exposure to ozone in both sensitive and resistant wild soybean accessions.

Results indicated many more genes responding to elevated ozone in accessions displaying an ozone-sensitive phenotype with a large amount of overlapping gene responses. Fewer genes responded in resistant accessions, and the responses were less correlated indicating two possible different response mechanisms to ozone exposure. Defense response genes were up regulated in both sensitive wild soybean accessions and down regulated in resistant accessions. A depletion of gene expression involved in photosynthetic activity was noted for both sensitive accessions as well as one resistant accession. Resistant accession PI 507656 did not show a significant decrease in expression of genes connected to photosynthetic activity. Genetic response to ozone in wild soybean showed a reduction of number of genes responding in resistant accessions.

Unique response to ozone was also noted for different leaf positions within each accession. Older leaf tissue displayed a consistent depletion of genes involved in

photosynthesis, while younger tissue had a pattern of more defense genes responding. All tissue seemed to display response in the form of oxidoreductase activity.

INTRODUCTION

The effect of ozone (O_3) on agriculture has been of increasing concern. O_3 levels have risen from < 10 parts per billion (ppb) in the late 1800s (Volz and Kley, 1988) to daytime levels of nearly 40 ppb in many areas of the Northern Hemisphere (Stevenson et al., 2006). As populations rise and technology advances, it is expected that these levels could increase to 75 ppb by the year 2050 (Ainsworth et al. 2012). Effects of these increased levels have proven detrimental to plant health; the damages to carbon assimilation, stomatal conductance, and plant growth result in losses to crop yield (Ainsworth et al., 2008; Ashmore et al., 2005; Feng et al., 2008; Fiscus et al. 2005). Losses in total crop global production have been estimated to be from \$14 billion to \$26 billion (Van Dingenen, 2009).

Ozone has been shown to react with many different molecules in the apoplast of leaves to produce different harmful products known as Reactive Oxygen Species (ROS), such as OH^\cdot radicals and hydrogen peroxide (Grimes et al., 1983; Heath, 1987), making the capacity of the apoplast to handle ROS the primary defense mechanism of the plant (Moldau, 1998). After exposure to levels of >150 ppb, stress is displayed by changes in ROS, phytohormone levels, calcium ion binding and changes in mitogen-activated protein kinase (MAPK) signaling cascades. The ozone response pathway shares many similarities to that of programmed cell death often seen in response to pathogens. Both processes involve rise in ROS production in response to stress, which further activates the ethylene, salicylic acid, and

jasmonic acid pathways for expression of defense genes (Ainsworth, 2012). This also emphasizes the dual role of ROS, both as a signaling molecule and potential toxin.

Responses to extreme levels of ozone that are unlikely to be seen in nature are easy to identify, but slightly lower levels encountered in nature do not cause such drastic responses. Chronic exposure to ozone has been shown to cause decreases in photosynthesis and plant mass as well as early leaf senescence (Fiscus et al., 2005). Exposure to acute treatments with ozone has been observed for different species. *Arabidopsis* displayed a rapid temporary decrease in stomatal conductance with a rise in ROS followed by a slower recovery to normal levels of conductance; however, this was not thought to affect overall photosynthetic activity (Kollist et al., 2007). Chronic exposure causes irreversible lower stomatal conductance, which can directly damage plant productivity (Wittig et al., 2007). Stomatal sensitivity to abscisic acid may also be affected by O₃ stress causing transpiration levels to increase when exposed to drought (Mills et al. 2009). Lower stomatal conductance is likely to indicate a decrease in photosynthesis (Kusumi et al., 2012).

Recent studies done on *Glycine max* showed a reduction of up to 25% in yield largely due to smaller seed size and fewer pods/plant (Betzberger et al., 2010). This is consistent with other studies showing reduced photosynthesis in soybean as well in other crops such as wheat and rice (Morgan et al. 2003, Feng, 2008; Ainsworth, 2008). Increased cellular respiration may lead to a further reduction in productivity (Gillespie et al., 2012).

Burkey and Carter (2009) conducted a screen of 30 North American ancestral lines to determine if resistance exists in the background of present cultivars. A ten-fold difference was discovered between responses in these lines indicating development of resistant soybean

varieties is possible. More in-depth studies on soybean response to ozone damage have been done in recent years. Whaley et al. (2015) conducted a gene expression analysis of a resistant *G. max* cultivar, Fiskeby III, and a sensitive cultivar, Mandarin (Ottawa) subjected to treatment with high [75 ppb] and low [25 ppb] ozone. This study shows that cultivated soybean exhibits differential expression among expected ROS and defense pathways, but there are cultivar-specific genes between resistant and sensitive cultivars. Expression of many genes varied greatly at different time points for the sensitive cultivar with different stress response genes responding at different times. In contrast, resistant soybean showed consistent defense response and an up-regulation in metabolism genes. The Whaley study also found a possible link between leaf architecture and ozone response as demonstrated by the regulation in expression of wax and cutin biosynthesis genes of the sensitive variety. In addition, Burton et al. (2016) identified several putative QTL for ozone resistance in a mapping population created from a cross between Mandarin (Ottawa) and Fiskeby III. In the same study, a different response between younger and more mature tissue was noted.

In this study, we identify wild soybean accessions with contrasting responses to ozone and characterize the gene expression response to elevated ozone of wild soybean (*Glycine soja* Zucc and Sieb). An RNA-Seq analysis was done using 4 plant introductions (PIs) of *Glycine soja* treated under high ozone (75 ppb) and charcoal-filtered (< 10 ppb) conditions at each of 3 different leaf development positions. Early response to ozone was characterized, and leaves of different maturities were utilized to determine effects of leaf age in response to ozone.

MATERIALS and METHODS

Identification of Phenotypic Response of *Glycine soja* to Increased Ozone

806 *G. soja* accessions from the USDA Soybean Germplasm Collection were identified as being >.01% genetically different than any other accession in the collection. From these accessions, clusters of ~10 were created and eventually a 10% Core Collection (81 accessions) was chosen by Dr. Qiagin Song as being the most genetically diverse (personal comm., 2014). Ozone responses for 70 of the 81 accessions chosen for the USDA *Glycine soja* Diversity Core had been previously evaluated in three annual trials from 2012 - 2014 in collaboration with Dr. Burkey's lab (**Table 3.1**). The remaining 11 accessions were not tested due to their early maturity and lack of relevance for Southern germplasm. The effect of ozone was retested on these 70 varieties. Accessions were planted in pots placed in 3 replicated blocks of continuous stirred tank reactors (CSTR) subjected to a treatment of 75 parts per billion (ppb) ozone exposure or a control treatment of charcoal filtered (CF) air for a period of 5 days. CSTRs are cylindrical chambers surrounded by Teflon film designed for containment and mixing of gases (Heck et al., 1978). Varieties identified as sensitive or resistant from earlier *Glycine max* studies (2012) were included in the chambers for validation of results.

The plants were rated by 5 – 9 individual raters on a scale from 1 (resistant) to 4 (sensitive) based on percent foliar damage in the method developed by Heagle (1979) for characteristic ozone damage (1 = 0-5%, 2 = 5-25%, 3 = 25-50%, and 4 = 50-100%). Plants were rated alongside corresponding control treated plants for verification of damage due to ozone and not some extraneous growth habit. These ratings are shown in **Table 3.1**.

Compiling data from two years of phenotypic evaluation allowed for selection of two genotypes that consistently exhibit ozone resistance (PI 424123 and PI 507656) as well as two sensitive to ozone (PI 424007 and PI 407179)

Collection of Ozone-Treated Samples

Ozone-sensitive wild soybean accessions PI 407179 (MG V; 37.78 N, 127.12 E) and PI 424007 (MG V; 37.21 N, 126.82 E) and ozone-resistant wild soybean accessions PI 424123 (MG V; 35.71 N, 129.21 E) and PI 507656 (MG VII; 33.00 N, 129.50 E) were germinated and grown in a greenhouse with charcoal-filtered air (< 10 ppb ozone) in June 2014 for a period of 14 days at the USDA-ARS facility in Raleigh, NC, USA. 6-inch pots filled with Fafard 2 Mix (Sungro, MA) and 1 teaspoon of Osmocote Plus (Scotts Company, SC) were sown with 5 seeds but was thinned down to the best two seedlings where appropriate. One week after planting, Marathon (OHP, Inc., PA) (.25 tsp./pot) was applied for thrips prevention. The plants were then moved to CSTRs in an adjacent greenhouse bay to acclimate to conditions two days prior to ozone exposure.

Earlier studies on *G. max* indicate that gene expression levels changed drastically within hours of exposure to ozone; therefore, individual trifoliates were collected separately following seven hours of treatment with 75 ppb [O₃] (Whaley et al., 2015). Each genotype also had control plants set into chambers under CF conditions. Actual ozone levels were between 3 – 10 ppb for CF chambers and 75 ± 3 ppb for treatment chambers. CSTR temperature, relative humidity, and light levels were 35 ± 2 °C, 67 ± 6%, and 300 ± 60 μmol m⁻² s⁻¹ PAR, respectively, during the exposure period.

This design was replicated in 4 blocks within the greenhouse bay. Leaves from 2-3 plants of the same genotype within the same CSTR corresponding to the same trifoliolate were pooled in order to obtain sufficient tissue for RNA isolation. The first three trifoliate (T1-3) were collected from each plant. T1 was the most mature leaf, while T3 was the least mature leaf. Though less mature than T1 and T2, T3 was still a completely unfolded trifoliolate in all plants. Samples were harvested into liquid nitrogen to prevent changes of gene expression during handling and to limit RNase activity. The samples were transferred to an -80 C freezer for storage.

Isolation of RNA and Sequencing

Expressed transcripts for a total of 95 samples were sequenced (4 genotypes * 2 ozone conditions * 3 leaf positions * 4 replicates) – (1 degraded sample). Total RNA was isolated using Plant RNeasy kits with on-column DNase I digestion (Qiagen, CA). Isolated samples were sent to the Genomic Sciences Laboratory (GSL) at North Carolina State University in Raleigh, NC to assure quality on Bioanalyzer RNA nano chips (Agilent, CA), prepared into Truseq libraries (Illumina, CA), and sequenced using the HiSeq 2500 (Illumina, CA). Samples were multiplexed in lanes to allow for ~10 samples to be sequenced in each. Between 18-32 million x 125 bp single-end reads were obtained for each sample. Sequencing was done in 2 rounds – 2 replicates in August 2014 and the other 2 replicates in May 2015.

Analysis of RNA-Seq Data

Samples were first analyzed for quality and nucleotide bias using the FastX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The sequences were trimmed to 108 bases to

reduce primer bias and quality-filtered to a PHRED score of 34. Trimmed sequences were then aligned to the Williams 82 soybean reference genome *Glycine max* Wm82.a2.v1 with the available GFF file annotations using Tophat2 (Trapnell, 2009). Alignment rates from the first set of sequencing were between 65-81%, while the second set aligned between 86-93% of input reads. In order to determine the cause of lower alignment for the first samples, alignment of the unmapped sequences from the first run of sequencing was run with the chloroplast and mitochondrial genomes; the results showed an additive alignment of 85-95%, indicating that lower genomic alignments were caused by a higher presence of chloroplastic and mitochondrial mRNA.

The resulting files were input into Cufflinks 2 (Trapnell et al., 2013) using the Williams 82.v2.a1 reference genome annotations as a guide for transcript assembly. The Cufflinks script *cuffmerge* was then used to merge the transcript assemblies from each sample in a single-expressed transcriptome based on all FASTQ files. *Cuffdiff* was subsequently run multiple times using the *-u* and *-b* options to correct for possible fragment and multi-read bias in order to determine differential gene expression between different leaf positions, genotypes, and each leaf position and ozone treatment combination within each genotype. Gene expression is reported in Fragments Per Kilobase of transcript per Million mapped reads (FPKM). FPKM normalizes counts of short sequences by read depth and transcript length. Differential expression was determined using *cuffdiff* output for genes exhibiting significance at a false discovery rate (FDR) of $p < 0.05$ coupled with a log₂-fold change of ≥ 1 or ≤ -1 . Recent work by the SEQC/MACQ-III Consortium has displayed a strong correlation between expression data from RNA-Seq and qPCR (>80%), so qPCR

validation was not performed on this data (2013); however, further experiments are being conducted using specific target genes on additional genotypes of *Glycine soja*.

Significant differentially-expressed genes at the \log_2 -fold change = ± 1 following ozone treatment were then analyzed for associated gene ontologies (GO) using the single enrichment analysis (SEA) tool on AgriGO (<http://bioinfo.cau.edu.cn/agriGO/index.php>). SEA results returned GO term categories enriched or depleted in response to ozone found in each sample based on an adjusted $p < 0.05$. Based on available gene model annotation information, gene numbers reported could fit multiple ontology categories. Absence of a reported gene number in an enrichment category may not signify zero genes in that category but rather a lack of significant change of gene expression in response to ozone. Ontology categories in this study were reduced to eliminate redundancy of identical gene sets matching similar categories. These were chosen to match the most detailed descriptions without presenting misleading gene numbers. Expression analysis was visualized with graphs and figures created using *CummeRbund*, an R package for data visualization of Cufflinks data (Goff, et al.), and the *VennDiagram* R Package (Chen).

RESULTS

Sequencing of Ozone Sensitive and Resistant Wild Soybean Lines

Although several wild soybean accessions screened showed varying responses to ozone exposure (**Table 3.1**), four genotypes were chosen to represent the extreme responses based on consistent phenotypic ratings: PIs 424123 and 507656 for resistance to ozone and PIs 424007 and 407179 for sensitivity to ozone. Ninety-five samples were collected from four biological replicates of the first three trifoliates from both ozone-treated and CF treated

plants of each genotype; RNA was isolated for use in sequencing. Between 18-32 million 125 bp single-end reads were obtained from each sample.

Analysis of RNA-Seq Data

Using *cuffdiff* (Trapnell, 2009) to obtain differential gene expression results, comparisons were made between genotypes, treatments, and leaf positions. The numbers of genes that were differentially expressed following treatment with ozone relative to the control were identified and evaluated. In these studies, we use differential expression of genes to refer to changes in the steady state level of gene transcripts measured by RNA-Seq at which there was a log₂-fold change of at least 1 or -1 observed between control and ozone-treated samples and a false discovery rate (FDR) of < 0.05. While the exact numbers of genes differentially expressed varied by genotype, the two resistant genotypes displayed similar numbers of differentially-regulated genes with PI 424123 having 2,100 and PI 507656 having 1,863 differentially-regulated genes. Analogously, the two sensitive genotypes had similar responses with PI 407179 showing 6,374 differentially expressed genes, and PI 424007 showing 6,457 (**Table 3.2**). A full listing of genes responding differentially between treatments with associated FPKM and log₂-fold changes can be found on the NCBI Gene Expression Omnibus (Project: GSE85146) (<http://www.ncbi.nlm.nih.gov/geo/>).

To assess the similarity of the genotypes, unique gene counts were created for all combinations of genotypes, and the counts of genes differentially regulated in response to ozone shared by different accessions are shown in Venn diagrams. Comparison of all genes differentially expressed in the resistant genotypes showed that the two shared responses of 732 genes, while PI 424123 maintained 1,368 unshared responsive genes compared to 1,131

genes shown by PI 507656 (**Figure 3.1**). The sensitive varieties showed more similarity with 3,934 shared differential genes, while in PI 424007 2,923 transcripts were uniquely expressed compared to 2,440 genes by PI 407179 (**Figure 3.2**). Total numbers of shared genes between all combinations of genotypes can be seen in **Figure 3.3**.

Spearman correlations of the log₂-fold change values for corresponding genes following ozone treatment were obtained to assess genotype/leaf position combination similarity. These revealed responses in gene expression at different leaf positions were more closely related in certain genotype/leaf position combinations than others (**Table 3.3**). For example, sensitive genotypes have a correlation of 0.29274 at the third trifoliolate, 0.462203 at the second trifoliolate, and 0.430592 at the first trifoliolate. In contrast, the resistant genotypes show lower correlations at every position with 0.118728 at the third trifoliolate, 0.169758 at the second trifoliolate, and 0.19454 at the first trifoliolate.

Ozone Response in Sensitive Genotypes PI 407179 and PI 424007

Gene Ontologies (GO) have been added to the annotation of the soybean reference genome to help make sense of large data sets. GO categorize genes through a species-independent vocabulary based on their connection to biological processes, molecular functions, and cellular components in the plant. In this section, genes that were differentially expressed in at least one leaf position in each sensitive genotype are evaluated to determine patterns of changes in transcript abundance. Insight into patterns of differential gene expression can be gained by measuring enriched gene ontologies connected to each of these categories. Single Enrichment Analysis (SEA) using the GO of the genes differentially expressed in both PIs 407179 and 424007 showed significance in several gene ontology

(GO) categories (95:119). As noted earlier, the responses in differentially-expressed genes of the two sensitive genotypes had significant overlap. Genes involved in cell redox homeostasis and photosynthesis were found to be down regulated in both genotypes. In addition, genes were down regulated with involvement in several cell macromolecule biosynthetic, metabolic, and catabolic processes (**Table 3.4**).

Down-regulation of genes enriched by biological process differed between sensitive genotypes as well. PI 407179 showed a reduction of expression in genes involved in heterocyclic metabolism, lipid metabolism, nucleotide metabolism and catabolism, protein folding and polymerization, as well genes annotated with roles in the light reactions of photosynthesis and photosystem II (PSII) stabilization that were not shared by its sensitive counterpart. PI 424007 showed enrichments for down-regulation of genes connected to response to hormone stimuli and steroid metabolism not shared by PI 407179. A full listing of the number of down-regulated genes connecting to the tests for significance of selected biological process enrichment can be seen in **Table 3.4**.

While a reduction of biological processes was discovered in sensitive genotypes, several enrichments were also found among genes identified as up regulated. Both genotypes shared positive enrichments for genes involved in cell communication, cell recognition, cell wall macromolecule metabolism and modification, defense response, as well as recognition of pollen, regulation of transcription and response to biotic stimuli. Differences in up-regulated enrichments included aminoglycan metabolism, carboxylic acid biosynthesis, chitin, chlorophyll and chorismate metabolisms and DNA replication, which were significantly up-regulated in PI 424007 but not in PI 407179. PI 407179 was not enriched for

any unique terms in this comparison. A full listing of significant GO up-regulated terms can be seen in **Table 3.5**.

Molecular function groupings of differentially-expressed genes also showed depletions among sensitive genotypes. Both were down regulated for genes involved in metal binding, specifically calcium, iron, and iron-sulfur complex, as well as hydrolase and transferase activity on glycosyl bonds and serine, and xyloglucan:xyloglycosyl transferase activity. Alkylbase DNA N-glycosylase activity, FAD binding, serine-type endopeptidase activity, and translation termination factor activity categories were depleted in PI 407179, while PI 424007 had reduction in anion (sulfate) transmembrane transporter activity and sigma factor activity. A full listing of significant enrichments of down-regulated genes can be found in **Table 3.4**.

Molecular function ontology enrichments of up-regulated genes shared by ozone-sensitive accessions included: ammonia-lyase activity, calcium ion binding, sequence-specific DNA binding, steroid dehydrogenase activity, transcription regulation, ubiquitin-protein ligase activity, and macromolecular metabolism and biosynthesis. PI 407179 was only uniquely enriched with up-regulation of genes connected to acyl-group transferase activity; however, PI 424007 had several unique enrichments including: carbon-nitrogen lyase activity, chitin binding and chitinase activity, coenzyme bonding, heme binding, O-methyltransferase activity, and unfolded protein binding. The numbers of genes matching significant categories of up-regulation can be found in **Table 3.5**.

Genes in the oxidation-reduction biological process as well as the oxidoreductase molecular function displayed both up-regulation and down-regulation in the sensitive

genotypes. In the case of this study, a clear pattern of relocation of oxidoreductase capability in response to ozone did not present itself. A sampling of the pattern of varied expression amongst these genes is shown using a heatmap in **Figure 3.4**. Cellular component enrichments for down-regulated genes for sensitive were connected to the apoplast (**Figure 3.5**), cell wall, chloroplast, extracellular region, oxygen evolving complex, photosynthetic membrane, as well as both photosystems I & II. PI 407179 also showed depletion for genes in the cytochrome b6f complex. Both genotypes showed enrichment of up-regulated genes in the ubiquitin ligase complex, while PI 424007 showed positive enrichment in the endoplasmic reticulum (ER) and extracellular matrix. Down- and up-regulated cellular component gene numbers can be seen in **Tables 3.4** and **3.5**, respectively

Ozone Response in Resistant Genotypes PI 424123 and PI 507656

One of the most important aspects of this research is to identify patterns of gene expression that distinguish the response of ozone-resistant accessions from ozone sensitive accessions. Genes that were differentially expressed in at least one leaf position in each resistant genotype are evaluated to identify patterns of changes in transcript abundance. An evaluation of genes affected by ozone for each leaf position is presented later in this chapter. Fewer genes were responsive in either resistant genotype as compared to the sensitive genotypes; however, Single Enrichment Analysis (SEA) of the genes differentially expressed in both PIs 424123 and 507656 returned significance in multiple gene ontology (GO) categories (66:53). In the biological process category, both resistant genotypes were enriched for down-regulated genes involved in cellular glucan metabolism, lipid biosynthesis and metabolism and polysaccharide metabolism. PI 424123 was uniquely reduced in genes

belonging to photosynthesis, while PI 507656 had genes involved in steroid metabolism depleted. The number of genes connected to significant down-regulated terms can be seen in **Table 3.4**.

Significant up-regulated biological processes by way of single gene enrichment analysis for resistant genotypes included: cell recognition, phosphate metabolism, post-translational protein modification, recognition of pollen, and transcription regulation. PI 424123 did not show any significant unique up-regulated biological processes; however, PI 507656 was enriched for genes associated with aminoglycan metabolism, cell wall macromolecule metabolism, chitin metabolism, defense response, lipid biosynthesis, and response to biotic stimuli. Up-regulated gene enrichments can be seen in **Table 3.4**.

Resistant accessions displayed shared enrichment of down-regulated genes corresponding to molecular functions of coenzyme and cofactor binding, electron carrier activity, endopeptidase and enzyme inhibitor activity, FAD binding, heme binding, hydrolase activity on glycosyl bonds, iron ion bonding, UDP-N-acetylmuramate dehydrogenase activity, and xyloglucan:xyloglucosyl transferase activity. Also under molecular function, PI 424123 was uniquely enriched for down-regulation of genes linked to alkylbase DNA N-glycosylase activity, while PI 507656 had enrichments for steroid metabolism, O-methyltransferase activity, and steroid dehydrogenase activity. The full list of genes depleted in ozone resistant accession in response to ozone treatment can be seen in **Table 3.4**.

Molecular functions enriched among resistant accessions were more accession-specific. Shared enrichment was only present in sequence-specific DNA binding and unfolded protein binding. PI 424123 showed significant up-regulation for ATPase activity,

calcium ion binding, endoribonuclease activity, acid anhydride hydrolase activity, protein kinase activity and ribonuclease, specifically RNase III, activity. PI 507656 had up-regulated enrichments for chitinase activity, O-methyltransferase activity, steroid dehydrogenation, and transcription regulation. A complete list of up-regulated genes in significant gene ontology categories can be seen in **Table 3.5**. Genes involved in oxidoreductase activity again showed differential expression in both directions without a cellular component of localization. A heatmap showing relative expression of specific oxidoreductase genes shows the lack of a pattern of expression amongst these genes (**Figure 3.6**).

Resistant genotypes shared cellular component enrichments for down-regulated genes in the apoplast (**Figure 3.7**), cell wall, and extracellular regions; PI 424123 displayed additional significance for oxygen evolving complex, photosynthetic membrane, as well as both PSI & PSII. Neither resistant accession showed significance for cellular components enriched with up-regulated genes. Gene numbers for significant GO categories of down-regulation can be seen in **Table 3.4**.

Comparison of Resistant and Sensitive Genotype Response to Ozone

The most important goal of the research is to identify changes in gene expression associated with resistance to ozone. Therefore, the responses that contrast between sensitive and resistant accessions are most likely to explain the differences in the response of these accessions to ozone. In this section, genes differentially expressed in at least one leaf position in each genotype were evaluated to determine resistant-specific patterns of changes in transcript abundance in response to ozone. Comparison of significant gene ontology categories between resistant and sensitive varieties showed differences in processes and

functions affected by ozone exposure. Sensitive accessions shared significant down-regulations in cell redox homeostasis and negative regulation of catalytic activity that were absent in resistant genotypes. In contrast, resistant genotypes did not have any significant down-regulated categories absent in the sensitive genotypes. Up-regulated enrichment categories of biological process unique to sensitive varieties include: cell communication, cell wall macromolecule metabolism, and protein ubiquitination. Resistant varieties shared no unique enrichments for up-regulation of genes connected to any one biological process.

In the category of molecular function, fructose 1,6-bisphosphate 1-phosphatase activity, calcium ion binding, identical protein binding, iron-sulfur cluster binding and protein disulfide oxidoreductase activity were differentially regulated in both sensitive accessions but by neither resistant accession. Resistant genotypes shared enrichments for down-regulation of genes associated with enzyme regulation, heme binding, and UDP-N-acetylmuramate dehydrogenase activity that are absent in either sensitive genotype. Up-regulated gene molecular function enrichments unique to sensitive accessions include: ammonia-lyase activity, cofactor binding, fatty acid synthase activity, intramolecular lyase activity, and ubiquitin-protein ligase activity. Resistant genotypes shared no unique molecular function enrichments for up-regulated gene activity.

The 29 genes differentially expressed in both PIs 424123 and 507656 but in neither sensitive genotype (**Figure 3.3**) did not return any significant ontology terms; though, some of the individual genes were annotated with genes corresponding to possible defense responses. These genes were all down regulated in the resistant accessions.

Glyma.03G246800, Glyma.08G114700, and Glyma.09G138600 are all metal-binding genes.

Glyma.18G289100 matched with ontologies of “heat shock cognate protein” and “defense response.” Glyma08G341100, Glyma.08G341400, and Glyma.08G342100 are all putative trypsin inhibitors (Koiwa, 1997), and Glyma.17G065400 is an additional xyloglucan:endo-transglucosylase gene. The entire list of available annotations can be seen in **Table 3.6**.

Only two genes matched criteria of being down regulated in both sensitive accessions and up regulated in both sensitive accessions. One of the genes, Glyma.06G297200, had no annotation, while Glyma.16G170100 is described as a hydroxyproline-rich glycoprotein. Sixteen genes matching the opposite criteria, up regulated in sensitive accessions and down regulated in resistant accessions, showed more descriptive results of genes connected to known stressors and defense responses (**Table 3.6**).

Ozone Response Varies in Genotype Based on Leaf Position

Intra-genotypic Spearman correlations were obtained to characterize plant response to ozone based on age of leaf tissue. PI 424007 displayed the most similar gene expression following ozone exposure between leaf positions with values from 0.444517-0.536984; PI 407179 was also highly correlated with values from 0.37761-0.522321. PI 424123 showed similarity between the first two trifoliates (0.372723), but neither T1 nor T2 had a correlation with T3 (0.084234; 0.010379, respectively). Correlations between leaf positions for PI 507656 were more similar but were lower than those seen among sensitive genotypes (0.13003-0.24135) (**Table 3.3**).

Gene ontology analyses of genes differentially expressed in all three leaf positions as well as each individual position were conducted to explain these correlations at the gene

enrichment level. As it was shown that direction of gene regulation did not vary between leaf positions, differential expression was used for this analysis to show total response. Genes matching significant ontologies in either direction are listed in **Tables 3.8 – 3.11**. Categories enriched followed the same directional patterns as mentioned in the previous section for genotype analysis.

Within the sensitive accessions, all three trifoliate in PI 407179 responded with genes connected to P-P-bond-hydrolysis transmembrane transporter activity, oxidoreductase activity as well as additional transferase, hydrolase and ATPase activity. The first trifoliate (most mature leaf) responding with unique genes linked to photosynthesis and associated chloroplast and photosystem cellular components, cell redox homeostasis, oxidoreductase activity, translation termination factor activity, and iron-sulfur cluster binding. The second trifoliate was enriched for genes associated with protein modification, as well as phosphatase and phosphorous transport activity. The third trifoliate contained unique genes for defense response and response to biotic stimulus. The filtered list of significant ontology matches with associated gene numbers can be seen in **Table 3.8**.

The other sensitive accession, PI 424007, had the highest correlation coefficient, and as such, revealed many significant ontology categories connected to all three trifoliate. These included: recognition of pollen, oxidoreductase activity, lipid and steroid biosynthesis, phosphate metabolism, cell communication, defense response, calcium ion binding, response to biotic stimulus, and xyloglucan:xyloglucosyl transferase activity. The first trifoliate responded uniquely with genes connected to photosynthesis and the chloroplastic region as well as both photosystems I & II. The second oldest trifoliate was uniquely enriched only for

genes connected to photosystem II, while the third trifoliolate had unique genes involved in cellular glucan metabolism, xyloglucan:xyloglucosyl transferase activity, and copper binding. The filtered list of enrichments by leaf position with direction of regulation for PI 424007 can be seen in **Table 3.9**.

In the resistant accessions, the genes shared by all three trifoliate were far fewer as was expected from the lower correlation coefficients. PI 424123 was only enriched for transferase activity of hexosyl groups. The first trifoliolate displayed the most enrichments of unique genes with connections to photosynthesis, oxidation-reduction, calcium ion binding as well as the extracellular matrix and thylakoid regions of the cell. Trifoliolate 2 was enriched for genes involved in oxidoreductase activity, endoribonuclease activity, coenzyme binding, and a component enrichment for the extracellular region, and Trifoliolate 3 had significant unique genes involved in oxidoreductase activity. A full list of gene numbers with associated direction of regulation and relevant ontology enrichments can be seen in **Table 3.10**.

PI 507656 showed a low correlation consistently at each leaf position and did not show any significant enrichment shared by all three trifoliate. However, there were significant categories returned for each individual leaf position. Trifoliolate 1 was enriched for cellular glucan metabolism, oxidation-reduction processes, endopeptidase inhibitor activity, glycosyl bond hydrolase activity, and transferase activity of both glycosyl groups and xyloglucan:xyloglucosyl. In addition, the genes in the first trifoliolate were significantly enriched in the apoplast, cell wall, and extracellular regions. Trifoliolate 2 was again enriched for oxidation-reduction activity in addition to lipid metabolism, coenzyme, FAD, and iron-ion binding as well response in the extracellular region. The third trifoliolate was uniquely

enriched for genes involved in aminoglycan, cell wall macromolecule, chitin, phosphate and steroid metabolisms, in addition to lipid biosynthesis, protein amino acid phosphorylation and O-methyltransferase activity. The third trifoliolate was also uniquely enriched for genes connected to defense response and response to biotic stimulus. Complete numbers of genes and directions of regulation connected to each category for PI 507656 can be found in **Table 3.11**.

DISCUSSION

Many more genes were responsive to ozone exposure in sensitive genotypes than in resistant genotypes. It would seem that wild soybean ozone resistance, at least in these selected accessions, is associated with a diminished plant response at the transcriptome level rather than a large response to ozone (**Table 3.2**). In fact, sensitive genotypes were enriched for activation of defense response genes, but resistant genotypes displayed far fewer. Genes associated with photosynthetic activity were significantly down-regulated in both sensitive genotypes as well as older leaves of resistant genotype PI 424123, but not for other resistant accession PI 507656.

There was also a clear trend showing a reduced number of genes responding at younger leaf tissue. This is a trend that has been noted before by Burton et al. (2016) in a QTL-study for ozone response in cultivated soybean. The trend is not demonstrated, however, for PI 507656. A parsimonious explanation for this different response to ozone is that the third trifoliolate was at a slightly different stage of development in this particular accession, possibly related to genotypic differences among these diverse accessions. Different enrichments show that genes involved processes like photosynthesis and metal ion

binding are affected at older leaf positions but that genes involved in defense and stress response occur in younger leaf tissue as well.

Unfolded protein response (UPR), which describes the process of proteins not being properly folded, has been previously associated with plants responding to increased production of ROS in plants (Martinez et al., 2003). It is known to affect several important cellular processes, and the up-regulation of genes with connections to unfolded protein binding following ozone treatment in both resistant genotypes as well as PI 424007 could suggest a reduction in efficient protein folding and the plant's response to eradicate the unfolded proteins. In addition, PI 424007 shows enrichment for genes being up regulated in the endoplasmic reticulum (ER), where protein folding typically occurs. This connection of ER stress has been noted before in Arabidopsis following increased ROS production (Ozgun et al., 2015). This suggests that the protein folding response is more severe in this sensitive genotype. Further demonstrating the connected to improper protein folding was the enrichment of disulfide oxidoreductase activity in both of the sensitive genotypes, as this activity is thought to responsible for creation of disulfide bonds often seen in protein folding (Martinez et al., 2003).

Both sensitive and resistant genotypes showed an enrichment for genes responding to oxidation-reduction processes, which confirms the efficacy of the ozone treatment. Many of these genes were not annotated beyond their role in oxidoreductase activity; though, some had more specific functions predicted. Glyma.08G238100 is linked to cytochrome P450 and electron carrier activity in addition to its oxidation-reduction activity. Glyma.07162900 is a gene for production of allene oxide synthase, a precursor of jasmonic acid and a previously

documented gene involved in plant defense response (Sivasankar, 2000). However, the response of the collective group of genes involved in oxidation-reduction was sporadic, and there was not a clear pattern of up- or down-regulation. In addition, no specific pattern of up- or down-regulation among genes involved in oxidoreductase activity appeared to be centered by cellular compartment. This could be due to the fact that the responses occur at a nonlinear rate, and the point of tissue collection could show very different levels of expression. We conclude it is more informative to track the presence of differential expression in this case rather than the direction of regulation, as the kinetics are often changing as the plant responds (Burkey personal comm., 2016).

Though many processes in plants are known to produce ROS, photosynthesis is well documented as being a contributor to these molecules (Mittler, 2002). We hypothesize that down-regulation of genes involved in photosynthetic activity is a response to ROS in other parts of the cell. This could explain the sporadic response of genes involved oxidoreductase activity following treatment with ozone, notably in apoplasts and extracellular regions as well as chloroplastic and membrane areas. Interestingly the ozone resistant accession PI 507656 was the only one that did not appear to down-regulate photosynthetic genes. Photosynthesis of this PI is robust against elevated ozone levels and, consequently, plant yields may be higher in an elevated-ozone atmosphere. It will be interesting to determine if this phenotype is inherited among progeny of this PI with a *G. max* parent.

The recurring enrichment of metal-ion binding genes, notably calcium and iron-sulfur complex was noted in both sensitive and resistant accessions. Fluctuation of metal ion binding genes is of note as sequestering of metal ions has been documented previously in the

quenching of reactive oxygen species (ROS), specifically when caused by heavy metal toxicity (Rodriguez-Serrano, 2009). Therefore, it could be hypothesized that ozone causes a similar response in wild soybean accessions. Interestingly, genes annotated with functions involved in reproductive processes such as pollen recognition were significantly up regulated in both sensitive genotypes and PI 424123. Though, as the plants had not yet reached reproductive development, these genes are likely part of some other related process or have been incorrectly annotated.

Gene expression analysis of resistant and sensitive cultivated soybean (*Glycine max*) - by Whaley et al. (2015) revealed cultivar-specific responses to ozone, and this study confirms a similar trend in soybean's wild relative. Similar to the study in *G. max*, more genes responded in sensitive accessions versus resistant counterparts; however, due to the differences in experimental design, these studies were not directly comparable. Whaley et al. noted an enrichment of wax and cutin biosynthetic genes and theorized that a difference in leaf architecture could be responsible for response to ozone; however, that same trend was not revealed from study in wild soybean. It could relate to cultivar-specific responses are the cause of this trend, or the difference could be due to the level of "low" ozone (27 ppb) used in that study compared to the CF treatment (<10 ppb) ozone used in this experiment.

It was clear that enrichments consistently occurred in extracellular spaces and that the genes connected to the apoplast in general were affected in most leaf positions and in every genotype tested. Links to cell walls, extracellular compartments, and apoplasts indicate that a significant amount of the response is occurring in the outer components of the plant cell. Apoplastic ROS quenching has been noted previously as a plant defense mechanism in

response to pathogen attack and to other abiotic stresses (Mittler, 2002). It is possible that wild soybean responds by inducing a series of reactions possibly aimed at inducing a hypersensitive response when ozone enters plant tissue to localize and prevent damage from spreading. In addition, the heatmap of gene response in the apoplast region included genes annotated as xyloglucan endotransglucosylase/hydrolase genes. Up-regulation of these genes has been linked to ozone damage in soybean previously and is thought to play a later role in pod dehiscence (Leisner, 2014). The down-regulation of these genes at this earlier developmental stage could possibly be a prevention of future yield losses.

CONCLUSION

Wild soybean shows a variety of responses when exposed to increased ozone levels. The most characteristic response of ozone resistance is that fewer genes are differentially expressed in response ozone than sensitive accessions. The correlation of genes expressed is much higher among the ozone-sensitive accessions than ozone-resistant accessions suggesting that there might be more than one mode of resistance to ozone. Multiple modes of ozone resistance are also supported by the fact that the transcriptome associated with the photosynthetic apparatus of PI507656 is not differentially expressed in response to ozone, while similar transcripts in the other resistant accession (PI 424123) are affected. Greater similarity exists between leaves of different maturities in sensitive genotypes, consistent with previous reports of leaf specific responses to ozone. Sources of identified resistance unique to wild soybean could become useful in developing soybean varieties capable of tolerating increases in ozone levels predicted for future years.

REFERENCES

- Ainsworth, E. A., Yendrek, C. R., Sitch, S., Collins, W. J., & Emberson, L. D. (2012). The effects of tropospheric ozone on net primary productivity and implications for climate change. *Annual Review of Plant Biology*, *63*, 637.
- Ainsworth, E. A. (2008). Rice production in a changing climate: A meta-analysis of responses to elevated carbon dioxide and elevated ozone concentration. *Global Change Biology*, *14*(7), 1642-1650.
- Ashmore, M. R. (2005). Assessing the future global impacts of ozone on vegetation. *Plant, Cell & Environment*, *28*(8), 949-964.
- Betzlberger, A. M., Gillespie, K. M., McGrath, J. M., Koester, R. P., Nelson, R. L., & Ainsworth, E. A. (2010). Effects of chronic elevated ozone concentration on antioxidant capacity, photosynthesis and seed yield of 10 soybean cultivars. *Plant, Cell & Environment*, *33*(9), 1569-1581.
- Burkey, K. O., & Carter Jr., T. E. (2009). Foliar resistance to ozone injury in the genetic base of U.S. and Canadian soybean and prediction of resistance in descendent cultivars using coefficient of parentage. *Field Crops Research*, *111*(3), 207-217.
- Burton, A. L., Burkey, K. O., Carter, T. E., Orf, J., & Cregan, P. B. (2016). Phenotypic variation and identification of quantitative trait loci for ozone tolerance in a Fiskeby III x Mandarin (Ottawa) soybean population. *Theoretical and Applied Genetics*. 1-13.
- Feng, Z., Kobayashi, K., & Ainsworth, E. A. (2008). Impact of elevated ozone concentration on growth, physiology, and yield of wheat (*Triticum aestivum* L.): A meta-analysis. *Global Change Biology*, *14*(11), 2696-2708.

- Fiscus, E. L., Booker, F. L., & Burkey, K. O. (2005). Crop responses to ozone: Uptake, modes of action, carbon assimilation and partitioning. *Plant, Cell & Environment*, 28(8), 997-1011.
- Gillespie, K. M., Xu, F., Richter, K. T., McGrath, J. M., Markelz, R. J. C., Ort, D. R., et al. (2012). Greater antioxidant and respiratory metabolism in field-grown soybean exposed to elevated O₃ under both ambient and elevated CO₂. *Plant, Cell & Environment*, 35(1), 169-184.
- Goff, L., Trapnell, C., & Kelley, D. cummeRbund: Analysis, exploration, manipulation, and visualization of cufflinks high-throughput sequencing data.
- Grimes, H. D., Perkins, K. K., & Boss, W. F. (1983). Ozone degrades into hydroxyl radical under physiological conditions: A spin trapping study. *Plant Physiology*, 72(4), 1016-1020.
- Heagle, A. S. (1979). Ranking of soybean cultivars for resistance to ozone using different ozone doses and response measures. *Environmental Pollution (1970)*, 19(1), 1-10.
- Heath, R. (1987). The biochemistry of ozone attack on the plasma membrane of plant cells. In J. Saunders, L. Kosak-Channing & E. Conn (Eds.), (pp. 29-54) Springer US.
- Heck, W., & Philbeck, R. (1978). In Dunning J. A., North Carolina Agricultural Experiment Station, Philbeck R. B. and United States. Agricultural Research Service. Southern Region. (Eds.), *A continuous stirred tank reactor (CSTR) system for exposing plants to gaseous air contaminants: Principles, specifications, construction, and operation*. New Orleans, La.: Dept. of Agriculture, Agricultural Research Service, Southern Region.

- Koiwa, H., Bressan, R. A., & Hasegawa, P. M. (1997). Regulation of protease inhibitors and plant defense. *Trends in Plant Science*, 2(10), 379-384.
- Kollist, T., Moldau, H., Rasulov, B., Oja, V., Rämme, H., Hüve, K., et al. (2007). A novel device detects a rapid ozone-induced transient stomatal closure in intact *Arabidopsis* and its absence in *Abi2* mutant. *Physiologia Plantarum*, 129(4), 796-803.
- Kusumi, K., Hirotsuka, S., Kumamaru, T., & Iba, K. (2012). Increased leaf photosynthesis caused by elevated stomatal conductance in a rice mutant deficient in *SLAC1*, a guard cell anion channel protein. *Journal of Experimental Botany*,
- Leisner, C. P., Ming, R., & Ainsworth, E. A. (2014). Distinct transcriptional profiles of ozone stress in soybean (*Glycine max*) flowers and pods. *BMC Plant Biology*, 14(1), 335.
- Martinez, I. M., & Chrispeels, M. J. (2003). Genomic analysis of the unfolded protein response in *Arabidopsis* shows its connection to important cellular processes. *The Plant Cell*, 15(2), 561-576.
- Mills, G., Hayes, F., Wilkinson, S., & Davies, W. J. (2009). Chronic exposure to increasing background ozone impairs stomatal functioning in grassland species. *Global Change Biology*, 15(6), 1522-1533.
- Mittler, R. (2002). Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science*, 7(9), 405-410.
- Moldau, H. (1998). Hierarchy of ozone scavenging reactions in the plant cell wall. *Physiologia Plantarum*, 104(4), 617-622.

- Morgan, P. B., Ainsworth, E. A., & Long, S. P. (2003). How does elevated ozone impact soybean? A meta-analysis of photosynthesis, growth and yield. *Plant, Cell & Environment*, 26(8), 1317-1328.
- Ozgun, R., Uzilday, B., Sekmen, A. H., & Turkan, I. (2015). The effects of induced production of reactive oxygen species in organelles on endoplasmic reticulum stress and on the unfolded protein response in *Arabidopsis*. *Annals of Botany*, 116(4), 541-553.
- Rodríguez-Serrano, M., Romero-Puertas, M., Pazmiño, D. M., Testillano, P. S., Risueño, M. C., del Río, L. A., et al. (2009). Cellular response of pea plants to cadmium toxicity: Cross talk between reactive oxygen species, nitric oxide, and calcium. *Plant Physiology*, 150(1), 229-243.
- SEQC/MAQC-III Consortium. (2014). A comprehensive assessment of RNA-Seq accuracy, reproducibility and information content by the sequencing quality control consortium. *Nat Biotech*, 32(9), 903-914.
- Sivasankar, S., Sheldrick B FAU - Rothstein, S.J., & Rothstein, S. J. (2000). *Expression of allene oxide synthase determines defense gene activation in tomato* Retrieved from 2000 Apr database.
- Stevenson, D. S., Dentener, F. J., Schultz, M. G., Ellingsen, K., Van Noije, T., Wild, O., et al. (2006). Multimodel ensemble simulations of present-day and near-future tropospheric ozone. *Journal of Geophysical Research. D. Atmospheres*, 111(D8).
- Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., & Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotech*, 31(1), 46-53.

- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics*, 25(9), 1105-1111.
- Van Dingenen, R., Dentener, F. J., Raes, F., Krol, M. C., Emberson, L., & Cofala, J. (2009). The global impact of ozone on agricultural crop yields under current and future air quality legislation. *Atmospheric Environment*, 43(3), 604-618.
- Volz, A., & Kley, D. (1988). Evaluation of the montsouris series of ozone measurements made in the nineteenth century. *Nature*, 332(6161), 240.
- Whaley, A., Sheridan, J., Safari, S., Burton, A., Burkey, K., & Schlueter, J. (2015). RNA-Seq analysis reveals genetic response and tolerance mechanisms to ozone exposure in soybean. *BMC Genomics*, 16, 426-015-1637-7.
- Wittig, V. E., Ainsworth, E. A., & Long, S. P. (2007). To what extent do current and projected increases in surface ozone affect photosynthesis and stomatal conductance of trees? A meta-analytic review of the last 3 decades of experiments. *Plant, Cell & Environment*, 30(9), 1150-1162.

TABLES and FIGURES

Table 3.1. Average ozone damage ratings for selected USDA *Glycine soja* diversity core collection. Average represents the averaged leaf injury score, while SE represents the associated standard error of the measurement for 3 replications. Lower values indicate resistance, while greater values indicate sensitivity. Accessions highlighted in green indicated those identified as sensitive or resistant and used to RNA-Seq sequenced in this study.

Accession	Ozone Rating	SE	Accession	Ozone Rating	SE
PI424116	1.36	0.36	PI378697A	3.38	0.13
PI245331	1.63	0.38	PI597448D	3.38	0.38
PI424123	1.75	0.25	PI507624	3.50	0.25
PI507656	2.13	0.13	PI522226	3.50	0.50
PI562551	2.25	0.50	PI549032	3.50	0.25
PI378690	2.38	0.13	PI424004B	3.73	0.07
PI407240	2.38	0.38	PI424082	3.61	0.17
PI407096	2.63	1.38	PI407059	3.63	0.13
PI369621	2.63	0.13	PI407157	3.63	0.13
PI339871A	2.77	0.23	PI407206	3.63	0.13
PI562561	2.88	0.63	PI522209B	3.63	0.13
PI424025B	2.88	0.88	PI597460A	3.63	0.13
PI639588B	2.88	0.38	PI407195	3.70	0.14
PI407038	3.00	0.50	PI407156	3.75	0.25
PI407228	3.00	0.50	PI407020	3.75	0.25
PI562553	3.00	0.50	PI424007	3.75	0.25
PI424102A	3.00	0.25	PI479752	3.75	0.25
PI407287	3.13	0.38	PI522233	3.75	0.25
PI407300	3.13	0.38	PI342622A	3.75	0.25
PI407314	3.13	0.13	PI424083A	3.75	0.00
PI407085	3.53	0.29	PI464890B	3.75	0.25
PI424035	3.60	-	PI407179	3.88	0.13
PI507761	3.13	0.13	PI407191	3.88	0.13
PI562547	3.13	0.13	PI424045	3.88	0.13
PI378696B	3.13	0.38	PI447003A	3.88	0.13
PI507581	3.25	0.25	PI522235B	3.88	0.13
PI507618	3.25	0.25	PI366122	4.00	0.00
PI378684B	3.25	0.50	PI407231	4.00	0.00
PI424070B	3.25	0.75	PI468916	4.00	0.00
PI549046	3.25	0.26	PI479768	4.00	0.00
PI407214	3.38	0.63	PI483466	4.00	0.00
PI407248	3.38	0.13	PI639623A	4.00	0.00
PI378686B	3.38	0.13	PI479746B	4.00	0.00

Table 3.2. Differential expression of selected *Glycine soja* genotypes following ozone exposure. The table lists the number of genes responding in both directions at least 2-fold between ozone treated and the control in the genotypes from each of the three leaf positions collected (T1, T2 and T3) as well as the total number in each genotype accounting for genes expressed in multiple leaves.

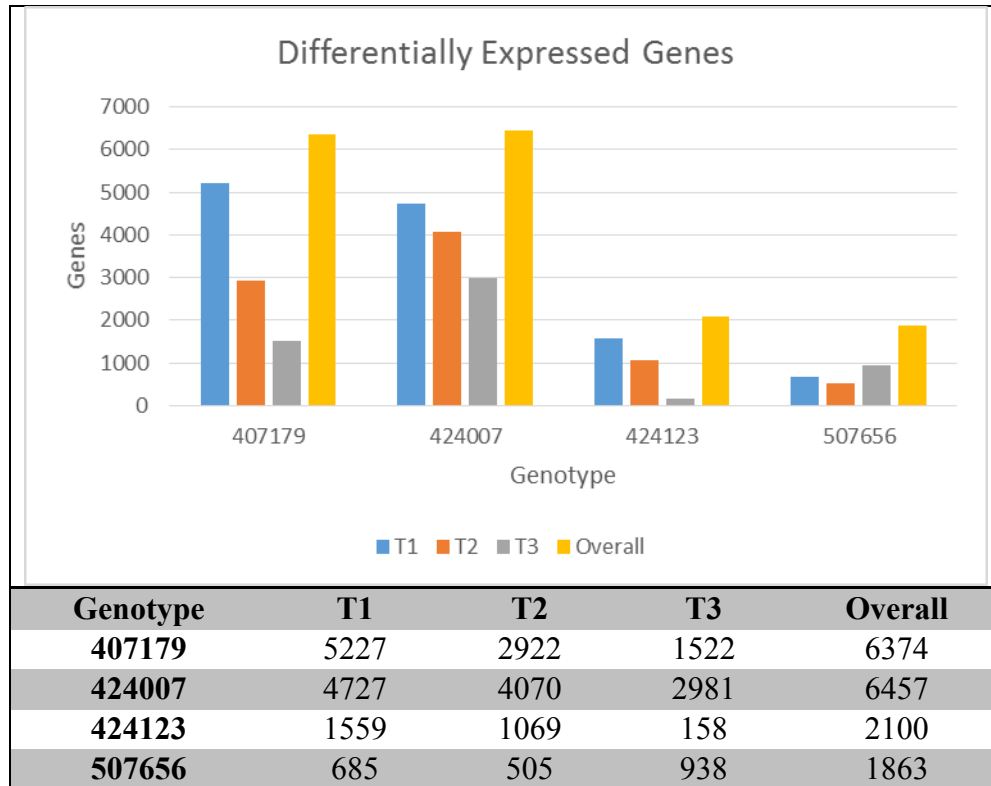


Table 3.3. Spearman correlation coefficients of log2-fold changes between Fragments Per Kilobase of transcript per Million mapped reads (FPKM)-values of CF and ozone-treated leaf samples both within genotype at each different leaf collection point as well as between the two sensitive and two resistant genotypes at each leaf position. Higher values indicate more similarity in gene response. Geno 1 & 2 represent the genotype for which the comparison is made.

Sensitive Genotypes			Resistant Genotypes		
Leaf/Geno 1	Leaf/Geno 2	Correlation	Leaf/Geno 1	Leaf/Geno 2	Correlation
T2/407179	T1/407179	0.522321	T1/424123	T3/424123	0.084234
T3/407179	T2/407179	0.38629	T2/424123	T1/424123	0.372723
T3/407179	T1/407179	0.37761	T2/424123	T3/424123	0.010379
T2/424007	T1/424007	0.536984	T1/507656	T2/507656	0.24135
T3/424007	T2/424007	0.473316	T1/507656	T3/507656	0.167825
T3/424007	T1/424007	0.444517	T2/507656	T3/507656	0.13003
T3/407179	T2/424007	0.312596	T1/507656	T1/424123	0.19454
T3/407179	T3/424007	0.29274	T3/507656	T2/424123	0.174712
T3/407179	T1/424007	0.28312	T1/424123	T3/507656	0.220342
T1/424007	T2/407179	0.435989	T1/424123	T2/507656	0.14186
T1/424007	T1/407179	0.430592	T2/424123	T2/507656	0.169758
T2/424007	T2/407179	0.462203	T2/424123	T1/507656	0.149973
T2/424007	T1/407179	0.452457	T3/424123	T3/507656	0.118728
T3/424007	T1/407179	0.424815	T3/424123	T1/507656	0.111249

Table 3.4. Filtered gene ontology categories for down-regulated genes in each genotype following exposure to increased ozone levels. Gene numbers are displayed under the column for designated by the corresponding genotype name. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Absence of a number in a column does not necessarily refer to zero responding genes but rather lack of significant enrichment (FDR < 0.05).

Category	Gene Ontology Description	Sensitive		Resistant	
		407179	424007	424123	507656
P	cell redox homeostasis	42	33	-	-
P	cellular glucan metabolic process	30	31	22	20
P	cellular homeostasis	44	34	-	-
P	heterocycle metabolic process	43	-	-	-
P	lipid biosynthetic process	59	48	26	25
P	lipid metabolic process	124	-	51	46
P	negative regulation of catalytic activity	19	17	-	-
P	nucleotide catabolic process	12	-	-	-
P	nucleotide metabolic process	33	-	-	-
P	oxidation reduction	263	199	116	100
P	photosynthesis	71	62	20	-
P	photosynthesis, light reaction	9	-	-	-
P	photosystem II stabilization	5	-	-	-
P	polysaccharide metabolic process	48	47	28	25
P	protein folding	28	-	-	-
P	protein polymerization	12	-	-	-
P	regulation of photosynthesis, light reaction	5	-	-	-
P	response to hormone (endogenous) stimulus	-	14	-	-
P	steroid metabolic process	-	22	-	11
F	alkylbase DNA N-glycosylase activity	6	-	5	-
F	anion transmembrane transporter activity	-	13	-	-
F	calcium ion binding	50	40	-	-
F	coenzyme binding	83	62	39	32
F	cofactor binding	97	-	44	39
F	electron carrier activity	94	68	35	32
F	endopeptidase inhibitor activity	-	-	18	19
F	enzyme inhibitor activity	-	-	20	23
F	enzyme regulator activity	-	-	23	24

Table 3.4 Continued.

Category	Gene Ontology Description	Sensitive		Resistant	
		407179	424007	424123	507656
F	FAD binding	32	-	21	17
F	fructose 1,6-bisphosphate 1-phosphatase activity	7	6	-	-
F	heme binding	-	-	33	34
F	hydrolase activity of glycosyl bonds	98	81	45	53
F	hydrolase activity, hydrolyzing O-glycosyl compounds	88	73	40	47
F	identical protein binding	20	19	-	-
F	iron ion binding	89	-	34	35
F	iron-sulfur cluster binding	23	20	-	-
F	O-methyltransferase activity	-	-	-	7
F	oxidoreductase activity	306	236	130	111
F	oxidoreductase activity, acting on CH-OH group of donors	57	-	26	26
F	oxidoreductase activity acting on NADH or NADPH	8	-	-	-
F	oxidoreductase activity acting on paired donors	85	-	-	33
F	oxidoreductase activity, acting on sulfur group of donors	20	-	-	-
F	oxidoreductase activity, acting on the aldehyde or oxo group of donors	-	-	-	6
F	oxidoreductase on CH-OH donors	50	-	-	23
F	protein disulfide oxidoreductase activity	18	15	-	-
F	serine hydrolase activity	42	38	-	-
F	serine-type endopeptidase activity	26	-	-	-
F	sigma factor activity	-	6	-	-
F	steroid dehydrogenase activity acting on the CH-OH group of donors	22	21	-	11
F	sulfate transmembrane transporter activity	-	9	-	-
F	tetrapyrrole binding	77	-	34	34
F	transferase activity, transferring acyl groups	53	-	27	-
F	transferase activity, transferring acyl groups other than amino-acyl groups	46	-	24	-

Table 3.4 Continued.

Category	Gene Ontology Description	Sensitive		Resistant	
		407179	424007	424123	507656
F	transferase activity, transferring glycosyl groups	82	79	41	39
F	translation termination factor activity	6	-	-	-
F	UDP-N-acetylmuramate dehydrogenase activity	-	-	7	8
F	xyloglucan:xyloglucosyl transferase activity	23	22	18	16
C	apoplast	23	22	18	16
C	cell wall	34	31	22	21
C	chloroplast	15	15	-	-
C	cytochrome b6f complex	5	-	-	-
C	extracellular region	31	27	24	20
C	oxygen evolving complex	32	27	8	-
C	photosynthetic membrane	68	59	20	-
C	photosystem I	17	18	6	-
C	photosystem I reaction center	8	7	-	-
C	photosystem II	47	40	13	-
C	thylakoid	73	61	21	-

Table 3.5. Filtered gene ontology categories for up-regulated genes in each genotype following exposure to increased ozone levels. Gene numbers are displayed under the column for designated by the corresponding genotype name. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Absence of a number in a column does not necessarily refer to zero responding genes but rather lack of significant enrichment (FDR < 0.05).

Category	Gene Ontology Description	Sensitive		Resistant	
		407179	424007	424123	507656
P	aminoglycan metabolic process	-	6	-	5
P	carboxylic acid biosynthetic process	-	33	-	-
P	cell communication	33	44	-	-
P	cell recognition	29	42	12	9
P	cell wall macromolecule metabolic process	12	20	-	8
P	cellular nitrogen compound biosynthetic process	-	35	-	-
P	chitin metabolic process	-	6	-	5
P	chlorophyll metabolic process	-	10	-	-
P	chorismate metabolic process	-	9	-	-
P	defense response	17	24	-	10
P	dicarboxylic acid metabolic process	-	9	-	-
P	DNA-dependent DNA replication	-	5	-	-
P	lipid biosynthetic process	47	54	-	20
P	oxidation reduction	163	218	-	-
P	phosphate metabolic process	231	325	79	71
P	pigment metabolic process	-	13	-	-
P	polysaccharide catabolic process	-	-	-	5
P	protein ubiquitination	21	26	-	-
P	recognition of pollen	29	42	12	9
P	response to biotic stimulus	14	17	-	9
F	ammonia-lyase activity	5	8	-	-
F	ATPase activity	-	-	24	-
F	calcium ion binding	45	69	24	-
F	chitinase activity	-	6	-	5
F	coenzyme binding	-	67	-	-
F	cofactor binding	66	85	-	-
F	endoribonuclease activity	-	-	8	-
F	fatty-acid synthase activity	8	7	-	-
F	galactosyltransferase activity	-	10	-	-

Table 3.5 Continued.

Category	Gene Ontology Description	Sensitive		Resistant	
		407179	424007	424123	507656
F	heme binding	-	73	-	-
F	intramolecular lyase activity	8	9	-	-
F	O-methyltransferase activity	-	17	-	8
F	oxidoreductase activity	-	250	-	-
F	oxidoreductase activity, acting on CH-OH group of donors	45	57	-	18
F	oxidoreductase activity, acting on NADH or NADPH	-	-	7	-
F	oxidoreductase activity, acting on NADH or NADPH, with oxygen as acceptor	-	6	5	-
F	oxidoreductase activity, acting on the CH-OH group of donors	43	55	-	17
F	P-P-bond-hydrolysis-driven transmembrane transporter activity	-	-	10	-
F	polysaccharide binding	14	23	-	9
F	primary active transmembrane transporter activity	-	-	10	-
F	protein serine/threonine kinase activity	-	-	5	-
F	ribonuclease III activity	-	-	5	-
F	sequence-specific DNA binding	72	99	32	30
F	steroid dehydrogenase activity, acting on the CH-OH group of donors	21	25	-	15
F	transcription regulator activity	105	138	-	39
F	transferase activity, transferring acyl groups	41	-	-	-
F	transferase activity, transferring acyl groups other than amino-acyl groups	34	-	-	-
F	ubiquitin-protein ligase activity	35	43	-	-
F	unfolded protein binding	-	16	9	6
C	endoplasmic reticulum	-	21	-	-
C	extracellular matrix	-	12	-	-
C	ubiquitin ligase complex	35	43	-	-

Table 3.6. List of specific differentially-expressed genes shared by both resistant genotypes (PIs 424123 and 507656) but by neither sensitive genotype (PIs 407179 and 424007). Genes in this list were all down-regulated; no unique genes were up-regulated and shared by both resistant genotypes. Genes without a known annotation were excluded from this table.

Gene Model	Gene Ontology Biological Process Descriptions
Glyma.01G209500	cell wall modification; "histone phosphorylation"
Glyma.03G005400	cysteine biosynthetic process; "drug transmembrane transport"; "response to symbiotic fungus"; "transmembrane transport"
Glyma.03G107500	biological process
Glyma.03G246800	"detection of calcium ion"; "mitochondrion localization"; "peroxisome localization"; "response to mechanical stimulus"
Glyma.04G027800	anthocyanin accumulation in tissues in response to UV light;
Glyma.06G270500	microtubule-based process
Glyma.06G303200	carboxylic acid metabolic process; "metabolic process"; "seed dormancy process"
Glyma.08G114700	biological process, iron ion binding
Glyma.08G189600	defense response
Glyma.08G190700	transmembrane transport; "transport"
Glyma.08G340200	MAPK cascade; "detection of biotic stimulus";
Glyma.08G341100	defense response to bacterium; "nitrate transport"; "programmed cell death";
Glyma.08G341400	nitrate transport; "protein folding"; "response to endoplasmic reticulum stress"
Glyma.09G014100	regulation of transcription, DNA-dependent
Glyma.09G220600	defense response to bacterium; "response to auxin stimulus"; "response to cold"
Glyma.10G021400	coumarin biosynthetic process; "response to wounding"
Glyma.10G252800	cellular response to nitrogen starvation; "protein phosphorylation"
Glyma.11G022300	metabolic process
Glyma.13G042200	gibberellin metabolic process; "oxidation-reduction process"
Glyma.13G203900	brassinosteroid biosynthetic process
Glyma.14G140900	response to abscisic acid stimulus; "response to cold"; "response to desiccation"; "response to salt stress"
Glyma.15G015100	innate immune response; "response to fungus"
Glyma.15G229400	circadian rhythm; "response to abscisic acid stimulus"; "response to salt stress"
Glyma.17G072200	cell wall biogenesis; "cell wall macromolecule metabolic process"; "cellulose biosynthetic process"
Glyma.17G234700	actin filament bundle assembly; "cellular membrane fusion"
Glyma.17G065400	xyloglucan:endo-transglucosylase activity

Table 3.7. List of genes that were down regulated in both resistant genotypes and up regulated in both sensitive genotypes.

Gene Name	Gene Ontology Biological Process Descriptions
Glyma.01G056100	"oxidation-reduction process"; "response to abscisic acid stimulus"
Glyma.01G126600	N-terminal protein myristoylation; "defense response";
Glyma.01G217600	"defense response"; "incompatible interaction"; "response to salt stress"
Glyma.03G044900	"defense response"; "lignan biosynthetic process"
Glyma.03G046000	N-terminal protein myristoylation; "defense response"
Glyma.03G247500	"defense response" "systemic acquired resistance"
Glyma.08G235400	"programmed cell death"; "response to hydrogen peroxide"
Glyma.09G049100	oxidation-reduction process
Glyma.09G281800	cellular modified amino acid biosynthetic process; "response to wounding"
Glyma.10G161500	biological process
Glyma.15G128700	"response to oxidative stress"; "response to salt stress"; "response to zinc ion"
Glyma.15G203500	cellular response to iron ion; "oxidation-reduction process"; "response to zinc"
Glyma.15G211500	"protein folding"; "response to ER stress"; "response to hydrogen peroxide"
Glyma.16G019900	"cell wall macromolecule metabolic process"; "oxidation-reduction process"
Glyma.17G030400	defense response; "mRNA modification"
Glyma.20G205800	defense response; "proline transport"; "response to wounding"

Table 3.8. Filtered gene ontology categories for differentially-expressed genes in sensitive accession PI 407179 following exposure to increased ozone levels. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Shared by all positions signifies genes differentially expressed in each of the 3 trifoliates, while trifoliolate subheadings refer to genes that were only expressed at that position. “DE” refers to significant genes pooled in either direction, while “Up” and “Down” refer to the direction of genes matching each category when run separately.

PI 407179				
Category	Gene Ontology Description	DE	Up	Down
Shared by All Leaf Positions				
F	ATPase activity	20	17	-
F	ATPase activity, coupled to transmembrane movement of substances	9	7	-
F	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement	9	-	-
F	oxidoreductase activity	68	-	31
F	P-P-bond-hydrolysis-driven transmembrane transporter activity	9	7	-
F	transferase activity, transferring other than amino-acyl groups	16	-	-
F	iron ion binding	-	-	13
1st Trifoliolate				
P	cell redox homeostasis	32	-	29
P	cellular homeostasis	35	-	30
P	oxidation reduction	172	-	136
P	photosynthesis	45	-	42
F	iron-sulfur cluster binding	15	-	14
F	oxidoreductase activity	198	-	160
F	translation termination factor activity	6	-	6
C	chloroplast	11	-	11
C	extrinsic to membrane	20	-	19
C	oxygen evolving complex	19	-	19
C	photosynthetic membrane	42	-	41
C	photosystem I	15	-	15
C	photosystem I reaction center	7	-	7
C	photosystem II	25	-	24
C	thylakoid	45	-	44
P	protein folding	-	-	18

Table 3.8 Continued.

PI 407179				
Category	Gene Ontology Description	DE	Up	Down
2nd Trifoliolate				
P	protein amino acid phosphorylation	48	30	-
P	phosphate metabolic process	49	31	-
P	post-translational protein modification	50	32	-
F	protein kinase activity	48	30	-
F	phosphotransferase activity, alcohol group as acceptor	48	30	-
F	transferase activity, transferring phosphorus-containing groups	50	31	-
C	cell wall	-	-	7
F	hydrolase activity, hydrolyzing O-glycosyl compounds	-	-	14
3rd Trifoliolate				
P	defense response	8	8	-
P	response to biotic stimulus	7	7	-

Table 3.9. Filtered gene ontology categories for differentially-expressed genes in sensitive accession PI 424007 following exposure to increased ozone levels. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Shared by all positions signifies genes differentially expressed in each of the 3 trifoliates, while trifoliolate subheadings refer to genes that were only expressed at that position. “DE” refers to significant genes pooled in either direction, while “Up” and “Down” refer to the direction of genes matching each category when run separately.

PI 424007				
Category	Gene Ontology Description	DE	Up	Down
Shared by All Leaf Positions				
P	recognition of pollen	25	25	-
P	protein amino acid phosphorylation	146	123	-
F	calcium ion binding	40	31	-
P	phosphate metabolic process	151	128	-
F	phosphotransferase activity, alcohol group as acceptor	156	132	-
F	3-beta-hydroxy-delta5-steroid dehydrogenase activity	21	16	-
P	lipid biosynthetic process	42	31	-
P	steroid biosynthetic process	21	16	-
F	oxidoreductase activity, acting on CH-OH group of donors	39	29	-
F	oxidoreductase activity	150	107	-
F	coenzyme binding	48	33	-
P	oxidation reduction	133	95	-
F	intramolecular lyase activity	6	5	-
P	response to biotic stimulus	11	11	-
P	cell communication	25	25	-
F	ammonia-lyase activity	5	5	-
F	sequence-specific DNA binding	51	49	-
F	transferase activity, transferring acyl groups other than amino-acyl groups	29	21	-
F	xyloglucan:xyloglucosyl transferase activity	10	-	10
C	extracellular region	17	-	13
F	cofactor binding	52	-	-
C	apoplast	10	-	10
F	O-methyltransferase activity	11	11	-
F	transferase activity, transferring acyl groups	32	-	-
P	lipid metabolic process	59	-	-
P	cell wall macromolecule metabolic process	9	9	-

Table 3.9 Continued.

PI 424007				
Category	Gene Ontology Description	DE	Up	Down
P	cellular carbohydrate metabolic process	42	-	19
F	carbon-nitrogen lyase activity	6	6	-
P	defense response	12	12	-
1st Trifoliolate				
P	photosynthesis	27	-	24
C	photosynthetic membrane	23	-	23
C	photosystem I	11	-	11
C	photosystem II	12	-	12
C	chloroplast	6	-	5
F	anion transmembrane transporter activity	9	-	9
F	phosphotransferase activity, phosphate group as acceptor	5	-	-
F	sulfate transmembrane transporter activity	7	-	7
P	anion transport	10	-	10
P	response to hormone (endogenous) stimulus	8	-	-
P	porphyrin biosynthetic process	7	-	-
2nd Trifoliolate				
C	photosystem II	6	-	6
P	amino acid transport	-	5	-
P	photosynthesis	-	-	8
3rd Trifoliolate				
P	cellular glucan metabolic process	10	-	9
F	copper ion binding	11	-	9
F	transferase activity, transferring glycosyl groups	25	-	21
F	transferase activity, transferring hexosyl groups	23	-	19
F	xyloglucan:xyloglucosyl transferase activity	7	-	7
C	apoplast	7	-	7
C	chloroplast	5	-	5
F	protein kinase activity	-	31	-
F	ribonucleotide binding	-	47	-
C	extracellular region	10	-	8

Table 3.10. Filtered gene ontology categories for differentially-expressed genes in resistant accession PI 424123 following exposure to increased ozone levels. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Shared by all positions signifies genes differentially expressed in each of the 3 trifoliates, while trifoliolate subheadings refer to genes that were only expressed at that position. “DE” refers to significant genes pooled in either direction, while “Up” and “Down” refer to the direction of genes matching each category when run separately.

PI 424123				
Category	Gene Ontology Description	DE	Up	Down
Shared by All Leaf Positions				
F	transferase activity, transferring hexosyl groups	5	-	-
1st Trifoliolate				
P	oxidation reduction	73	-	59
P	photosynthesis	18	-	18
F	calcium ion binding	18	-	-
F	endopeptidase inhibitor activity	10	-	10
C	extracellular matrix	5	-	-
C	extrinsic to membrane	8	-	8
C	photosystem II	13	-	13
C	thylakoid	19	-	19
2nd Trifoliolate				
F	coenzyme binding	16	-	12
F	endoribonuclease activity, producing 5'-phosphomonoesters	5	5	-
F	oxidoreductase activity, acting on CH-OH group of donors	15	-	10
F	ATP binding		36	-
F	UDP-N-acetylmuramate dehydrogenase activity	6	-	-
C	extracellular region	8	-	5
3rd Trifoliolate				
P	oxidation reduction	14	-	12
F	oxidoreductase activity	15	-	13

Table 3.11. Filtered gene ontology categories for differentially-expressed genes in resistant accession PI 507656 following exposure to increased ozone levels. Category refers to the Gene Ontology classification category: biological process (P), molecular function (F), or cellular component (C). Shared by all positions signifies genes differentially expressed in each of the 3 trifoliates, while trifoliolate subheadings refer to genes that were only expressed at that position. “DE” refers to significant genes pooled in either direction, while “Up” and “Down” refer to the direction of genes matching each category when run separately.

PI 507656				
Category	Gene Ontology Description	DE	Up	Down
First Trifoliolate				
P	cellular glucan metabolic process	14	-	14
P	oxidation reduction	45	-	44
F	endopeptidase inhibitor activity	7	-	8
F	FAD binding	10	-	9
F	hydrolase activity, acting on glycosyl bonds	32	-	32
F	oxidoreductase activity	51	-	50
F	transferase activity, transferring glycosyl groups	24	-	24
F	xyloglucan:xyloglucosyl transferase activity	11	-	11
C	apoplast	11	-	11
C	cell wall	12	-	12
C	extracellular region	14	-	13
Second Trifoliolate				
P	lipid metabolic process	19	-	21
P	oxidation reduction	42	-	45
F	coenzyme binding	12	-	14
F	FAD binding	7	-	7
F	iron ion binding	17	-	17
F	oxidoreductase activity	45	-	49
F	oxidoreductase activity acting on paired donors	17	-	-
F	oxidoreductase activity, acting on CH-OH group of donors	10	-	13
F	tetrapyrrole binding	15	-	17
F	transferase activity, transferring glycosyl/hexosyl groups	13	-	-
C	extracellular region	6	-	-

Table 3.11. Continued.

PI 507656				
Category	Gene Ontology Description	DE	Up	Down
Third Trifoliolate				
P	aminoglycan metabolic process	5	5	-
P	cell wall macromolecule metabolic process	7	7	-
P	chitin metabolic process	5	5	-
P	defense response	10	10	-
P	lipid biosynthetic process	20	18	-
P	phosphate metabolic process	73	64	-
P	protein amino acid phosphorylation	72	63	-
P	response to biotic stimulus	9	-	-
P	steroid metabolic process	12	-	-
F	O-methyltransferase activity	9	8	-
F	hydrolase activity, acting on ester bonds	-	-	17

Resistant Genotypes

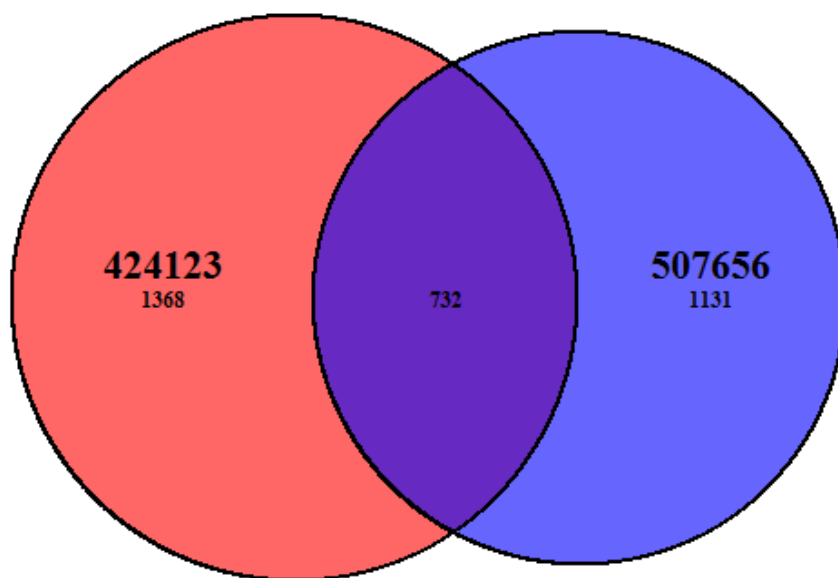


Figure 3.1. Venn diagram showing comparison of number of genes differentially expressed between ozone-resistant genotypes 424123 and 507656.

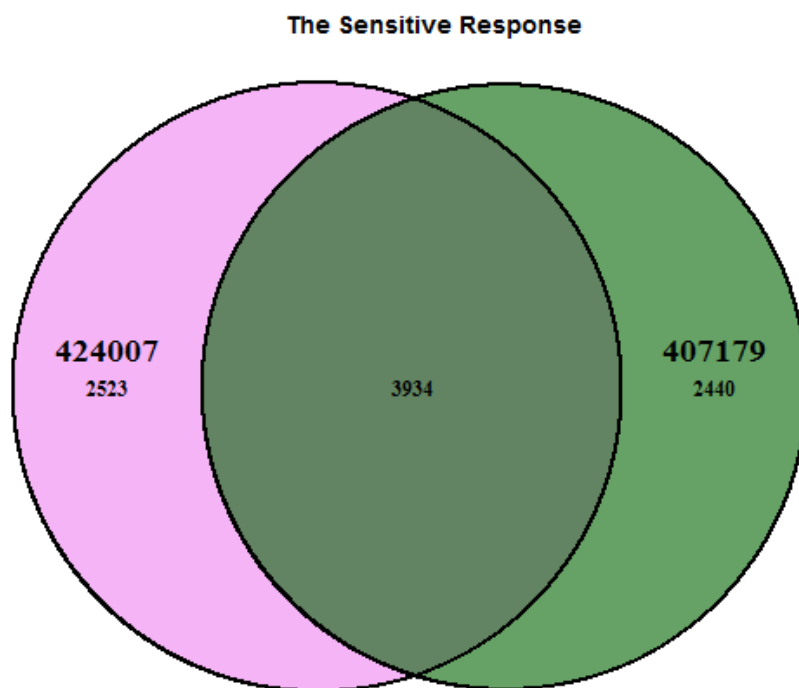


Figure 3.2. Venn diagram showing comparison of number of genes differentially expressed between ozone-sensitive genotypes 407179 and 424007.

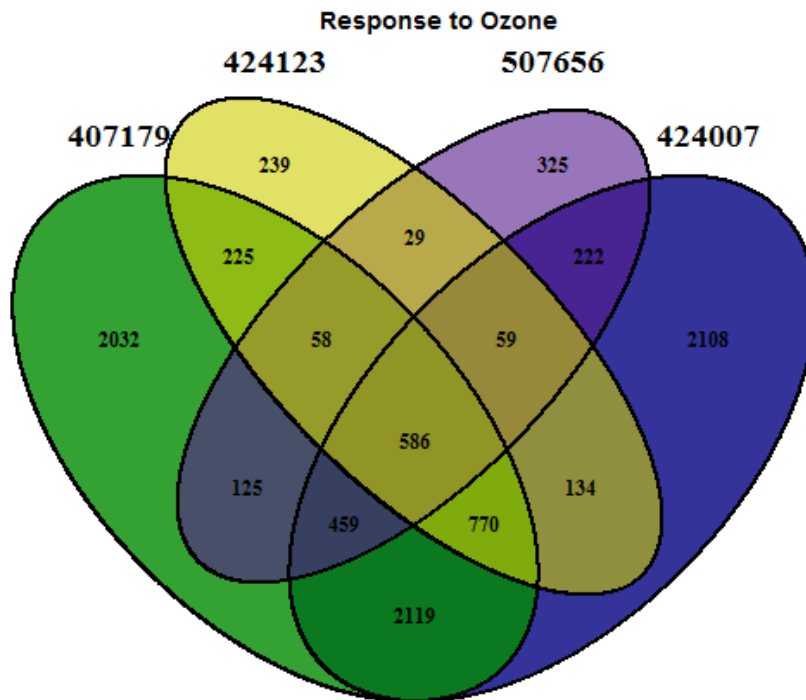


Figure 3.3. Venn diagram showing comparison of number of genes differentially expressed between all genotypes used in the *Glycine soja* Ozone Study. This diagram shows several interesting numbers such as 2,119 genes shown to be uniquely expressed by both sensitive genotypes but not by either of the resistant genotypes and that 29 genes are shared by both of the resistant sources but neither sensitive genotype.

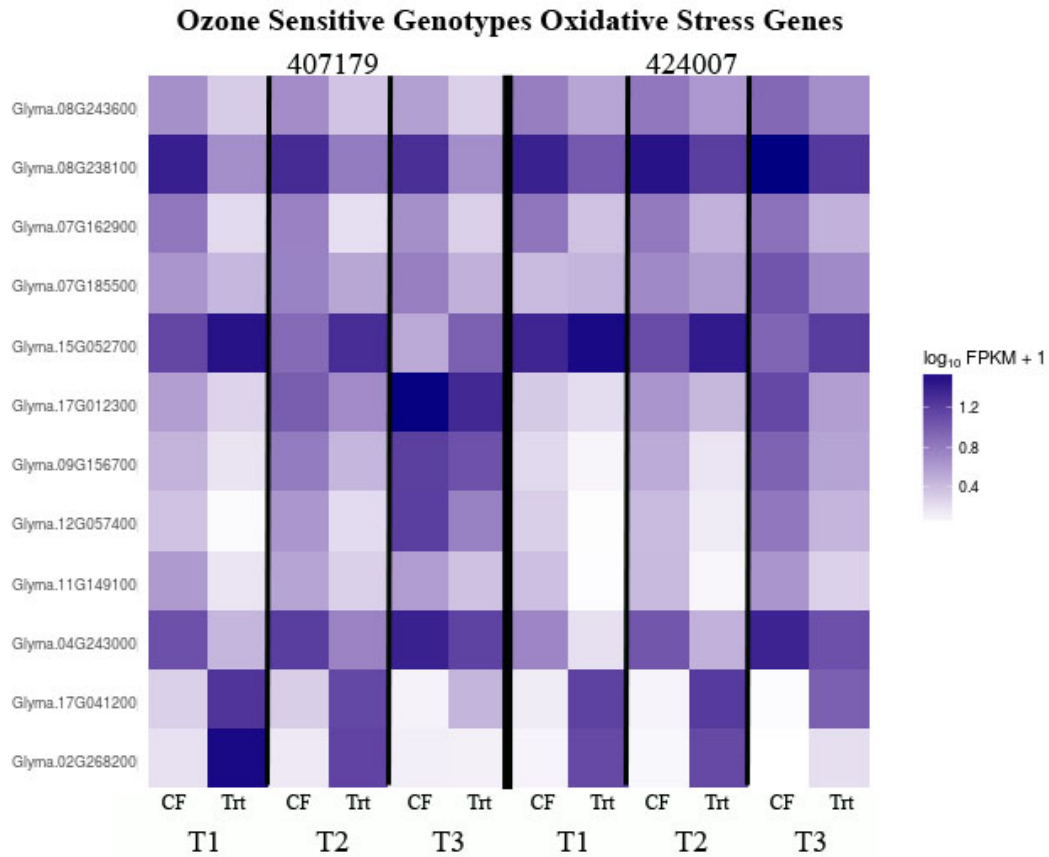


Figure 3.4. Heatmap of FPKM values for selected oxidative stress genes in ozone sensitive genotypes with twelve differentially-expressed genes related to oxidation/reduction reactions in ozone sensitive genotypes PIs 407179 and 424123 at each trifoliolate (T1, T2, T3). CF indicates a sample that was controlled in charcoal-filtered conditions, while Trt signifies exposure to 75 ppb ozone. Level of expression (FPKM) is noted by the intensity of color in each cell with high expression being dark purple and low expression being white. Glyma.08G238100 and Glyma.08G243600 refer to iron- and heme-binding genes, Glyma.09G156700 and Glyma.15G052700 are genes with peroxidase activity, Glyma.02G269200 and Glyma.17G041200 are genes connected to ATP binding and movement of substances, Glyma.17G012300 is a gene for copper ion binding, Glyma.12G057400 is annotated a gene for NAD activity, Glyma.11G149100 is linked with catalytic activity, Glyma.07G185500 relates to FAD binding, and Glyma.04G243000 and Glyma.07G1929100 are not annotated beyond oxidoreductase activity.

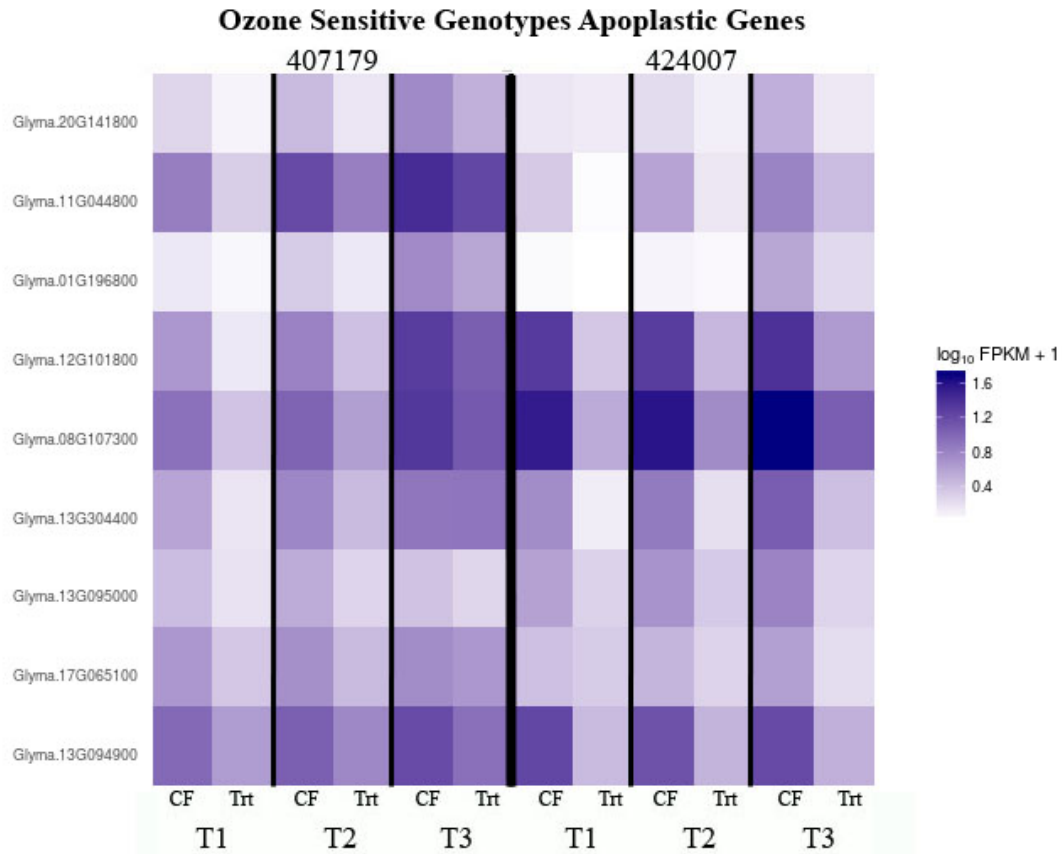


Figure 3.5. Heatmap of FPKM values for selected apoplastic genes in ozone sensitive genotypes of nine differentially-expressed genes in the apoplasts of ozone sensitive genotypes PIs 407179 and 424007 at each trifoliolate (T1, T2, T3). These genes correspond to genes thought to code for xyloglucan endotransglucosylases/hydrolases. CF indicates a sample that was controlled in charcoal-filtered conditions, while Trt signifies exposure to 75 ppb ozone. Level of expression (FPKM) is noted by the intensity of color in each cell with high expression being dark purple and low expression being closer to white.

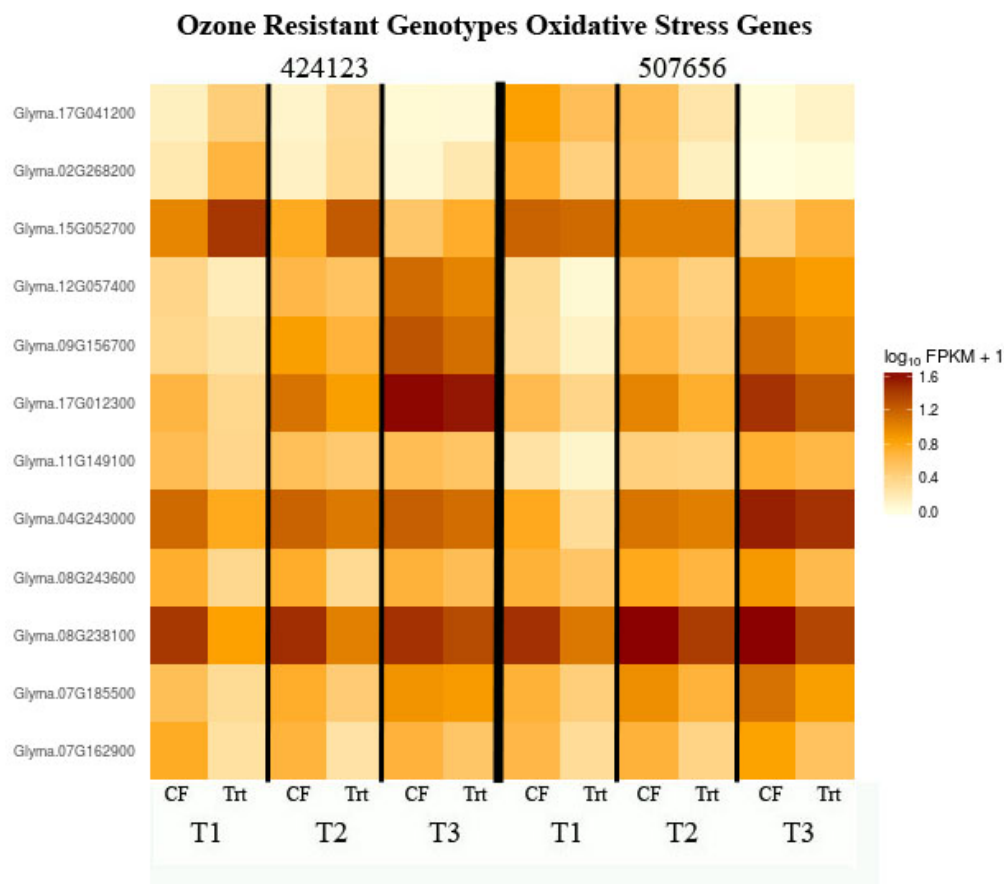


Figure 3.6. Heatmap of FPKM Values for selected oxidative stress genes in ozone resistant genotypes with twelve differentially-expressed genes related to oxidation/reduction reactions in ozone resistant genotypes PIs 424123 and 507656 at each trifoliolate (T1, T2, T3). CF indicates a sample that was controlled in charcoal-filtered conditions, while Trt signifies exposure to 75 ppb ozone. Level of expression (FPKM) is noted by the intensity of color in each cell with high expression being dark red and low expression being light yellow. Glyma.08G238100 and Glyma.08G243600 refer to iron- and heme-binding genes, Glyma.09G156700 and Glyma.15G052700 are genes with peroxidase activity, Glyma.02G269200 and Glyma.17G041200 are genes connected to ATP binding and movement of substances, Glyma.17G012300 is a gene for copper ion binding, Glyma.12G057400 is annotated a gene for NAD activity, Glyma.11G149100 is linked with catalytic activity, Glyma.07G185500 relates to FAD binding, and Glyma.04G243000 and Glyma.07G1929100 are not annotated beyond oxidoreductase activity.

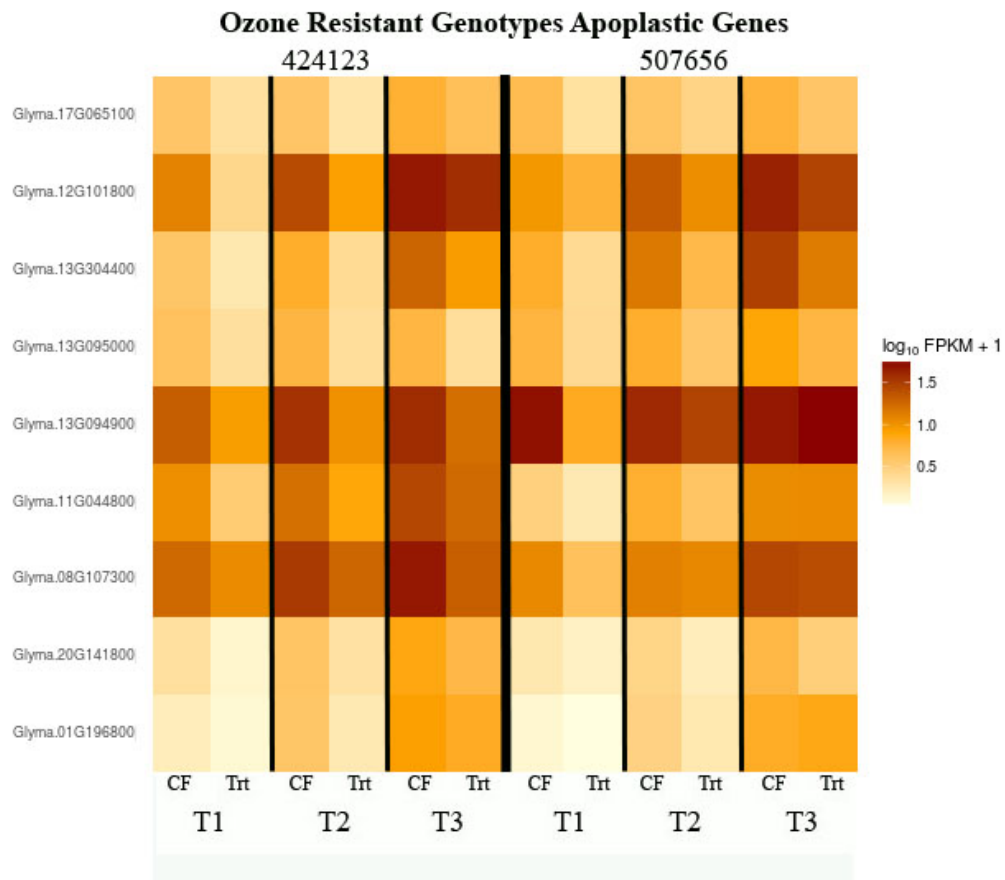


Figure 3.7. Heatmap of FPKM values for selected apoplastic genes in ozone resistant genotypes with nine differentially-expressed genes in the apoplasts of ozone resistant genotypes PIs 424123 and 507656 at each trifoliolate (T1, T2, T3). These genes have connections to xyloglucan:xyloglucosyl transferase activity. CF indicates a sample that was controlled in charcoal-filtered conditions, while Trt signifies exposure to 75 ppb ozone. Level of expression (FPKM) is noted by the intensity of color in each cell with high expression being dark red and low expression being light yellow.

**Chapter IV: Incorporation of Abiotic Stress Resistance from *Glycine soja* into
Germplasm Releases for Soybean Breeding**

ABSTRACT

Phenotypic evaluations of genetic materials are only a first step in germplasm improvement. Application of this information to applied plant breeding results in improved genetic materials. This study incorporates phenotypic ozone response in wild soybean with publically available genetic marker information to identify a putative marker for ozone response on Chromosome 19 in soybean. This chapter also highlights the process for creation of crosses between wild and domesticated soybean to produce soybean germplasm resources with increased genetic diversity. These crosses can also be used for study of heritability of traits from either wild or domesticated parental lines that will be useful in further development of genetically diverse soybean lines. This study suggests a correlation between ancestral and progeny seed sizes as well as selection of agronomically suitable *G. max* x *G. soja* based on seed size.

INTRODUCTION

Soybean is an early example of crop domestication and has been bred for improved yields, seed compositions, and stress resistance. However, domestication and breeding efforts have created several genetic bottlenecks for soybean (Carter et al., 2004). As a result, 84% of the genetic base of elite North American soybean cultivars can be traced back to only seventeen ancestral lines, while an estimated 81% of rare alleles have been lost. It is believed that these lost alleles could help in future breeding efforts (Gizlice et al., 1994; Hyten et al., 2006). Concerns have been raised that soybean improvement will reach plateaus due to lack of genetic variability; thus, expansion of soybean germplasm resources has become of recent interest (Egli, 2008).

The putative wild ancestor of soybean, *Glycine soja* Succ & Zieb, is a source of much more diversity than cultivated soybean due to its natural adaptation and freedom from artificial selection (Singh and Hymowitz, 1999). Examples of favorable alleles identified in wild soybean include those associated with yield, increased seed protein, and stress resistance (Diers et al., 1992; Li et al., 2008; Wang et al., 2007; Luo et al., 2005). In addition, *G. soja* is part of the soybean primary gene pool and is capable of easily producing fertile offspring with *G. max* (Singh and Hymowitz, 1999).

While a potential source of genetic improvement, wild soybean also brings with it unfavorable traits such as vining growth habit, lodging susceptibility, lack of leaf abscission at maturity, as well as small, hard black seeds (Carpenter and Fehr, 1986). Many attempts have been made to backcross wild soybean with a recurrent cultivated soybean and select against undesirable traits from wild soybean; however, these strategies indirectly select against desirable wild parent alleles, as well. Delheimer (2012) developed a method that, while space intensive, works to maximize genetic diversity retained following crosses between wild and cultivated soybean. This method creates “mega populations” of progeny and does not include any backcrossing to the cultivated soybean parent, which retains more alleles donated from *G. soja*. Through this process, a set of core lines can be created to provide sources of new alleles in a more favorable phenotype as germplasm resources.

As part of a grant aimed at increasing soybean genetic diversity funded by the United Soybean Board, 806 *G. soja* accessions from the USDA Collection were identified as being >.01% genetically different than any other accession in the collection. From these accessions, clusters of ~10 were created and eventually a 10% core collection (81 accessions) was

chosen by Dr. Qiagin Song as being the most genetically diverse (personal comm, 2014) – referred to as the “diversity core” throughout this chapter. These accessions can be characterized and used for trait identification to assess their use in improving genetic diversity in soybean. Additionally, progeny derived from crosses between *G. max* and members of the wild soybean diversity core are in various stages of development.

In this study, we work to utilize the phenotypic variation we have identified in the wild soybean “diversity core” that we have presented in other chapters. Subsequently, we create genetically diverse populations for use in soybean breeding. Independent of phenotypic evaluations for response to abiotic stressors, several crosses to increase genetic diversity were made; some of these crosses can now be connected with associated stress response observations. Development of a population now connected with observations from ozone response is followed in this study.

MATERIALS and METHODS

Genome Wide Association Study for Ozone Resistance

The same ozone damage ratings of 70 of the relevant Southern subset of the diversity core used to identify accessions resistant to ozone (**Table 3.1**) were used as input of phenotypic data into TASSEL 5 (Bradbury et al., 2007). Publically available Soybean 50K SNP data for each of the lines screened for ozone damage (Song et al., 2015) was used as the genotypic input. Markers were filtered in TASSEL to require a minor allele frequency of ≥ 0.1 ; a total of 28,907 markers were retained. Following filtering, genotypic information was output in a text file and loaded into JMP Genomics 7 (SAS Institute, Cary, NC) for Tag SNP Analysis based on linkage disequilibrium to identify markers with correlations of ≤ 0.9 . A

distinct set of 23,749 markers was returned to TASSEL, where a principal components analysis was done using 2 principal components. A general linear model (GLM) with 1,000 permutations was run using ozone ratings, raw marker data, and principal components in TASSEL; Manhattan plots showing markers of significance were subsequently produced. Further analysis of markers was done with available tools on Soybase (www.soybase.org)

Development and Progression of Genetically Diverse Soybean Lines

Population 23 was created in 2012; NC Raleigh was chosen as the maternal parent because of its overall vigor, while PI 424007 was initially chosen for its diverse genotype (Taliercio personal comm., 2014), F₁ seeds from successful crosses in Clayton, NC were collected and sent to the USDA-ARS Tropical Agricultural Research Station (TARS) winter nursery in Isabella, Puerto Rico for generation advancement from December 2012 – April 2013. The F₂ seeds were then scarified using sulfuric acid (H₂SO₄) and planted in 3 row plots grouped by original F₁ in Hugo, NC in June 2013. The F₃ progeny were then bulk harvested using a plot combine and packaged again according to original F₁. Seeds were then sieved into 5 seed size classes (>12/64”, < 12/64”, < 11.5/64”, < 11/64” & < 8.75/64”).

In June 2014, the seed was planted in replicate in 1.2-hectare plots in Rock Mount, NC and Clayton, NC blocked according to original F₁ and again by seed size. Single plant selections were made among mature F₃ plants based on uprightness and lack of shattering. An example of this selection process can be seen in **Figure 4.1**. In spring 2015, F_{3:4} seeds were planted in progeny rows in Clayton, NC, and single plant selections were made based on uprightness and general vigorous appearance. F_{4:5} selections were planted in progeny rows in Clayton, NC in spring 2016 and will be harvested for further study. The current breeding

scheme can be seen in **Figure 4.2**. Subsequent data analysis of variance for association between original sieve size and 100-seed weight of current progeny was done in SAS 9.4 (Cary, NC) using PROC GLM with the PLOTS = ALL statement. Visualization of the number of selections based on original sieve size was done using R 3.3.0.

RESULTS

Genome Wide Association Study for Ozone Resistance

Evaluation of the population structure based on allele frequencies at 23,749 polymorphic SNP markers by principal components analysis showed a 23.53% explanation of variance from the first principal component (PC1), 5.11% explained by PC2. Explanation of variance reduced on further components. Ozone response of 70 *G. soja* had significant association ($p = 3.39 \times 10^{-7}$, permutation $p = 0.07$) with one marker on Chromosome 19, ss715636367. This marker returned an R^2 value of 0.01238. In addition, markers on Chromosome 3 ($p = 1.04 \times 10^{-6}$, permutation $p = 0.146$) and Chromosome 14 ($p = 1.82 \times 10^{-6}$, permutation $p = 0.199$) showed high initial significance but did not pass the criterion of permutation $p \leq 0.10$. These markers had R^2 values of 0.00662 and 0.02232, respectively. All other markers did not approach significance. Markers of significance can be seen in **Figure 4.4** and with related statistics in **Table 4.1**.

The putative ozone marker ss715636367 corresponds to an intergenic SNP located on Chromosome 19 with alleles T/C. T is listed as the minor allele with frequencies ranging from 0.231 in cultivated soybean landraces to 0.492 in wild soybean accessions. The SNP is located at base 8227790 with closest genes Glyma.19g050700, an uncharacterized ribosomal protein, and Glyma.19g050800, a GDSL-like Lipase/Acylhydrolase. A full list of the genes

within 1 Mbp of the marker on chromosome 19 can be seen in **Table 4.2**. The marker on Chromosome 3, ss715586544, is located in an intron at position 44542037 with alleles A/G. Minor allele G is present at 0.229 in cultivated soybean landraces and 0.143 in *G. soja* accessions. The SNP is located within Glyma.03g249200, a myosin family protein. The third marker, ss715619614, is located in the coding region of gene model Glyma.14g062200, a gene thought to code for Zuotin and related molecular chaperones. The SNP has G/T alleles T being the minor allele in *G. max* landraces and elite cultivars (0.115 – 0.205) and T representing the minor allele among *G. soja* accessions (0.300).

Development and Progression of Genetically Diverse Soybean Lines

First plant selections in Population 23 were made out of bulk plantings of F₃ materials in Clayton, NC and Rocky Mount, NC in 2014. Two hundred thirty-five selections were made in Clayton, NC and an additional 40 selections were made in Rocky Mount, NC. From these F₃ selections, 63 subsequent progeny rows had plants selected for advancement during the following growing season. These F_{3:4} selections were descendants from 6 of the original F₃ selections in Rocky Mount and 28 of the original F₃ selections from Clayton made in 2014. 166 single plant selections were made from amongst these 63 F_{3:4} progeny rows. In late spring 2016, 166 F_{4:5} progeny rows were planted from these selections.

The number of progeny selected from each of the sieve sizes of original F₃ plantings shows that 67 of the 166 F_{4:5} progeny are descendants of plants belonging to largest sieve size class (> 12) and another 42 came from the next largest class (< 12). 23 came from the 11 1/2 sieve, 6 from the 11 1/12 sieve, 20 from the < 11 sieve, and 8 came from the smallest sieve size class, 8 3/4, (**Figure 4.5**).

Though the population is still in development, weights (g) of 100 seeds were collected at F_{3:4} as well as the F_{4:5} generation stage. Using this data for the current 166 F_{4:5} progeny, a correlation of 0.46108 was determined between parent and offspring seed size. Subsequently, an ANOVA was run to test for association between the different seed sieve size classes of the original F₃ ancestor (< 8 3/4, < 11, < 11 1/12, < 11 1/2, < 12, > 12) and the 100-seed weight of the current progeny. Significance in this test may provide an early idea about the heritability of seed size in this cross. The distributions of 100-seed weight are presented in boxplots (**Figure 4.6**) to assess variance within each group as well as to identify influential data points. The following ANOVA identified sieve size as a significant influence on seed weight (p = 0.0282); however, the differences in LSMEANS estimates identifies that weights of those with ancestry from sieve sizes > 12 and < 11 were statistically similar and have the larger average seed weights at 5.73 g and 5.83 g, respectively. In contrast, those with ancestors belonging to the < 11 1/2 and the < 8 3/4 have the smallest seed weights at 4.76 g and 4.48 g, respectively. The other groups were not statistically different from one another.

DISCUSSION

Genome Wide Association Study for Ozone Resistance

Putative ozone markers gained from the study were used to make crosses between a known resistant *Glycine soja* parent (PI 424123) and known sensitive *Glycine max* parent (Haberlandt) with contrasting alleles at the marker of interest on Chromosome 19. These crosses will be advanced and analyzed to determine the association between the alleles and ozone resistance. These crosses may provide a tractable method for breeding for improved

ozone damage response. A mapping study to identify the heritability of this marker will help validate the association between the marker and ozone response in progeny.

Genes within 1 Mbp of the significant marker on Chromosome 19,ss715636367, show associations to possible defense responses such as peroxidases (Glyma.19G051400). However, gene expression response shown from the RNA-Seq study of ozone response in wild soybean shows down-regulation of the nearest gene to the putative marker (Glyma.19G050700) in multiple leaf positions in both sensitive accessions as well as resistant PI 424123 (Chapter III). Unfortunately, this gene does not have a known annotation. No other genes showed any pattern of expression linked to the marker.

While significant results were obtained from this study, it should be noted that the experiment was not initially designed for an association study, and additional data would help prove the validity of the results. For example, the phenotypic input taken from only 70 wild soybean accessions is relatively small and could be strengthened by more phenotypic ratings to assess ozone response. In addition, selecting more wild soybean accessions to rate was not done as the ones used in the study were chosen as a 10% representation of maximum genetic diversity present in *Glycine soja* maintained in the USDA germplasm collection. Perhaps utilizing a 20% or more representation would give better resolution to the study.

Several methods were attempted to run the statistical model for association, it should be noted that a general linear model containing a PCA to account for population structure and one only containing a kinship matrix (K) to account for family relatedness returned similar results; however, when the two were used together in a mixed linear model (MLM), significance was not returned for any marker. A plausible explanation for this could be due to

the small sample size and overcorrection of relatedness. However, both population structure as well as kinship are known to influence association studies, so a way to incorporate the two measures would be useful in future work.

Though this study does not provide a definitive link between a SNP marker and ozone response in wild soybean, it does serve as a starting point for marker-trait association studies between wild soybean and ozone response. Combination of PCA, removal of markers in strong linkage disequilibrium and the use of permutation testing returns a significant marker ($p = 0.07$). Interestingly, it is a marker that was not identified in the ozone QTL study done in cultivated soybean by Burton et al. (2016). This could indicate a unique allele contributed by wild genetic sources to improve soybean quality in future crop improvement efforts. Both cultivars used in creation of Burton's mapping study contained the same allele at this locus, while resistant *G. soja* lines had the minor allele; this gives further evidence of a novel source of resistance from wild soybean. Future work on this study could help strengthen the use of wild soybean in germplasm improvement as well as the need for greater genetic diversity in commercial plant breeding programs.

Development and Progression of Genetically Diverse Soybean Lines

Genetically diverse crosses involving wild soybean will help provide genetic resources for future soybean improvement and could contribute beneficial alleles lost in modern soybean production by breeding efforts. Following characterization of ozone response in wild soybean lines, it would have been ideal to study a population with identified ozone tolerance from wild sources in a cultivated soybean line. Unfortunately, none of the populations already in development using an ozone-resistant *G. soja* parent were ready for

the selection of “mega” populations that is our preferred method for breeding with wild soybean. Population 23 has ancestry of a parent known to be ozone sensitive; however, it was already in development and offered a chance to follow heritability of wild soybean response to ozone. While ozone response has not yet been characterized in this population, it could be used for a mapping study in the future.

Population 23 holds promise for its ozone response characterization, but it will also be used to evaluate the incorporation of wild soybean genome into agronomically-useful soybean germplasm resources. This study shows an early example of seed weights of selected progeny being connected to ancestral seed size and demonstrates evidence for seed size being a highly heritable trait. More surprisingly, it appears that agronomically-adapted or upright progeny come largely from larger seed size classes. This is in agreement with an observation made by David Eickholt on another *G. max* by *G. soja* population (personal comm., 2016).

It should be noted that the maximum 100-seed weight measurement was collected from the sieve class of > 12. While the average 100-seed weight was larger for < 11, the variance was much greater in the > 12 class. This could relate to the number of selections used in each measure with 67 in the > 12 category and only 20 in the < 11 category. This analysis should be run again once the final selections for germplasm are released following further selection and yield testing. However, the connection of sieve size to the number of selections is of note as this process is done without knowledge of the sieve size. If this relationship holds, planting only larger seed sizes may improve efficiency of selection of the “mega” populations.

CONCLUSION

A putative marker for ozone tolerance was found in *Glycine soja*, which may provide a marker to track ozone response as introduced into soybean germplasm. Similarly, phenotypic information of developed progeny regarding response to ozone may have added benefits as wild soybean is used in breeding programs to broaden genetic diversity. Populations derived from wild soybean characterized in this study are under development. These progeny will be evaluated for inheritance of the resistance to ozone has been identified. Inheritance of these traits may provide genetic resources to deal with increased abiotic stress pressure and secure the food supply.

REFERENCES

- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics (Oxford, England)*, *23*(19), 2633-2635.
- Burton, A. L., Burkey, K. O., Carter, T. E., Orf, J., & Cregan, P. B. (2016). Phenotypic variation and identification of quantitative trait loci for ozone tolerance in a Fiskeby III x Mandarin (Ottawa) soybean population. *Theoretical and Applied Genetics*, , 1-13.
- Carpenter, J., & Fehr, W. (1986). Genetic-variability for desirable agronomic traits in populations containing *Glycine soja* germplasm. *Crop Science*, *26*(4), 681-686.
- Carter Jr, T., Burton, J., Villagarcia, M., Cui, Z., Zhou, X., Fountain, M., et al. (2003). Registration of 'N7103' soybean. (Registration of cultivars). *Crop Science*, *43*(3), 1128-1129.
- Carter, T. E., Nelson, R. L., Sneller, C. H., & Lee, Z. L. (2004). Genetic diversity in soybean. In H. R. Boerma, & J. E. Specht (Eds.), *Soybeans: Improvement, production, and use* (3rd ed., pp. 303-304-416). Madison, WI, USA: ASA - CSSA - SSSA.
- Delheimer, J. C. (2012). *Agronomic performance of breeding lines developed from a single cross of glycine max X glycine soja*. Unpublished Doctoral dissertation, North Carolina State University.
- Diers, B., Keim, P., Fehr, W., & Shoemaker, R. (1992). RFLP analysis of soybean seed protein and oil content. *Theoretical and Applied Genetics*, *83*(5), 608-612.
- Egli, D. B. (2008). Comparison of corn and soybean yields in the united states: Historical trends and future prospects. *Agronomy Journal*, *100*(3), S79-S88.

- Gizlice, Z., Carter, T., & Burton, J. (1994). Genetic base for North-American public soybean cultivars released between 1947 and 1988. *Crop Science*, 34(5), 1143-1151.
- Hyten, D. L., Song, Q., Zhu, Y., Ik-Young Choi, Nelson, R. L., Costa, J. M., et al. (2006). Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(45), 16656-16671.
- Li, D., Pfeiffer, T. W., & Cornelius, P. L. (2008). Soybean QTL for yield and yield components associated with *Glycine soja* alleles. *Crop Science*, 48(2), 571-581.
- Luo, Q., Yu, B., & Liu, Y. (2005). Differential sensitivity to chloride and sodium ions in seedlings of *Glycine max* and *G. soja* under NaCl stress. *Journal of Plant Physiology*, 162(9), 1003-1012.
- Singh, R. J., & Hymowitz, T. (1999). Soybean genetic resources and crop improvement. *Genome*, 42(4), 605-616.
- Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2015). Fingerprinting soybean germplasm and its utility in genomic research. *G3: Genes|genomes|genetics*, 5(10), 1999-2006.

TABLES and FIGURES

Table 4.1. Tests of significance for ozone association in Single Nucleotide Polymorphism markers (SNP) in wild soybean diversity core accessions following GLM procedure in TASSEL 5 sorted by permutation p-value. Top 20 markers are shown in this table.

Marker	Chr*	Position	F-Statistic	p-value	Permutation p-value	R ² Value
ss715636367	19	8227790	18.97247	3.39E-07	0.07	0.01238
ss715586544	3	44542037	17.2152	1.04E-06	0.146	0.00662
ss715619614	14	5072807	16.42096	1.82E-06	0.199	0.02232
ss715604295	9	42719437	14.71672	5.52E-06	0.37	0.00168
ss715580932	2	10649706	14.02941	9.42E-06	0.481	0.01331
ss715605211	9	5948655	13.51916	1.27E-05	0.569	0.00122
ss715638007	20	39163661	13.42408	1.35E-05	0.59	-
ss715636443	19	8558219	13.32897	1.49E-05	0.613	-
ss715597622	7	37579786	13.31791	1.50E-05	0.614	-
ss715612691	12	36584840	12.94545	1.90E-05	0.669	-
ss715631550	18	50011276	12.971	2.14E-05	0.691	-
ss715586005	3	38960226	12.7537	2.29E-05	0.703	-
ss715630873	18	44188691	12.70701	2.31E-05	0.706	-
ss715637005	20	24615630	12.67987	2.70E-05	0.753	-
ss715594303	6	38104910	12.4489	2.78E-05	0.758	-
ss715628149	17	7184248	12.43874	2.80E-05	0.758	-
ss715633021	19	1063261	13.7729	3.17E-05	0.784	-
ss715607176	10	42295191	12.22048	3.20E-05	0.787	-
ss715631159	18	5133462	12.14496	3.38E-05	0.795	-
ss715628086	17	6217253	12.15733	3.43E-05	0.797	-

* Chromosome Number

Table 4.2. List of genes in 1 Mbp region around putative ozone SNP marker ss715636367 at position 8227790 on Chromosome 19. Gene models without annotations were excluded; protein length (AA) refers to the number of 3 base pair amino acids in each gene model.

Gene Name	Protein Length (AA)	Gene Description
Glyma.19G048900	83	F-box domain
Glyma.19G049000	386	F-box domain
Glyma.19G049100	328	EamA-like transporter family
Glyma.19G049200	192	Ribosomal protein S7e
Glyma.19G049300	336	Phosphomannose isomerase type I
Glyma.19G049400	367	"Response regulator receiver domain"
Glyma.19G049500	455	Enhancer of polycomb-like
Glyma.19G049700	540	"Phosphoglycerate kinase"
Glyma.19G050000	371	GDSL-like Lipase/Acylhydrolase
Glyma.19G050100	415	Nucleotidyl transferase
Glyma.19G050200	316	Retrotransposon gag protein
Glyma.19G050300	552	Sugar (and other) transporter
Glyma.19G050500	284	Fatty acid hydroxylase superfamily
Glyma.19G050800	372	GDSL-like Lipase/Acylhydrolase
Glyma.19G050900	371	GDSL-like Lipase/Acylhydrolase
Glyma.19G051000	371	GDSL-like Lipase/Acylhydrolase
Glyma.19G051100	371	GDSL-like Lipase/Acylhydrolase
Glyma.19G051200	458	"Phosphoribosyl transferase domain"
Glyma.19G051300	617	PIF1-like helicase
Glyma.19G051400	91	Peroxidase
Glyma.19G051500	497	Copine
Glyma.19G051800	124	Response regulator receiver domain
Glyma.19G051900	109	Ribosomal Proteins L2, C-terminal domain
Glyma.19G052000	126	Myb-like DNA-binding domain
Glyma.19G052100	451	PPR repeat
Glyma.19G052300	212	Ras family
Glyma.19G052400	448	"Elongation factor Tu GTP binding domain"
Glyma.19G053100	120	Mitochondrial domain of unknown function (DUF1713)
Glyma.19G053200	360	GDSL-like Lipase/Acylhydrolase
Glyma.19G053300	368	alpha/beta hydrolase fold
Glyma.19G053400	109	Ribosomal protein L16p/L10e

Table 4.3. ANOVA on significance of sieve size on Average 100 Seed Weight for F_{4:5} *G. max* x *G. soja* progeny lines in Population 23.

Source	DF*	SS†	Mean Square	F Value	p-value
Sieve Size	5	27.6935395	5.5387079	2.58	0.0282
Error	160	343.2325448	2.1452034		
Corrected Total	165	370.9260843			
R ² Value	CV‡	Root MSE††	Average 100 Seed Weight		
0.074661	27.04172	1.464651	5.416265		

*Degrees of Freedom

†Sum of Squares

‡ Coefficient of Variation

†† Mean Square Error



Figure 4.1. Selections are made on individual F_3 plants from a *G. max* X *G. soja* cross. On the left is an example of a viney, unfavorable plant while the plant on the right displays a cultivated soybean appearance and would be selected for advancement.

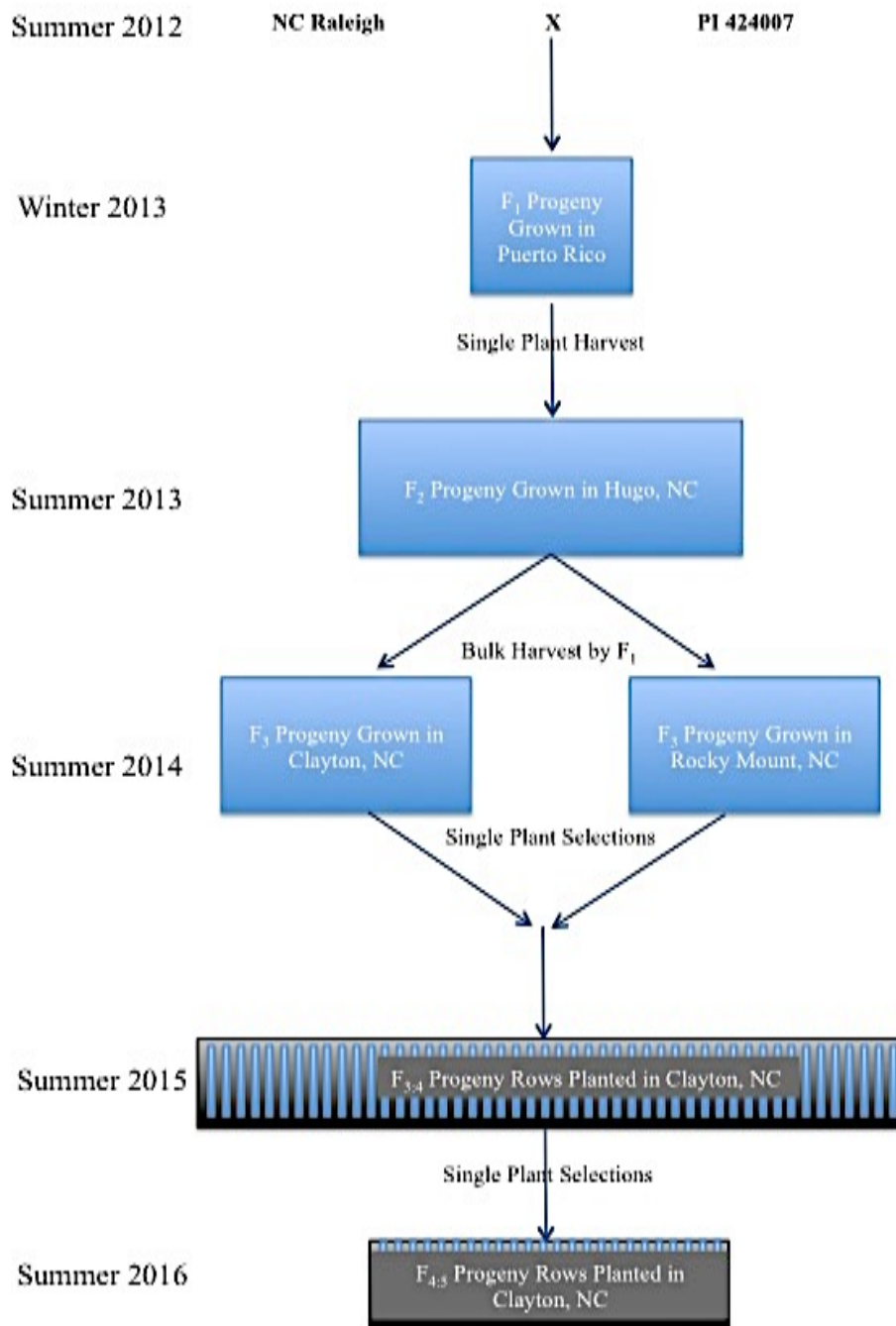


Figure 4.2. Breeding scheme for Population 23, a cross developed to increase genetic diversity in soybean now connected to response to elevated ozone.

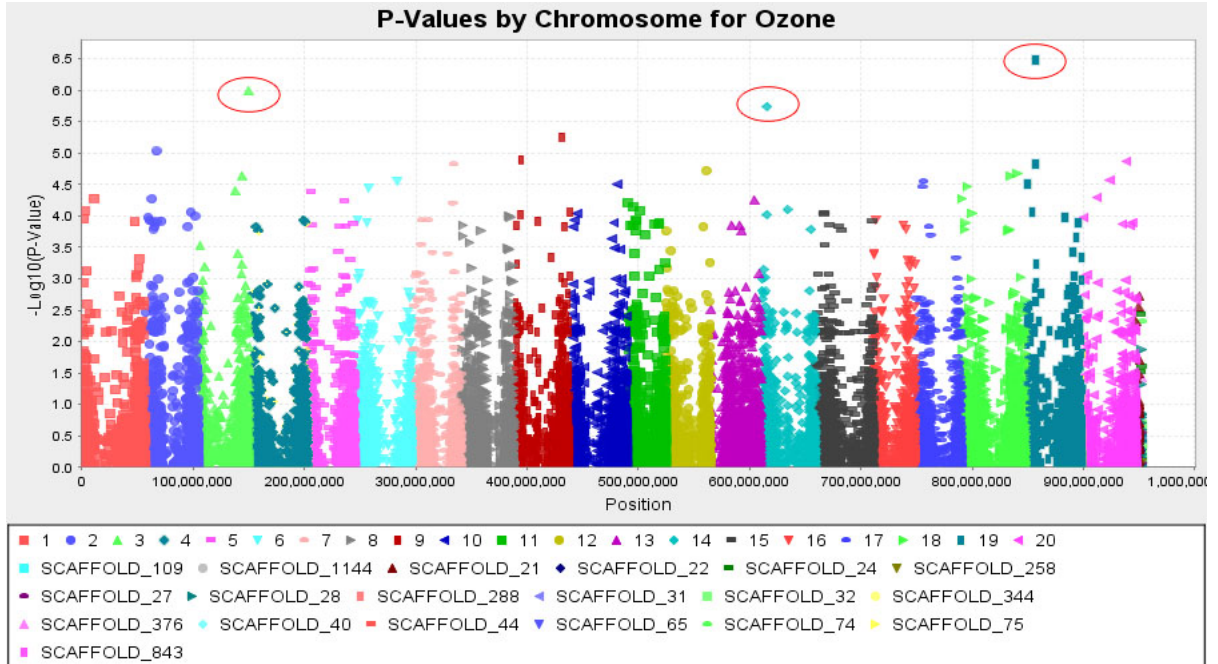


Figure 4.3. Manhattan plot of markers associated with ozone response in *Glycine soja*. Markers circled in red are those that were potentially most significant, while the marker on Chromosome 19 proved significant after permutation testing.

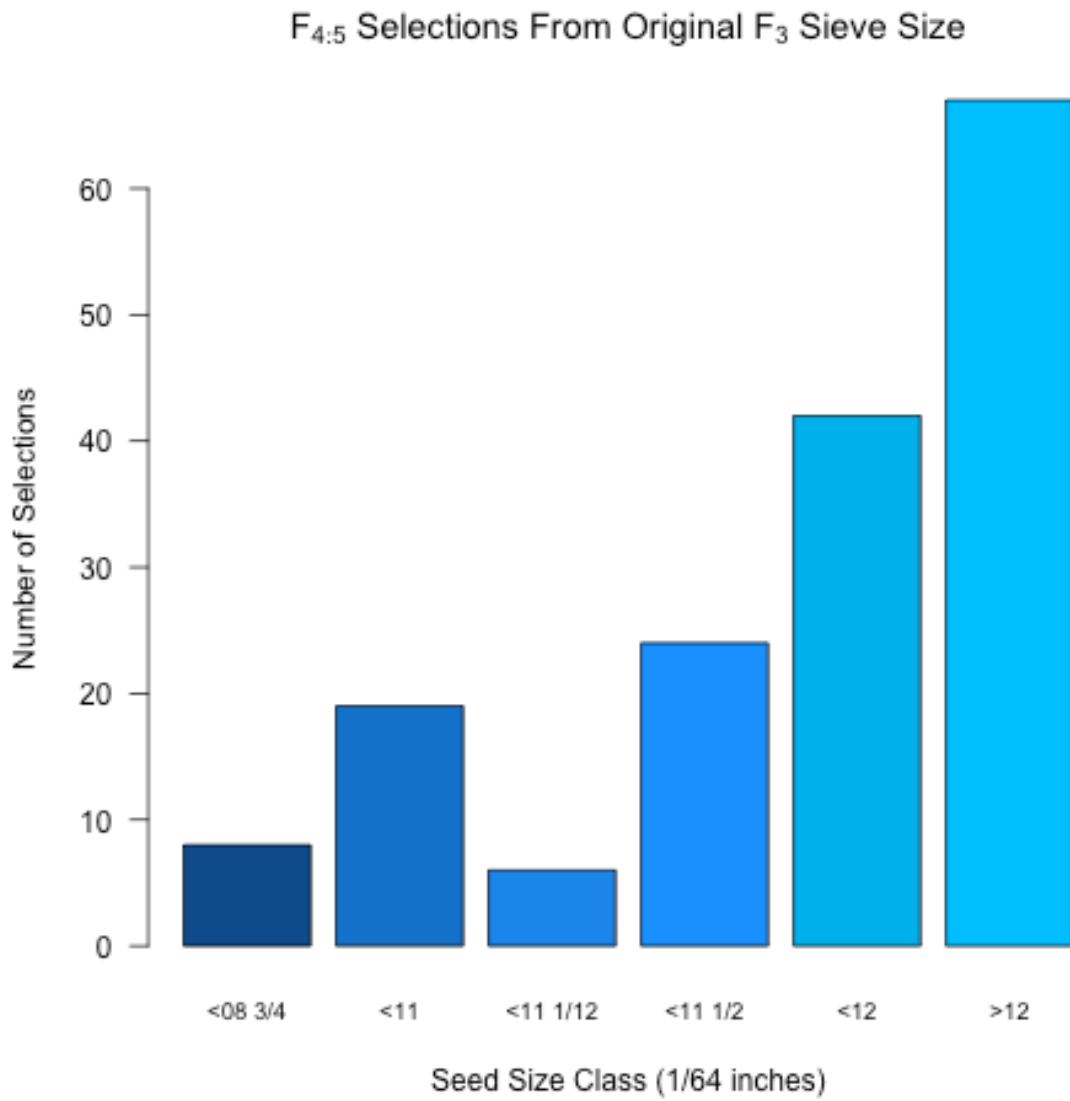


Figure 4.4. Bar chart of the number of F_{4:5} selections from original seed sieve size of F₃ plants for Population 23

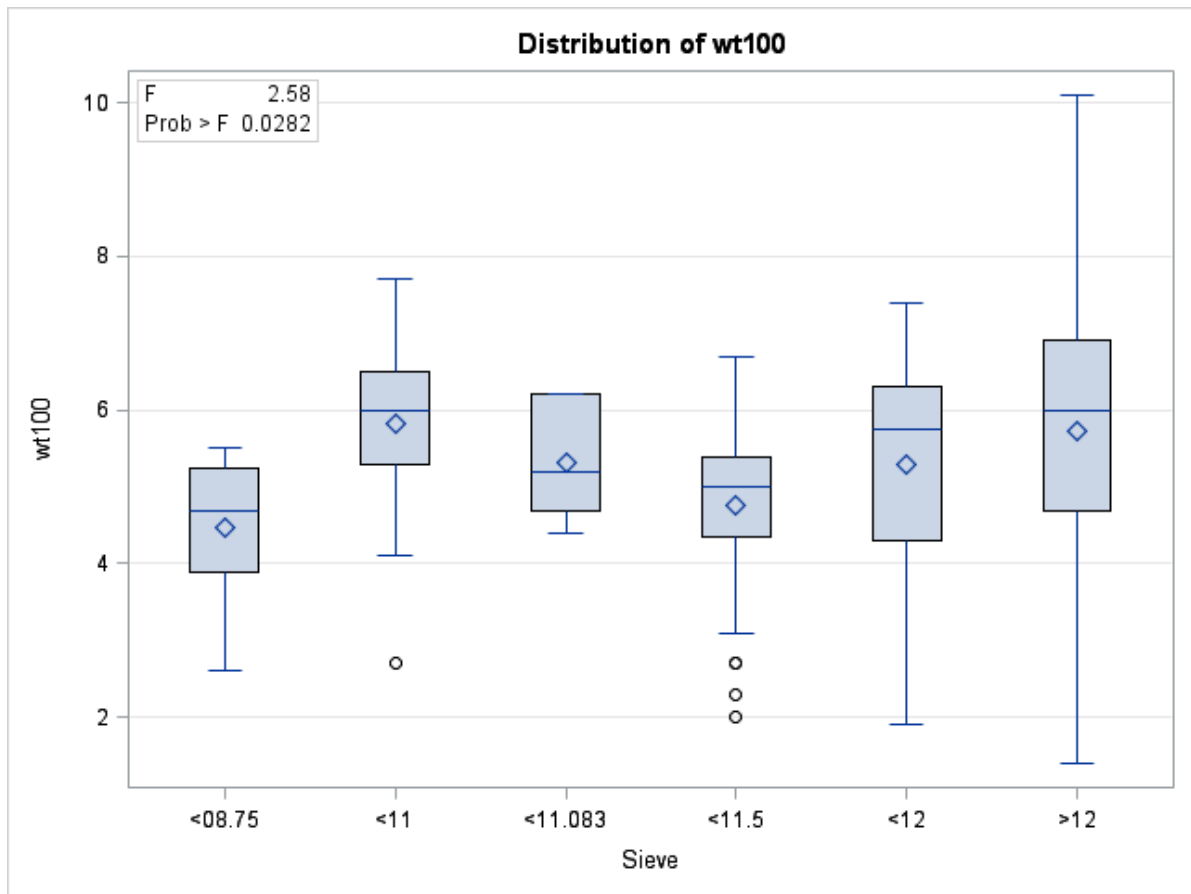


Figure 4.5. Box plots of 100 seed weights (wt100) of $F_{4:5}$ Progeny grouped by sieve class of each F_3 ancestor. The legend displays the associated F-Value and Prob > F (p-value) resulting from the ANOVA for association between 100 seed weight and sieve size.

APPENDIX

Appendix C. LSMEAN estimates of percent germination and average root length for each genotype tested at each of the temperatures used in the cold germination experiment. Summary statistics for each temperature are listed at the bottom of the table.

Genotype	MG	Percent Germination				Mean Root Length			
		11 C	17 C	22 C	28 C	11 C	17 C	22 C	28 C
366122	IV	0.67	0.81	0.80	0.88	0.50	1.65	7.22	9.88
378683	VI	0.50	0.62	0.64	0.61	0.08	1.55	3.46	6.08
378690	VII	0.67	0.80	0.78	0.71	0.19	1.31	2.35	3.42
407020	V	0.92	0.96	0.89	0.88	0.35	1.73	2.67	5.43
407029	IV	0.72	0.94	1.01	0.95	0.45	1.89	4.98	7.96
407038	V	0.75	0.90	0.95	0.87	0.18	1.77	4.47	7.01
407042	V	0.90	0.92	0.94	0.86	-0.20	1.62	2.37	5.94
407052	V	0.39	0.88	0.87	0.82	0.07	1.49	3.82	5.79
407059	VII	0.78	0.92	0.75	0.87	0.29	1.44	4.82	9.44
407060	VI	0.44	0.81	0.73	0.76	0.10	2.00	3.84	8.32
407085	VI	0.79	0.83	0.89	0.67	-0.24	1.96	5.33	7.66
407096	VII	0.71	0.87	0.97	0.93	0.10	2.10	5.05	9.58
407156	VI	0.27	0.70	0.85	0.75	-0.09	1.20	4.69	6.07
407157	VI	0.57	0.97	0.77	1.07	0.17	2.28	3.45	6.27
407171	IV	0.67	0.85	0.87	0.86	0.31	2.05	5.71	7.97
407179	V	0.82	0.94	0.97	0.92	0.19	2.25	4.70	6.76
407191	V	0.84	0.89	0.92	0.73	0.69	1.94	7.08	7.06
407195	IV	0.85	0.76	0.90	0.77	0.33	2.06	5.89	7.16
407206	V	0.63	0.85	0.91	0.88	0.13	1.62	5.91	8.75
407214	V	0.76	0.92	0.91	0.89	0.30	2.04	5.01	7.90
407228	V	0.64	0.96	0.98	1.01	0.29	2.40	4.72	7.33
407231	V	0.62	0.97	0.91	0.92	0.29	1.54	3.95	6.86
407240	V	0.66	0.88	0.85	0.90	-0.45	1.06	3.35	7.60
407248	V	0.70	0.90	0.88	0.85	0.29	1.88	5.48	8.62
407287	V	0.77	0.93	0.90	1.01	0.71	2.14	7.22	8.79
407300	V	0.62	0.91	0.92	0.86	0.22	2.24	4.16	8.04
407314	V	0.75	0.98	0.97	0.93	0.23	2.40	5.02	8.43
424007	V	0.62	0.79	0.79	0.83	0.13	1.61	4.89	7.59
424035	V	0.67	1.02	0.98	0.94	0.35	2.45	4.69	8.15
424045	V	0.83	0.89	0.98	0.97	0.26	2.75	5.22	8.21
424052	V	0.41	0.83	0.79	0.83	0.12	1.58	4.56	7.82
424082	V	0.70	0.84	0.88	0.89	0.29	1.53	2.85	7.50
424116	IV	0.79	0.96	0.97	0.97	0.30	2.34	6.89	10.26
424123	V	0.67	0.90	0.95	0.92	0.29	2.32	3.66	7.12
468916	III	0.50	0.88	0.92	0.93	0.17	1.29	5.01	8.01

Appendix A Continued.

Genotype	MG	Percent Germination				Mean Root Length			
		11 C	17 C	22 C	28 C	11 C	17 C	22 C	28 C
479751	III	0.27	0.43	0.42	0.48	0.12	0.85	3.39	5.28
479752	E	0.85	0.99	1.11	1.05	0.45	1.98	4.96	7.90
483466	V	0.77	0.96	0.94	0.98	0.14	1.80	4.28	6.33
483467	IV	0.75	0.85	0.81	0.81	0.34	2.36	5.59	10.29
507618	V	0.56	0.95	0.96	0.88	0.10	1.98	4.65	6.95
507624	VII	0.53	0.96	0.97	0.85	-0.07	1.95	4.87	8.52
507630	VII	0.48	0.88	1.02	0.85	0.73	2.04	4.55	5.79
507641	V	0.65	0.88	0.94	0.98	0.15	1.86	5.46	7.44
507656	VII	0.23	0.88	0.97	0.84	0.00	1.67	5.06	8.82
549032	III	0.41	0.63	0.68	0.60	0.27	1.29	3.00	5.51
549046	III	0.82	0.89	0.87	0.82	0.35	2.98	7.65	7.01
549048	III	0.92	0.89	0.84	0.94	0.19	2.33	7.00	11.85
562547	V	0.76	0.90	0.84	0.83	0.08	1.76	5.24	6.06
562551	V	0.65	0.96	0.93	0.94	0.09	2.68	6.53	9.67
562553	V	0.74	0.90	0.93	0.92	0.25	2.02	4.82	6.68
562561	V	0.83	0.95	0.94	0.95	0.23	2.06	5.21	9.30
562564	IV	0.54	0.58	0.48	0.72	0.31	0.85	5.14	1.47
562565	IV	0.41	0.88	0.94	0.89	0.22	2.09	6.98	9.06
593983	III	0.55	0.53	0.60	0.59	0.31	0.90	4.29	7.76
339871A	V	0.70	0.90	0.96	0.94	0.54	1.81	5.28	8.89
378684B	VI	0.69	0.95	0.97	0.88	0.13	2.15	4.60	7.80
378686B	VI	0.81	0.86	0.92	0.91	-0.22	1.65	6.99	8.34
378696B	VI	0.56	1.00	0.90	0.84	0.20	1.73	4.13	8.77
378697A	V	0.64	0.89	0.95	0.97	0.40	1.58	4.02	6.32
424025B	V	0.79	0.92	0.93	0.85	0.23	2.56	6.41	8.80
424070B	V	0.81	0.92	0.96	0.85	0.20	1.72	4.63	5.92
424083A	V	0.71	0.84	0.88	0.81	0.16	2.06	4.81	7.42
424102A	V	0.69	0.90	0.94	0.83	0.06	2.13	3.59	7.40
522211A	III	0.95	1.11	1.15	1.09	0.49	2.37	5.50	8.03
532453A	III	0.26	0.65	0.76	0.76	0.06	2.18	4.72	10.87
597458C	V	0.83	0.91	0.91	0.93	0.16	1.77	5.60	8.66
597460A	IV	0.61	0.69	0.65	0.68	0.23	1.67	3.19	8.23
597461B	V	0.81	0.86	0.84	0.88	0.27	1.79	5.00	7.93
597462B	IV	0.53	0.58	0.56	0.65	0.35	1.67	5.59	6.64
	Mean	0.66	0.86	0.88	0.86	0.22	1.88	4.89	7.63
	Variance	0.07	0.03	0.04	0.03	0.02	0.53	4.26	6.33
	DF	66	66	66	66	66	66	66	66

Appendix B. LSMEAN estimates of percent germination and average root length for each genotype tested at each of the temperatures used in the hot germination experiment. Summary statistics for each temperature are listed at the bottom of the table.

Genotype	MG	Percent Germination				Mean Root Length			
		28 C	33 C	38 C	43 C	28 C	33 C	38 C	43 C
366122	IV	0.55	0.41	0.59	0.08	3.79	4.26	6.41	0.55
378683	VI	0.66	0.50	0.46	0.14	2.83	3.42	2.14	1.66
378690	VII	0.64	0.49	0.24	0.05	4.87	3.40	0.33	0.58
407020	V	0.45	0.64	0.50	0.00	2.13	2.89	2.19	-
407029	IV	0.39	0.28	0.16	0.26	4.66	2.32	4.17	1.43
407038	V	0.86	0.62	0.31	0.07	3.05	4.38	1.57	-0.06
407042	V	0.89	0.83	0.57	0.41	3.76	5.11	2.54	2.98
407052	V	0.72	0.51	0.51	0.11	3.12	3.64	2.50	0.69
407059	VII	0.82	0.53	0.45	-0.01	3.07	3.21	1.05	-0.05
407060	VI	0.71	0.71	0.43	0.24	4.98	6.93	2.10	1.86
407085	VI	0.89	0.91	0.73	-0.01	3.52	6.64	2.28	0.70
407096	VII	0.86	0.87	0.77	0.28	7.11	7.54	2.53	4.06
407156	VI	0.73	0.68	0.62	0.07	3.43	3.91	1.95	0.22
407157	VI	0.75	0.61	0.60	0.40	4.88	4.13	2.89	3.48
407171	IV	0.86	0.91	0.74	0.41	4.98	8.57	4.05	0.84
407179	V	0.90	0.64	0.75	0.35	4.69	6.35	3.51	1.20
407191	V	0.89	0.70	0.48	0.24	6.37	5.43	6.03	1.10
407195	IV	0.86	0.72	0.66	0.29	5.42	5.38	2.29	4.40
407206	V	0.82	0.69	0.53	0.29	5.02	7.14	3.84	2.16
407214	V	0.72	0.60	0.55	0.17	4.70	4.94	4.01	2.11
407228	V	0.69	0.73	0.72	0.27	4.35	7.32	5.38	3.90
407231	V	0.66	0.62	0.58	0.13	4.16	5.38	6.14	0.46
407240	V	0.87	0.70	0.47	0.36	4.24	6.67	3.68	1.51
407248	V	0.68	0.54	0.50	0.17	3.60	6.01	2.63	1.74
407287	V	0.71	0.56	0.48	0.20	4.07	4.36	2.62	6.12
407300	V	0.78	0.66	0.58	0.15	5.53	5.63	3.78	1.28
407314	V	0.92	0.85	0.63	0.28	4.91	7.47	3.41	2.47
424007	V	0.71	0.48	0.44	0.29	2.11	3.17	2.24	2.61
424035	V	0.70	0.77	0.68	0.31	3.64	6.90	5.44	-0.46
424045	V	0.87	0.76	0.67	0.40	3.75	4.17	4.59	0.42
424052	V	0.88	0.92	0.90	0.81	5.74	8.48	5.40	2.88
424082	V	0.71	0.60	0.76	0.38	3.53	4.89	3.99	0.29
424116	IV	0.88	0.71	0.65	0.33	6.58	10.25	3.39	2.57
424123	V	0.74	0.54	0.64	0.26	3.24	3.83	4.43	2.55
468916	III	0.91	0.88	0.87	0.45	4.64	7.17	4.15	2.91
479752	E	0.94	0.88	0.86	0.45	3.16	5.63	3.80	2.21

Appendix B Continued.

Genotype	MG	Percent Germination				Mean Root Length			
		28 C	33 C	38 C	43 C	28 C	33 C	38 C	43 C
483466	V	0.92	0.90	0.71	0.26	4.32	6.17	3.82	1.71
483467	IV	0.86	0.74	0.41	0.20	5.30	8.94	2.23	0.62
507618	V	0.82	0.76	0.73	0.40	3.41	4.37	3.15	1.69
507624	VII	0.80	0.87	0.49	0.37	2.04	7.56	1.23	3.28
507630	VII	0.57	0.49	0.54	0.15	3.05	2.20	2.08	0.62
507641	V	0.82	0.78	0.61	0.12	4.04	5.80	3.96	0.91
507656	VII	0.87	0.27	0.18	0.11	3.62	0.01	-0.20	0.00
549032	III	0.81	0.69	0.63	0.43	5.41	7.35	3.86	2.46
549046	III	0.66	0.67	0.51	0.07	4.01	7.21	3.50	0.51
549048	III	0.88	0.83	0.95	0.28	7.88	11.12	4.14	3.22
562547	V	0.78	0.65	0.70	0.62	3.94	5.33	4.62	1.13
562551	V	0.90	0.70	0.59	0.22	4.44	5.06	4.80	1.40
562553	V	0.81	0.60	0.58	0.25	3.80	5.55	3.10	1.87
562561	V	0.99	0.93	0.95	0.51	5.73	7.37	6.45	1.71
562564	IV	0.57	0.51	0.75	0.11	3.60	0.55	0.56	-
562565	IV	0.55	0.62	0.45	0.17	4.09	6.69	2.98	2.32
593983	III	0.57	0.57	0.52	0.28	5.39	8.38	3.22	1.66
339871A	V	0.93	0.76	0.64	0.26	4.18	7.77	4.86	4.65
378684B	VI	0.92	0.69	0.72	0.27	5.16	5.54	2.50	3.72
378686B	VI	0.81	0.46	0.34	0.04	4.10	4.80	0.24	0.38
378696B	VI	0.77	0.58	0.53	0.18	4.60	5.03	2.67	3.80
378697A	V	0.99	0.83	0.69	0.31	4.21	6.42	3.61	2.78
424025B	V	0.75	0.87	0.71	0.32	5.09	8.77	5.87	2.51
424070B	V	0.70	0.66	0.60	0.27	4.29	5.06	4.50	3.46
424083A	V	0.82	0.76	0.45	0.16	4.83	7.90	2.38	3.14
424102A	V	0.80	0.48	0.29	0.07	3.34	2.52	2.30	0.79
522211A	III	1.00	1.02	1.02	0.52	6.07	6.62	5.62	5.57
532453A	III	0.69	0.73	0.55	0.43	5.36	9.48	3.45	1.75
597458C	V	0.93	0.85	0.81	0.47	6.27	7.45	4.92	2.03
597460A	IV	0.67	0.45	0.33	0.17	2.90	3.39	2.17	1.95
597461B	V	0.78	0.79	0.82	0.41	5.63	8.32	4.88	0.90
597462B	IV	0.51	0.36	0.34	0.12	5.71	4.63	1.80	1.24
	Mean	0.78	0.68	0.59	0.26	4.40	5.74	3.36	1.93
	Variance	0.05	0.07	0.08	0.05	3.90	9.72	4.09	2.78
	DF	65	65	65	65	65	65	65	63

Appendix C. Correlation of means, screening Glycine soja genotypes at 8 temperatures for Percent Germination and Average Root Length.

		Percent Germination				Average Root Length				Percent Germination				Average Root Length			
		11 °C	17 °C	22 °C	28 °C	11 °C	17 °C	22 °C	28 °C	28 °C	33 °C	38 °C	43 °C	28 °C	33 °C	38 °C	43 °C
Percent Germination	11 °C	1.00															
	17 °C	0.46	1.00														
	22 °C	0.38	0.82	1.00													
	28 °C	0.40	0.82	0.73	1.00												
Average Root Length	11 °C	0.20	0.11	0.13	0.18	1.00											
	17 °C	0.36	0.48	0.44	0.42	0.26	1.00										
	22 °C	0.29	0.17	0.21	0.23	0.30	0.45	1.00									
	28 °C	0.11	0.15	0.09	0.26	0.06	0.35	0.56	1.00								
Percent Germination	28 °C	0.35	0.40	0.41	0.36	-0.21	0.15	0.04	0.16	1.00							
	33 °C	0.32	0.33	0.35	0.32	-0.10	0.26	0.10	0.17	0.68	1.00						
	38 °C	0.31	0.30	0.31	0.41	0.09	0.21	0.12	0.16	0.53	0.78	1.00					
	43 °C	0.08	0.16	0.14	0.27	0.01	0.05	-0.09	0.07	0.37	0.56	0.64	1.00				
Average Root Length	28 °C	0.24	-0.01	-0.05	0.07	0.16	0.18	0.29	0.34	0.24	0.36	0.38	0.32	1.00			
	33 °C	0.21	0.01	-0.02	0.08	-0.02	0.23	0.30	0.45	0.32	0.70	0.54	0.43	0.61	1.00		
	38 °C	0.30	0.30	0.25	0.36	0.35	0.25	0.22	0.17	0.12	0.43	0.61	0.50	0.34	0.39	1.00	
	43 °C	0.17	0.20	0.19	0.29	0.18	0.15	0.08	0.14	0.23	0.33	0.33	0.35	0.30	0.28	0.13	1.00

Appendix D. Geographical Information for USDA *Glycine soja* Diversity Core Collection used in characterization of abiotic stress response.

Accession	MG	Province of Collection	Country of Collection
PI101404A	II	Heilongjiang	China
PI245331	X	Unknown	Taiwan
PI339871A	V	Cheju	South Korea
PI342622A	I	Primorye	Russia
PI366122	IV	Fukushima	Japan
PI378683	VI	Ishikawa	Japan
PI378684B	VI	Saitama	Japan
PI378686B	VI	Kochi	Japan
PI378690	VII	Fukuoka	Japan
PI378696B	VI	Shimane	Japan
PI378697A	V	Aomori	Japan
PI407020	V	Akita	Japan
PI407038	V	Akita	Japan
PI407042	V	Akita	Japan
PI407052	V	Iwate	Japan
PI407059	VII	Aichi	Japan
PI407085	VI	Aichi	Japan
PI407096	VII	Hyogo	Japan
PI407156	VI	Yamanashi	Japan
PI407157	VI	Chiba	Japan
PI407171	IV	Kyonggi	South Korea
PI407179	V	Kyonggi	South Korea
PI407191	V	Kyonggi	South Korea
PI407195	IV	Kangwon	South Korea
PI407206	V	Kangwon	South Korea
PI407214	V	Chungchong Puk	South Korea
PI407228	V	Chungchong Nam	South Korea
PI407231	V	Chungchong Nam	South Korea
PI407240	V	Kyongsang Puk	South Korea
PI407248	V	Kyongsang Puk	South Korea
PI407287	V	Kanagawa	Japan
PI407300	V	Jiangsu	China
PI407314	V	Chungchong Puk	South Korea
PI424004B	II	Kyonggi	South Korea
PI424007	V	Kyonggi	South Korea

Appendix D Continued.

Accession	MG	Province of Collection	Country of Collection
PI424025B	V	Kyonggi	South Korea
PI424035	V	Kyonggi	South Korea
PI424045	V	Kyonggi	South Korea
PI424070B	V	Kangwon	South Korea
PI424082	V	Kangwon	South Korea
PI424083A	V	Kangwon	South Korea
PI424102A	V	Kyongsang Puk	South Korea
PI424116	IV	Kyongsang Puk	South Korea
PI424123	V	Kyongsang Nam	South Korea
PI447003A	Z	Nei Monggol	China
PI458536	Z	Heilongjiang	China
PI464890B	I	Jilin	China
PI468916	III	Liaoning	China
PI479746B	II	Jilin	China
PI479751	III	Jilin	China
PI479752	I	Jilin	China
PI479768	Z	Heilongjiang	China
PI483466	V	Shandong	China
PI507618	V	Nagano	Japan
PI507624	VII	Gifu	Japan
PI507641	V	Hyogo	Japan
PI507656	VII	Nagasaki	Japan
PI507761	I	Unknown	Russia
PI522209B	II	Primorye	Russia
PI522226	ZZZ	Primorye	Russia
PI522233	I	Primorye	Russia
PI522235B	I	Primorye	Russia
PI549032	III	Liaoning	China
PI549046	III	Shaanxi	China
PI549048	III	Beijing	China
PI562547	V	Chungchong Nam	South Korea
PI562551	V	Chungchong Nam	South Korea
PI562553	V	Chungchong Nam	South Korea
PI562561	V	Cholla Puk	South Korea
PI562565	IV	Cholla Puk	South Korea
PI593983	III	Hokkaido	Japan

Appendix D Continued.

Accession	MG	Province of Collection	Country of Collection
PI597448D	Z	Heilongjiang	China
PI597458C	V	Shandong	China
PI597460A	IV	Shandong	China
PI597461B	V	Shandong	China
PI597462B	IV	Shandong	China
PI639586	ZZZ	Amur	Russian Federation
PI639588B	ZZZ	Amur	Russian Federation
PI639621	ZZZ	Amur	Russian Federation
PI639623A	ZZZ	Amur	Russian Federation
PI639635	ZZZ	Primorye	Russian Federation