# GENOMIC DISSECTION OF COMPLEX TRAIT VARIATION IN WILD SOYBEAN, *GLYCINE SOJA*

by

Janice Kofsky

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Biology

Charlotte

2019

Approved by:

_____
Dr. Bao-Hua Song

_____
Dr. Larry Leamy

_____
Dr. Adam Reitzel

_____
Dr. Changbao Li

_____
Dr. Robert Reid

_____
Dr. Jennifer Weller

©2019

Janice Kofsky

ALL RIGHTS RESERVED

# ABSTRACT

JANICE KOFSKY. Genomic dissection of complex trait variation in wild soybean, *Glycine soja*. (Under the direction of DR. BAO-HUA SONG)

There is a considerable demand for crop improvement, especially considering the growing population, continuing climatic fluctuations, and rapidly evolving plant pests and pathogens. Crop wild relatives hold great potential in providing beneficial alleles for crop improvement. Wild soybean, *Glycine soja* (Siebold & Zucc.), the wild ancestor to the domesticated soybean (*Glycine max* (L.) Merr.), harbors a high level of genetic variation. Research on *G. soja* has been largely devoted to understanding the domestication history of the soybean, while little effort has been made to explore its genetic diversity for crop improvement. In this dissertation, genomic dissection of agronomically important complex trait variation in *G. soja* was carried out in order to determine the potentially novel genetic architecture of traits of interest, and to better understand the use of crop wild relatives for crop improvement. The genetic background of quantitative traits: early vigor, Soybean Cyst Nematode (SCN) resistance and the effects of abiotic and biotic combination stresses (drought and SCN) on gene regulation was investigated. These findings have the potential to enhance the cultivated soybean with cutting edge biotechnologies (genetic modification, gene editing) and promote the exploration and use of wild resources for crop improvement in order to meet future food requirements.

# DEDICATION

To Science: *"It works…bitches."*

- Richard Dawkins

# ACKNOWLEDGEMENTS

This research was made possible by the outstanding support I have received from my advisor, Dr. Bao-Hua Song. Thank you for the opportunity to work in your lab and your expert guidance along the way. I also thank my committee members, Dr. Larry Leamy, Dr. Adam Reitzel, Dr. Changbao Li, Dr. Rob Reid, and Dr. Jennifer Weller for their constructive input and encouragement. In addition, I would like to acknowledge the invaluable contributions of my colleague and mentor in the lab, Dr. Hengyou Zhang, and all graduate students that support me daily.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

# CHAPTER 1: THE UNTAPPED GENETIC RESERVOIR: THE PAST, CURRENT, AND FUTURE APPLICATIONS OF THE WILD SOYBEAN (*GLYCINE SOJA*)

## 1. Soybean and wild soybean

The cultivated soybean (*Glycine max* (L.) Merr.) is an economically important crop grown world-wide with diverse uses in oil and protein consumption for human and livestock, as well as feedstock for biofuel production. The cultivated soybean produces two-thirds of the world's calories for agriculture, and is the leading producer of oil seed. With the growing world population, the yield demand for soybean is at a current deficit of 1.2% annual production per year (1). Although soybean production has increased by approximately 40% in the last quarter century (1990 - 2015) (2), current soybean yield potential is restricted by the narrow genetic reservoir in cultivated soybean, hindering the potential for breeding soybean varieties with high environmental stress tolerance and resistance traits. These drawbacks regarding the current cultivars are increasing because of the ecological changes caused by climate fluctuations and other factors such as expanded drought or saline environments. Meanwhile, the rapid evolution of pests and pathogens is posing difficult challenges to soybean production. Because of the aforementioned factors, there is an urgent need to use the untapped genetic resources from the wild relatives of cultivated soybean to improve crop production.

The wild soybean, *Glycine soja* (Siebold & Zucc.), is the wild ancestor of the domesticated soybean, *G. max*. *G. soja* is native to East Asia with a broad geographic range, from East Russia to South China, and the species can grow in diverse habitats. Though *G. soja* and *G. max* are different in many phenotypic characteristics (Figure 1), they have the same number of chromosomes (2n = 40), exhibit normal meiotic chromosome pairing, and are cross-compatible to produce vigor hybrids (3). Also, *G. soja* was found to harbor more than half of the rare alleles, which might have been lost in *G. max* during their domestication and improvement (4).

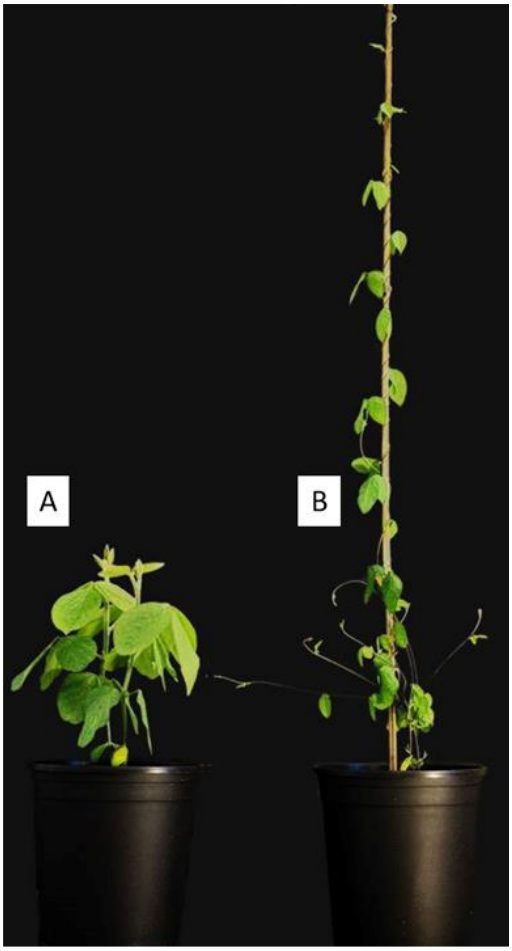



**Figure 1.** Phenotypic comparison of *Glycine max* (A) and *Glycine soja* (B).

Considering the higher level of genetic diversity retained in *G. soja*, as well as its adaptations to harsh environments, it is expected that *G. soja* has great potential to improve its agriculturally important domesticated relative (Figure 2), beyond what is currently known (5-8). Research on *G. soja* has been largely devoted to understanding the domestication history of the soybean, with comparatively little effort made to use it as a genetic reservoir for soybean improvement (5, 9-11). In

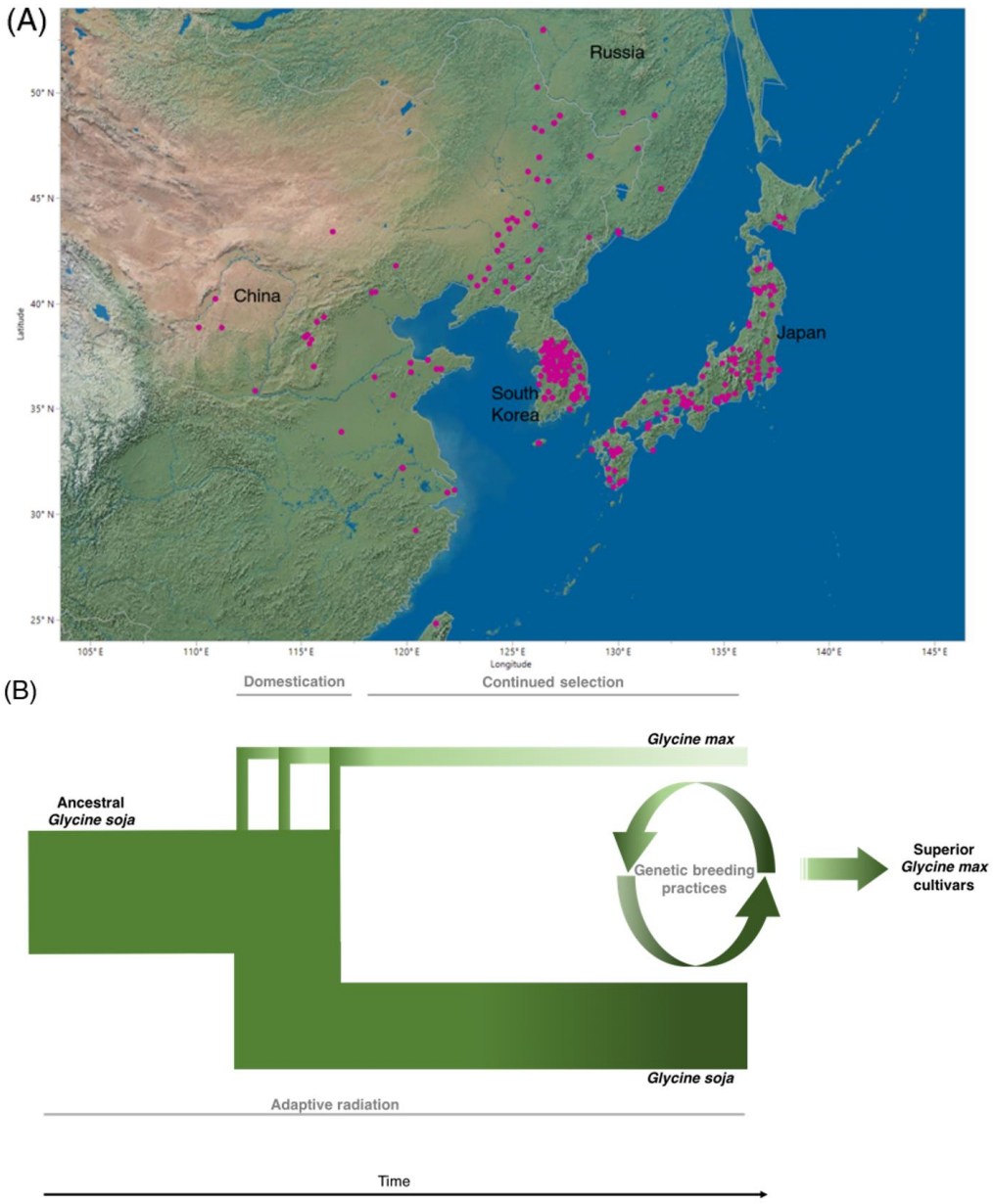



**Figure 2.** *Glycine max* diverged from ancestral *Glycine soja* as a result of multiple domestication events. *Glycine max* then underwent continued artificial selection for traits of agronomic importance, further reducing the genetic diversity found in *Glycine max*. During this time, *Glycine soja* continued to adapt to its various environments, maintaining and potentially increasing genetic variability. Genetic breeding practices incorporate select components from the *Glycine soja* gene pool to improve modern cultivars, creating superior *Glycine max* cultivars. Width and color represent genetic variability.

this review, we will discuss genomic diversity and the current research in *G. soja*, and we present a potential step forward in our application of *G. soja* to improve *G. max*.

## 2. Domestication history

Domestication of *G. soja* is reported to have occurred around 6,000-9,000 years ago in regions along the Yellow River or Huang-Huai Valley in Central China (3, 12, 13), resulting in landraces of *G. max*, and with further selection, the elite cultivars. However, the history of *G. soja* in relation to *G. max* is far more complex, with varying contradictory hypotheses. It is likely that the domestication process happened over a long period of time, allowing for frequent introgressions between the wild and cultivated populations during this time (14, 15). Although an opposing study of candidate domestication regions in Korean wild soybeans suggests a single selective sweep, detecting no evidence for multiple domestication events in East Asia (16). A summation of the contradictory hypotheses of soybean origin and domestication was recently presented in a review of the domestication history (17), in which three overarching hypotheses were provided: the single origin hypothesis, the multiple origin hypothesis, and the complex hypothesis. The single origin hypothesis states that *G. max* was diverged from *G. soja* from a single domestication event in Central China, no earlier than 9,000 years ago, supported by the observation that all the selected domesticated soybeans were clustered together by analyzing whole-genome SNPs of 302 wild, landrace and cultivated soybeans (18). The multiple origin hypothesis states that *G. max* was domesticated from *G. soja* during multiple events between 5,000 and 9,000 years ago. The complex hypothesis incorporates the results from two recent studies (19,

20), with a *G. soja*/*G. max* complex first diverging before multiple domestication events. The estimated age of the *G. soja*/*G. max* complex is 0.27 MYA (19) by whole genome comparison of one ecotype to one cultivar, or 0.8 MYA (20) by pan-genome comparison of 7 wild ecotypes. In this last hypothesis, the domestication would have stemmed from an already diverged *G. soja/G. max* complex (17). Individuals from either *soja* or *max* subpopulation were closely clustered based on their geographic origins (11, 21). A possible explanation could be that the early-domesticated *G. soja* or *G. soja/G. max* complex spread from China to Korea and Japan, and subsequently underwent varying degrees of domestication to meet local needs. Nevertheless, it is widely accepted that *G. max* was created from *G. soja* or *G. soja/G. max* complex through a long, slow, and complex domestication process by countless independent efforts (17).

**3. Advantages of using wild soybean**

*G. soja* gene pool is indisputably more diverse than *G. max* due to artificial selection during domestication and continued loss due to modern breeding practices. A comparison of 102 gene sequences from 26 *Glycine soja*, 52 landraces, 17 North American ancestor, and 25 elite cultivar isolates suggested that the most significant loss in genetic diversity occurred during the domestication bottleneck, and a secondary loss of diversity during modern breeding. The domestication bottleneck has resulted in an 81% loss of rare alleles, 60% gene allele frequency change, and almost halving the nucleotide diversity ($\pi$) from *G. soja* ($\pi = 2.47 \times 10^{-3}$) to landrace ($\pi = 1.47 \times 10^{-3}$). After domestication, intense selection toward the elite cultivars ($\pi = 1.17 \times 10^{-3}$) resulted in an additional loss of 23% nucleotide diversity

($\pi$), and a 21% loss of rare alleles (4).These artificial selection processes resulted in morphological differences of many agriculturally-important traits between *G. max* and *G. soja*, such as pod shattering resistance (22), determinate growth habit (23), and seed-related traits (24). Directional selection during modern soybean breeding practice also reduced the genetic diversity surrounding the regions conferring these agriculturally important traits, known as selective sweep. A recent genome-wide sequencing analysis showed that half of the resistance-related genes/loci in *G. soja* were not found in landraces or the domesticated soybean (18). Thus, modern breeding practices have further shrunk the gene pool by selecting from a small fraction of landraces to produce elite cultivars (25). These observations suggest that the morphological characters of the elite cultivars are genetically controlled by the combined effects of these genomic regions that were selectively swept or lost during domestication and artificial selection. It is possible that genetic diversity in the elite cultivars can be significantly increased by introgression of the favorable variation in wild or landrace in modern breeding program, while additional efforts are needed to balance the selection of agriculturally and adaptively important traits.

Natural populations of *Glycine soja* were strongly influenced by climatic fluctuations. The Quaternary glaciation around 2 MYA resulted in the death of many plant populations, likely including many from *Glycine* (26). Bottlenecks of most natural populations resulted in differentiation by genetic drift and environmental selection between these populations, and the genes acted on by selection (20, 27, 28). It is likely that *G. soja* populations then expanded rapidly through Asia, due to climatic shifts and *G. soja*'s high adaptability (29, 30). This evolutionary history accounts for the wide range of

genetic and phenotypic diversity found among wild soybean populations. The population size of *G. soja* is expanding, allowing for amplified genetic differentiation in varying environments, while *G. max* has been relatively constant as shown in Figure 3 (31). The diverse stress from varying environments caused *G. soja* to develop complicated mechanisms to tolerate many biotic and abiotic stressors. Adaptations to local environments has enabled *G. soja* to possess elaborated mechanisms. Many of the resistance-causal genes have been found to exist in manners of copy number variations (CNVs), such as soybean cyst nematode (SCN) (32), or presence-absence variation (PAVs), such as salt tolerance gene identified in *G. soja* (5). Environmental/niche isolation played a stronger role than isolation by geographic distance in the genetic differentiation of *G. soja*, suggesting the presence of many environmentally tailored adaptations in natural populations (28, 33). Lee, et al. (2015) found that the origin of the tandem duplication of the 31.2-kb segment at the *Rhg1* locus (34), a major QTL conferring SCN resistance, occurred prior to the divergence of *G. max* and *G. soja* or the formation of the ancestor of *G. max*/*G. soja* complex, implying that the copy number variation in *Rhg1* evolved from a *G. soja* population in East Asia, where the SCN populations most likely originated (35).

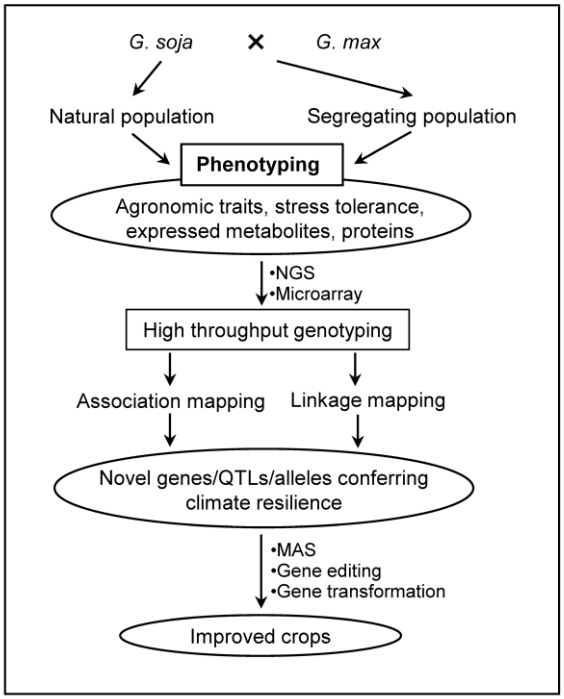



**Figure 3.** Pipeline of crop improvement of *Glycine max* using *Glycine soja*.

The concept of revitalizing the cultivated gene pool with a progenitor species is not novel to soybeans. Plant domestication limits the range of obtainable phenotypes by artificial selection, often removing unforeseen traits of interest down the line. During the 1970s, commercial corn crops (*Zea mays* L.) were devastated by blight affecting as much as 50% of the yield in the United States, until blight-resistant alleles from the wild relative (Mexican maize, *Tripascum dactyloides* L.) were introduced into the domesticated population (36). Examples of the use of crop wild relatives to improve abiotic stress tolerance, biotic stress resistance, and agronomic traits of cultivated crops dates back to the 1980's, seen in major crops: rice, barley, wheat, tomato, potato, and peanut (37). The artificial selection process that lead to the domestication of these crops was blind to the future demands of a growing population and environmental stressors. The development of superior soybean cultivars by incorporating genes/alleles from *G. soja* (Figure 2, Figure 3) is a promising and environmentally friendly solution for soybean improvement, moderating or negating the need for pesticides and fertilizers.

## 4. Discoveries from *Glycine soja*

In 2010, 18 ecotypes were sequenced using *Glycine max* as a reference, serving as the first reported whole-genome sequencing of *G. soja* (19, 38). Structural characteristics of the genome were identified two years later, creating a physical map, useful for investigating true genome architectural differences between *G. max* and *G. soja (39)*. The development and release of the SoySNP50K iSelect Illumina BeadChip in 2013 further promoted the use of wild soybean as a research tool (40). Information regarding soybean research, including all SNPs of USDA accessions are publicly available at the SoyBase Database, serving as an excellent tool for candidate gene discovery in wild soybeans.

Research on *G. soja* thus far has focused on the understanding of domestication and evolutionary history of the *G. soja* transition into the cultivated crop. Few studies have made full use of this wild genetic reservoir. Although the utilization of the genetic diversity contained in *G. soja* is lacking, the potential of *G. soja* as a tool in soybean breeding is beginning to be realized. Table 1 summarizes the recent research and findings using *G. soja* toward soybean improvement. We discuss some representative examples in the following paragraphs and provide an in-depth consideration of these findings. *G. soja* has also been used in many studies as a negative control to identify genes for traits present in the *G. max* population, for example in the study of seed weight or protein content (18, 41). In these cases, *G. soja* gene pool that was not used as a direct source of resistance, tolerance, or improvement is not presented here as such.

### 4.1 Biotic stress resistance

The soybean aphid is native to Asia and was introduced to the United States in 2000. Infestations of the aphids can directly affect biomass and yield, and indirectly affect yield with the transmission of the soybean mosaic virus. The screening of wild soybeans for soybean aphid resistance revealed that three wild soybean plants were resistant to the soybean aphid (*Aphis glycines*) (42). The follow-up studies identified two new QTLs related to aphid resistance, *Rag3c and Rag6*, using linkage mapping (43, 44). Map-based cloning of the two QTLs and the functional verification of candidate genes is needed. It is likely that the candidate genes within the QTL regions encode canonical nucleotide-binding site leucine-rich repeat (NBS-LRR), a resistance protein that plays an important role in potato aphid resistance (45).

The first report of the genetic identification of Foxglove aphid (*Aulacorthum solani*) resistance gene, *Raso2*, in *G. soja* was in 2015. Up to this point, five other aphid resistance genes had been identified, all mapped from *G. max* cultivars. *Raso2,* identified from *G. soja,* differs in response from those isolated from *G. max*, with strong antixenosis and antibiosis responses to the Foxglove aphid (46).

Soybean cyst nematode is the most devastating pest in cultivated soybeans. Breeding soybean cyst nematode (SCN) resistant soybean cultivars in the United States are dependent on very narrow genetic sources. In 2005, 94% of the SCN-resistant cultivars used in Illinois were sourced from a *G. max* accession (PI88788) that was no longer

resistant to the majority of SCN populations found in the soil (9). Nematodes have demonstrated the ability to adapt and overcome resistance in the *G. max*, and therefore new sources of SCN resistance are needed. *G. soja* has been shown to exhibit natural resistance to SCN populations currently affected crops in the United States(9, 11). A large portion of accessions are yet to be screened for resistance, and the investigation into the genetic mechanisms conferring resistance is still underway. So far, studies have reported candidate genes for SCN resistance using linkage mapping (9), and genome-wide association study (7, 11).Interestingly, the candidate genes/QTLs identified from *G. soja* are different from those conferring *G. max* SCN resistance, suggesting species-specific resistance mechanisms. Two SCN-resistance QTLs (cqSCN-006 and cqSCN-007) identified in *G. soja* genotype PI468916 differ from the two major SCN resistance QTLs identified in *G. max*, *rhg1* or *Rhg4* (Kim, et al. 2011). The majority of studies focused on resistance mechanisms to SCN HG type 0 (race3), which is prevalent in the central US. Few studies have worked on SCN HG type 2.5.7, which is prevalent in the southeast US. A recent study identified *G. soja* genotypes resistant to SCN HG type 2.5.7 (Zhang, et al. 2016). A genome-wide association study identified novel candidate genes in *G. soja* relating to HG type 2.5.7 resistance (11). Further analysis of one of the resistant genotypes revealed an entire defense regulatory network, with published RNA-seq data to aid in additional discoveries (7, 47). Given the variability and rapid evolution of SCN populations, it is critical to identify new source of resistance and develop soybean cultivars with broad-spectrum resistance to multiple SCN types with advanced technology, such as genome editing for gene stacking. The variability in the combination of these sources may provide

a unique source for resistance to SCN regionally, providing flexible solutions to mitigate the negative effects of SCN globally.

### 4.2 Abiotic stress tolerance

Soil salinity is a growing challenge for today's crops. As the most important protein-providing crop, cultivated soybean has been cultured in well-ploughed soil, therefore the crop is sensitive to salt stress. The realization of salt tolerance in *G. soja* was especially significant and led to the first investigation of this trait in 1997, but only weak correlation between the genetic markers chosen and the salt tolerance traits was found (48). Further investigation using randomly amplified polymorphism DNA markers revealed associations between 6 markers (OPF05-213, OPF19-4361, OPF19-1727, OPF19-14000-, OPF19-700, OPH02-1350) and salt tolerance in *G. soja*, but with no certainty of the genomic location of these markers (49). The first discovered salt tolerance gene in *G. soja* was *Ncl2*, which differs from the *G. max* salt tolerance gene (50) Screening and sequencing of another wild soybean ecotype W05 lead to the discovery of a different salt tolerance gene, *GmCHX1* (5), suggesting genotype-specific salt tolerance mechanisms. The mechanism of salt tolerance in the wild soybean, Tongyu06311, was found to be regulated by amino acid and organic acid metabolism where compatible solutes were accumulated instead of relying on the consumption of ATP (51). Most recently, aquaporin gene *GmTIP2;1*, was found associated with salt tolerance in *G. soja* (43). Given the role of aquaporins in water transport, it is likely that more aquaporin genes/alleles found in *G. soja* might be associated with drought and salt stress responses.

The discovery of drought tolerance gene *GsWRKY20* from *G. soja* was recently identified and validated by the development of a transgenic soybean overexpressing *GsWRKY20*. The transgenic soybean exhibits increased yield, plant height, and root length over the non-transgenic plant in the same stress conditions. The mechanisms governing the tolerance is related to stomatal density and closure speed (52). This study provides promising solutions to farming in arid and semi-air environments.

A landscape genomics investigation of the abiotic stress tolerance found in the various populations of *G. soja* revealed candidate QTLs associated with varying environmental factors, including monthly precipitation, substrate sand percentage and substrate silt percentage (53). These findings indirectly suggest these genes might play roles in certain abiotic stresses, functional evaluation is needed for each of these genes before they can be considered for soybean crop improvement.

### 4.3 Nutrition

Reduction of saturated fatty acid content of soybeans is highly desired due to its association with cardiovascular disease (54). A GWAS study on seed composition in *G. soja* revealed three new markers associated with palmitic acid levels, a saturated fatty acid, although increased resolution is needed to identify candidate genes. This result lead to the identification of a putative candidate gene, *Glyma.*07G112100, associated with biosynthesis of linoleic acid (55). Linoleic acid is a polyunsaturated fatty acid that is often

partially hydrogenated and attributed to cardiovascular health risks. This recent discovery of a gene controlling both unwanted fatty acids has the potential to make marked improvements on the nutritional profile of the cultivated soybean. The same study identified two candidate genes, *Glyma.14G121400.1* and *Glyma.16G068500.1*, which are associated with a saturated fatty acid, steric acid, and revealed a candidate gene associated with the "good" fatty acid, unsaturated oleic acid (55). In addition, seed protein content of *G. soja* has been shown to be higher than *G. max* on average, likely due to selection on increased yield and oil content (55-57)

### 4.4 Yield related traits

An investigation into the variation of early plant height in wild soybeans revealed significant differences among wild soybean accessions (58). This variation suggests that the trait has been selected for in the natural environment, and may have underlying genetic controls that can be introduced to the cultivated population. However, characterizing *G. soja* height beyond 30 days is especially challenging due to the fragile vining nature of the species, limiting growth rate phenotyping to early stages of maturity. Nevertheless, early growth rates are important indicators of a plant's eventual success and yield (59, 60).

Yield has been strongly selected in cultivars, often making *G. max* superior in this trait. However, one study has identified a QTL in *G. soja* correlated with improved yield. QTL mapping of a cross population of *G. soja* with *G. max* lead to the discovery of a yield

related QTL in *G. soja* on chromosome 14. This QTL is responsible for a 9.4% yield advantage and has been validated in two elite genetic backgrounds (61).

**Table 1.** Findings from *Glycine soja*.

| Trait | Gene or Loci found | Method | Significance | Source(s) |
|---|---|---|---|---|
| **Seed Protein Content** | Marker pA-245 on classic linkage group C | QTL mapping on *G. max* and *G. soja* cross population | This *G. soja* allele is associated with increased seed protein content and dominant over the *G. max* allele | (Diers et al., 1992) |
| **Yield** | Yield QTL B2(U26) with Satt168 being the most significant marker | QTL mapping and interval-mapping analysis | 9% yield advantage in *G. max* individuals carrying *G. soja* QTL | (Concibido et al., 2003) |
| **Soybean Cyst Nematode (*Heterodera glycines*) Resistance** | A245_1 on linkage group G and Satt598 on linkage group E | QTL mapping | QTLs associated with race 3 resistance, confirmed in a backcross population | (Wang et al., 2001) |
| | QTLs cqSCN-006 and cqSCN-007. candidate gene glyma.15g191200 in cqSCN-006 | QTL mapping and fine mapping | Increased resistance when in combination with *G. max* resistance alleles (rhg1, Rhg4, rhg1-b). glyma.15g191200 SNAP gene encodes SNAP proteins - similar to *RHg1* function. | (Kim et al., 2011; Kim and Diers, 2013; Yu and Diers, 2017) |
| | MAPK (*glymaa.18g106800*) and CDPK (*Glyma.18g064100*) | GWAS | Increased resistance to *Heterodea glycines* type 2.5.7 for plants carrying these alleles | (Zhang et al., 2016a) |

| | | | | |
|---|---|---|---|---|
| | Genes involved in pathogen recognition, calcium/calmodulin-mediated defense signaling, jasmonic acid (JA)/ethylene (ET) and sialic acid (SA)-involved signaling, the MAPK signaling cascade, and WRKY-involved transcriptional regulation | Transcroptomics and metabolomics | regulatory network of HG type 2.5.7resistance in wild soybean revealed by defenense associated gene differential expression | (Zhang and Song, 2017; Zhang et al., 2017d) |
| **Salt Tolerance** | *Ncl2* | QTL mapping | First report of salt tolerance gene in soybean separate from the S-100 derived *Glycine max* salt tolerance allele | (Lee et al., 2009) |
| | *GmCHX1* | Whole genome sequencing and QTL mapping | Salt tolerance via ion transporters role to maintain homeostasis during salt stress | (Qi et al., 2014) |
| | *GmTIP2;1 (interaction with GmTIP1;7 and GmTIP1;8)* | Genome wide identification of aquaporins and expression analysis | Overexpression of candidate gene increases salt stress response, likely also associated with drought response | (Zhang et al., 2017a) |

| | | | | |
|---|---|---|---|---|
| **Root Architecture** | *slow anion channel associated 1 like, Auxin responsive NEDD8-activating complex, peroxidase,* and *Apoptosis inhibitor 5-like*(epistatic only) | QTL mapping | SNPs in discovered genes are associated with shorter root or taproot in wild soybeans which is related to drought adaptations | (Prince et al., 2015) |
| **Foxglove Aphid (*Aulacorthum solani*) Resistance** | *Raso2* | QTL mapping | Antixenosis and antibiosis resistance to foxglove aphid | (Lee et al., 2015a) |
| **Alkalinity Tolerance** | ALMT, LEA, ABC transporter, GLR, NRT/POT and SLAH gene(s) | Transcriptomics | Genes upregulated during NaHCO3 stress | (Zhang et al., 2016b) |
| **Seed Saturated Fatty Acid Content** | ss71559532, ss715597684, ss715617910 | GWAS | QTLs associated with lower palmitic acid levels | (Leamy et al., 2017) |
| | *Glyma.14G121400.1* and *Glyma.16G068500.1* | GWAS | candidate genes associated with steric acid production | (Leamy et al., 2017) |

| | | | | |
|---|---|---|---|---|
| **Seed Unsaturated Fatty Acid Content** | *Glyma.16G014000* | GWAS | candidate gene associated with oleic acid levels | (Leamy et al., 2017) |
| | *Glyma.07G112100.1* | GWAS | candidate gene associated with linoleic acid levels | (Leamy et al., 2017) |
| **Drought Tolerance** | *GsWRKY20* | Transcriptomics and transgenic overexpression | overexpression of GSWRKY20 enhances tolerance to drought stress | (Ning et al., 2017) |
| **Soybean Aphid (*Aphis glycines*) Resistance** | *Rag3c and Rag6* | QTL mapping | Genes confer antibiosis resistance to aphids | (Zhang et al., 2017f) |

## 5. Limitations

*G. soja* and *G. max* have the same chromosome number and can be crossed to create fertile hybrids, which can transfer useful genes from *G. soja* to *G. max* by traditional breeding practices. But, introgression of *G. soja* into *G. max* can result in linkage drag by bringing in unwanted parts of the genome together with the selected genes due to linkage disequilibrium. Linkage drag usually result in drags of unfavorable traits, such as reduced yield, apt to shatter, and lodging etc. But this limitation can be resolved by the rapid progress of biotechnology, such as soybean genetic transformation and genome editing.

## 6. Perspective

*G. soja* holds great potential to provide novel genes/alleles in soybean and other legume species for crop improvement. A prerequisite for using *G. soja* to improve the agricultural potential of *G. max* is to dissect the genetic architecture underlying the traits of interest and uncover the molecular, physiological, and biochemical mechanisms involved. Some studies using GWAS and linkage mapping have facilitated the dissection of potentially useful traits, and an integration of these strategies coupled with investigation of gene expression (transcriptomics), protein expression (proteomics), and metabolite profiling (metabolomics) can be helpful to understand the mechanisms involved in phenotypic expression. High throughput sequencing technology and biotechnology (such as genome editing) can significantly facilitate novel gene discovery and transfer useful genes to soybean.

Thus far, *G. soja* has been used to determine the nature of domesticated traits in *G. max* (18, 20), while the associated genetic variation in *G. soja* has not been fully investigated. Knowledge of these genomic regions with selective sweeps in *G. soja* is helpful to understand the evolutionary mechanisms that produced and altered traits in *G. max* resulting from domestication, and can facilitate the use of *G. soja*-derived variations to improve soybean crop. Investigations of phenotypic variation should shift toward the screening for biotic resistance and abiotic tolerance under different environmental conditions and stressor regimes. Such a shift in research focus would produce results (and future directions) that are more relevant to current needs as they pertain to growing food needs and climate change. Although some studies have identified SCN-resistant *G. soja* accession by screening a small portion of *G. soja* collection at USDA germplasm (9, 11), a systematic screening is needed by collecting more ecotypes from many different natural populations, taking into account the observed association between the SCN resistance and *G. soja* origins.

Crop improvement by using wild soybean has been progressing, but is still lacking in some promising areas. Currently, no work on the Soybean Mosaic Virus (SMV) has been reported in *G. soja*, a native host to the virus. Natural fields of *G. soja* infected with SMV have been observed in South Korea (62). Given the observed resistance to biotic factors within the wild population, research in this area could be a promising avenue for developing SMV-resistant soybean crop. The observed variation in early vigor traits in

wild soybeans is also an understudied area with potential to improve the early success and eventual yield of soybean crop.

The chloroplast genome has been used in only a few studies with *Glycine*, and in those cases only to discuss diversity and domestication. Plastids regulate photosynthesis and the production of many metabolites; it would not be a great leap to assume that some of those attributes are a) important to cultivation and b) phenotypically found to vary in the wild population of soybean. The chloroplast genome of soybeans underwent multiple selection events, and diverged into two major haplotypes early on in domestication (63). Whole chloroplast genome assembly of *G. soja* compared to nine other *Glycine* species reported high conservation with no major rearrangements within *Glycine.* Phylogenetic topology is consistent between nuclear and plastid genomes, and an in-depth investigation into plastid genomic differences between *G. soja* and *G. max* may reveal attributes worth studying for crop improvement (64).

Efforts should be made to conserve the wild soybean populations. It has been shown that these diverse populations exhibit desired phenotypes and diverse genotypes that can be exploited for crop improvement. Conservation of these populations is required in order to continue benefiting from them. A recent study in Japan reported low risk of gene flow between genetically modified and wild soybeans (65). Soybeans self-fertilize, and therefore gene flow can easily be maintained by reducing seed transport and spillage and increasing the distance between crops and wild populations.

The natural populations of wild soybean are distributed in East Asia and Russia. Current research mainly uses a little over 1000 genotypes from USDA collections. To fully investigate the natural variation of this species and identify useful resources, international collaboration between scientists from diverse disciplines is needed. More findings from wild soybean are expected in the near future.

## 7. Conclusion

The wild soybean has been the source of many advancements in crop improvement and aided in research related to the evolution of the soybean. The domestication history of the soybean is now widely understood. Continued research and screening of more ecotypes for potentially useful traits in *G. soja* will help to reveal additional valuable genetic sources for crop improvement. More collaboration and effort should be put toward finishing what has already begun - to link the discovery of candidate genes with studies on relevant gene regulatory networks and ultimately, the molecular and practical procedures for crop improvement. Lastly, the conservation of natural soybean populations needs to be promoted to ensure the continued existence of adaptable, wild traits that could later be used for crop improvement. International collaboration is needed to carry out goal-directed, comprehensive research of *G. soja* to make full use of their untapped genetic reservoir.

# CHAPTER 2: GENOME WIDE ASSOCIATION STUDY OF EARLY VIGOR TRAITS IN WILD SOYBEAN

## 1. Introduction

The cultivated soybean, *Glycine max*, is an important legume crop that supplies the majority of the protein meal and oilseed worldwide (66). Crop improvement is a continuous necessity as the demands in agriculture increase due to a changing environment and rising population. Modern soybean crops are challenged by abiotic and biotic stress. In order for soybean production to keep up with the growing population, novel modifications must be made to the current crop beyond modern breeding practices (67). The wild counterpart to the cultivated soybean, *Glycine soja*, is analyzed for genetic associations to early vigor traits that are not present in the cultivated population.

*G. max* diverged from *G. soja* up to 0.8 MYA, leaving much of the wild genetic variability behind (17, 20). The cultivated soybean further diverged as a result of domestication in China 6000-9000 years ago, however the wild soybean continues to inhabit a wide range of areas across Eastern Asia (3, 19). The gene pool in wild soybean retains the genes and alleles lost during the process of artificial breeding practices, and more importantly, the initial domestication of the soybean, which contributes to 16.2% and 31% reduction in genetic diversity respectively (68). Selection and improvement within *G. max* alone has

been shown to have a significant effect on diversity on all twenty chromosomes (69, 70). By exploiting the genetic diversity harbored in the wild soybean we are able to discover novel sources of agronomical superiority.

Early vigor is a combination of traits that determine the eventual success and yield of a plant. The early vigor phenotypic traits studied in *G. soja* are node count, inter-node length, early growth rate, and early plant height. Early growth rate (EGR) and early plant height (EPH) are further analyzed for genotypic dissection using GWAS. These traits have not been studied or used for genetic association in wild soybean thus far. The node count and inter-node length are representative of the density of plant foliage, as each node results in a branching compound leaf. There is a significant correlation between the yield per soybean plant and the branches on that plant (71). Measurements of mature plant height in wild soybeans is relatively problematic due to its fragile stem and vining nature. Therefore, EPH can be used as an analysis of height while the plant is still in early growth stages. In cultivars of *G. max*, a positive linear relationship has been observed between plant height and seedling rate where seedling rate has a positive effect on the overall yield (59). EGR is an important trait, used to determine the ultimate success of the plant. Seed vigor is determined partially by an evaluation of seedling growth (60), which was initially established in the 1950's to represent the success and productivity of the resulting plant (72). All traits analyzed are representative of plant success and a measure of vigor.

A genome wide association study (GWAS) is used in this study to determine genetic associations with complex agronomically beneficial traits, EPH and EGR. GWAS has

significant advantages over the tradition approach of complex candidate loci discovery, quantitative trait loci (QTL) mapping. QTL mapping relies on a crossed mapping population from only two individuals to contain all genetic diversity for the trait being analyzed. In addition, QTL mapping is limited by its resolution in the mapping population that can be affected by recombination while crossing (73). Although first demonstrated for use in human disease (74), GWAS has been successfully used with plants for over a decade to identify genes responsible for quantitative traits (75). While the QTL mapping approach uses just two genotypes for genetic variation discovery, GWAS utilizes an unlimited sample size and takes advantage of the genetic and phenotypic diversity contained in the extensive natural population in order to accurately describe associations with higher resolution than QTL mapping (76, 77). The soybean is an ideal candidate for GWAS, as it can be maintained by self-fertilization, allowing for repeat screening of genetically similar individuals. GWAS has previously been used to dissect phenotypic traits related to biotic stress resistance (21, 78), abiotic stress tolerance (43, 53), and seed composition (55, 79) in wild soybean.

The value of crop wild relatives in crop improvement has been progressively recognized in the past decades (37). The wild soybean, *G. soja*, has been used in the genetic dissection of soybean growth (80, 81), yield (61), seed composition(55, 82), abiotic stress tolerance (10, 48, 50-53), and biotic stress resistance (7, 42-44, 47, 78, 83-88). However, using *G. soja* to study early vigor phenotypic traits, such as node count, inter-node length, early growth rate, and early plant height has never been reported. To explore the genetic diversity harbored in the wild soybean, we used GWAS to effectively characterize the genetic

architecture of early vigor traits. This study aims to make use of leading-edge quantitative trait loci discovery methods in order to dissect the genetic background of agronomically beneficial traits, EGR and EPH, and their relationships to inter-node length and node count, to further crop improvement techniques.

## 2. Results

### 2.1 Correlation in phenotypic traits

All trait relationships studied on 225 ecotypes of *G. soja*, inter-node length to node count, inter-node length to EGR, inter-node length to EPH, node count to EPH, node count to EGR, and EGR to EPH, expressed significant positive correlations, $p < 0.001$, by Spearman's correlation test (Table 2). As a trend, plants with higher growth rate and height had more nodes and the distance between each of the nodes was greater.

**Table 2.** Positive significant correlation between traits: EGR, EPH, Node count, and Inter-node Length by Spearman's ρ correlation coefficient. Mean, standard deviation (StDev), and coefficient of variance (CV), measurements by cm.

| | Mean | StDev | CV | EGR | EPH | Node Count | Inter-node Length |
|---|---|---|---|---|---|---|---|
| **EGR** | **16.02** | **8.34** | **52.05** | | 0.9861 | 0.7023 | 0.8674 |
| **EPH** | **244.68** | **127.2** | **51.99** | | | 0.6911 | 0.8838 |
| **Node Count** | **4.34** | **0.93** | **21.33** | | | | 0.3449 |
| **Inter-node Length** | **54.55** | **24.97** | **45.76** | | | | |

### 2.2 Population structure of ecotypes

An estimation of population structure for all 225 accessions resulted in three assumed populations. The inferred clusters, based on ΔK value (89), from STRUCTURE analysis account for 23.7%, 35.2%, and 41.1% of the populations respectively, visualized in Figure 4. The constructed Newick tree, from calculated genetic distance of 31, 726 SNPs, supports

the population estimation of three clusters (Figure 4A). The genetic distribution of the populations is seen to highly correlate with the geographic locations of these wild soybeans. Group 1 corresponds to collections from China and Russia, with only one contrary observation, PI424008A, collected from South Korea. Group 2 primarily includes South Korea accessions with just two individuals, PI549037 and PI479751, originating from China. Group 3 includes all accessions from Japan with one diverging individual from South Korea, PI407249.

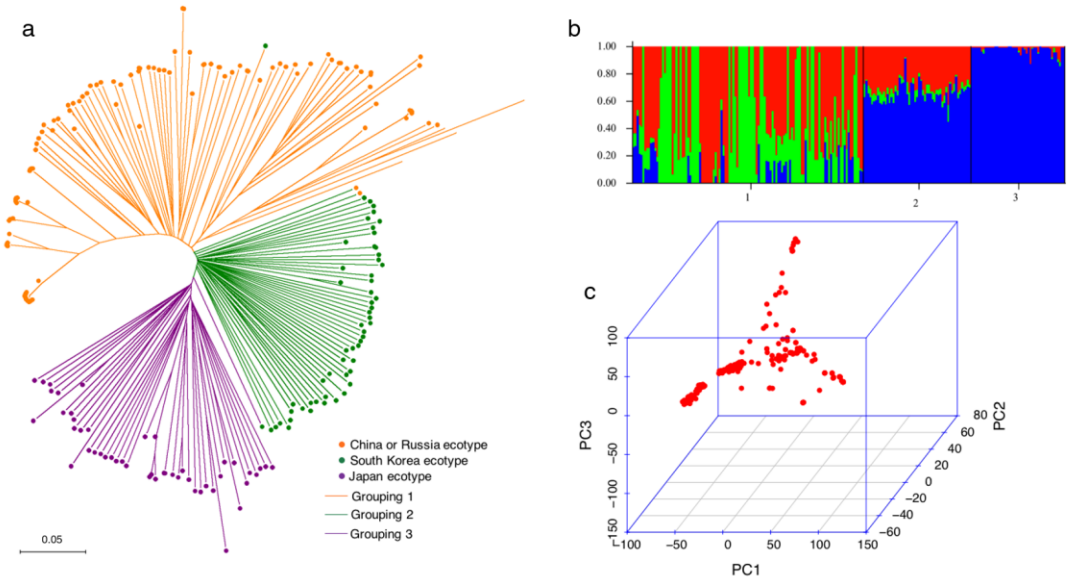



**Figure 4.** Population Structure Analysis. **(a)** Unrooted Newick Tree of ecotypes from 31726 SNPs. Groups primarily cluster by ecotype location. (**b)** STRUCTURE bar plot, K=3, inferred clusters 1 = 0.237, 2 = 0.352, 3= 0.411. **(c)** Plot of first three principle components with all 225 ecotypes.

The principle component analysis (PCA) results in the first three principle components accounts for 19.1% of the genetic variation in all populations (Figure 4c), which are used to control the population structure in the MLM for GWAS.

## 2.3 Genome wide association study

The MLM of 31,726 filtered SNPs combined with kinship and PC control, resulted in four significant SNPs associated with EGR and 12 significant SNPs associated with EPH with chromosome-wide FDR adjusted $p < 0.05$ (Table 3). All significant SNPs associated with EGR are also significant for EPH. The most highly associated SNP to EGR and EPH, ss715598271, is significant ($p < 0.05$) by all p-value adjustment and threshold tests (chromosome-wide Bonferroni, genome-wide Bonferroni, and genome-wide FDR). SNP marker ss715598271 explains 10.85% of EGR phenotype variation and 12.42% of EPH phenotype variation. Bracketing markers to ss715598271, ss715598270 and ss715598272, are also shown to be significant by chromosome-wide Bonferroni threshold adjustment in

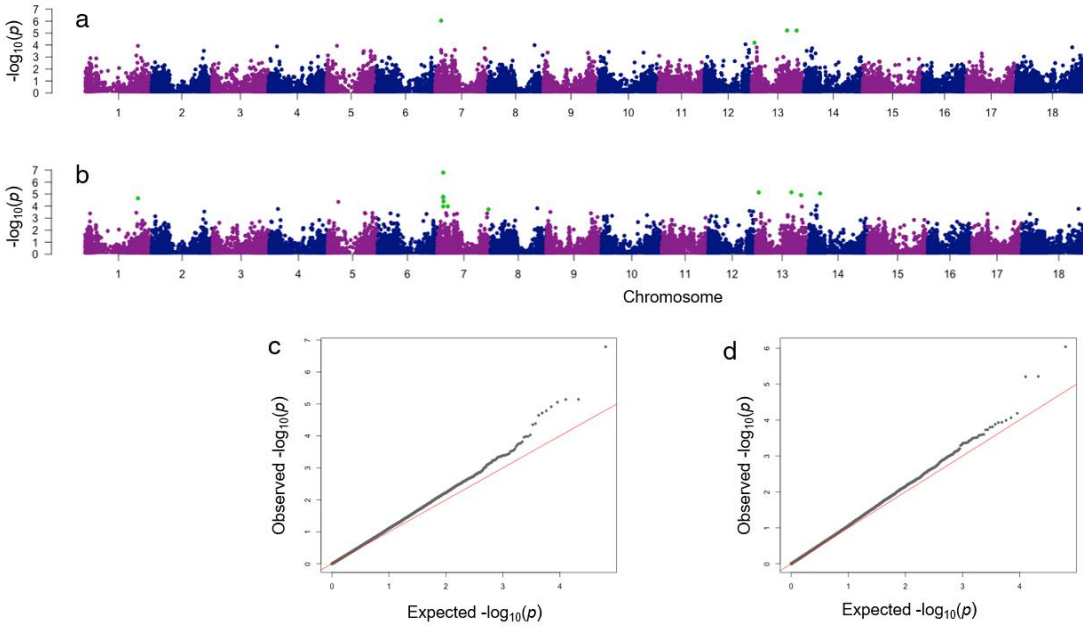



**Figure 5.** GWAS Results. **(a)** Manhattan plot of MLM for EGR. Significant SNPs from chromosome-wide FDR adjustment are highlighted in green. **(b)** Manhattan plot of MLM for EPH. Significant SNPs from chromosome-wide FDR adjustment are highlighted in green. **(c)** Quantile-quantile plot of MLM for EGR. **(d)** Quantile-quantile plot of MLM for EPH.

EPH, explaining 8.34% and 8.17% of the phenotypic variation respectively. Manhattan plots of EGR and EPH (Figure 5a, Figure 5b), visualize all significant SNPs by chromosome, with corresponding Q-Q plots (Figure 5c, Figure 5d).

### 2.4 In-depth candidate loci investigation

Highly associated SNP to EGR and EPH, ss715598271, and significant bracketing markers, ss715598269, ss715598270 and ss715598272 for EPH collocate with previously identified QTL, Plant Height 19-5(90). In addition, this candidate region is within 50kb adjacent to known QTLs, Plant Height 2-2 and Plant Height 2-4(91), which collocate with ss715598304, an intergenic marker significantly associated with EPH (Table 3). Significant markers on chromosome 13 correlate with two previously described QTL's related to plant height, Plant Height 25-6 (92) and Plant Height 26-11(93) (Table 3). Although this correlation gives confidence in the findings of this GWAS, the association of ss715598304 with EGR and EPH would be more reliable if the LD in the area of the marker was high and other markers in the area were also found to be significantly associated with the traits studied. Therefore, we do not have enough confidence to suggest that this area is as highly significant in plant early vigor as that on chromosome 7 in the wild soybean population.

The 65.3 kb region within significant markers from 4915929 bp to 4928272 bp on chromosome 7 is further analyzed by pairwise linkage disequilibrium, showing high linkage within this region (Figure 6). Two markers, ss715598269 and ss715598270, are within the intron and 3' untranslated region of gene *Glyma.07G055800.1* respectively. Two

markers, ss715598271 and ss715598272 are within the intron of gene *Glyma.07G055900.1*. No significant linkage disequilibrium is found near other significant

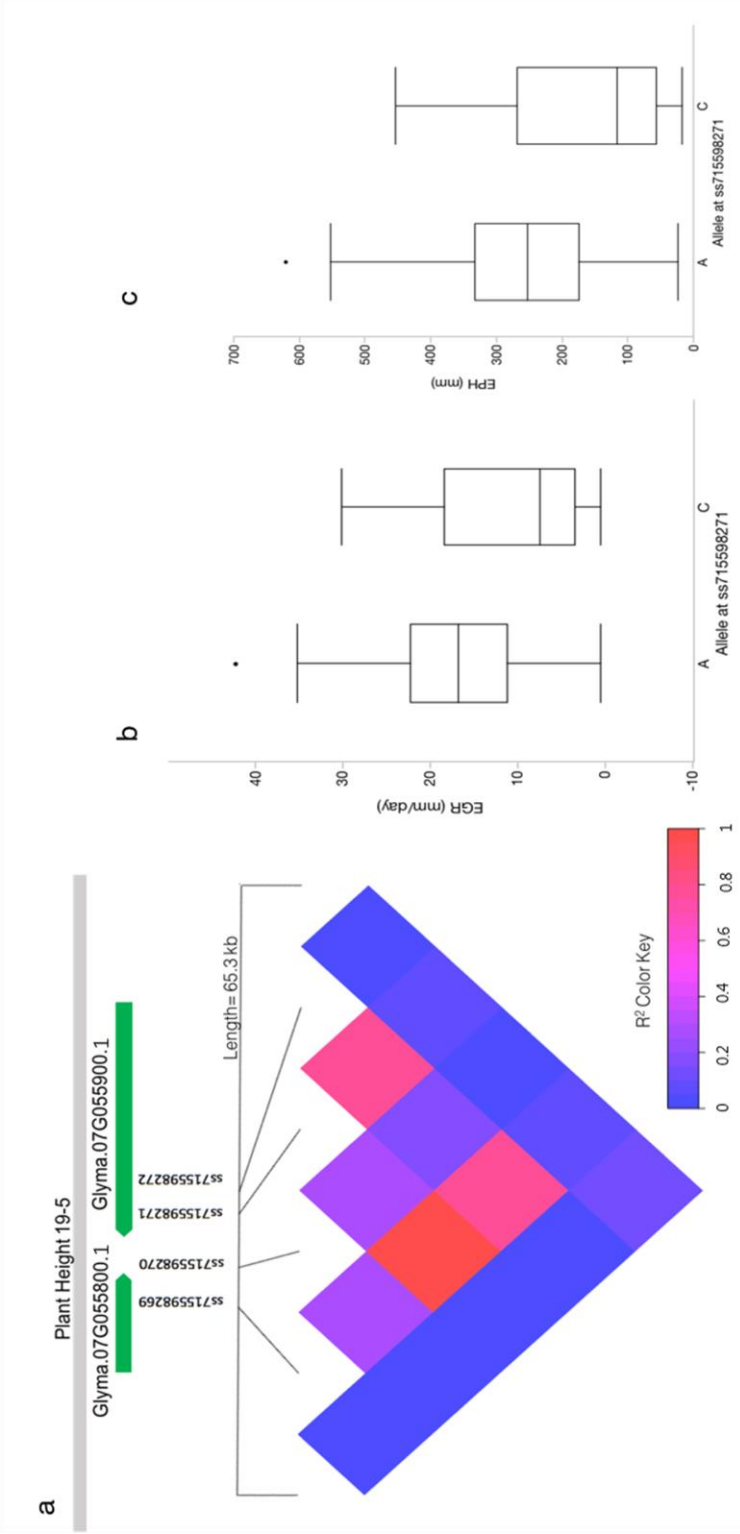



**Figure 6.** Pairwise LD with local genes of interest and Allele to Phenotype Relationship. **(a)** *Glyma.07G055800.1*, *Glyma.07G055900.1*. and known QTL, Plant Height 19-5 shown in relation to SNPs ss715598269, ss715598270, ss715598272, significantly associated with EPH and ss715598271, significantly associated with both EPH and EGR **(b)** *Glyma.07G055800.1* allele at ss715598271 A/C comparison to EPH (mm) at marker ss715598271, p-value <0.001. **(c)** *Glyma.07G055800.1* allele at ss715598271 A/C comparison to EGR (mm/day) at marker ss715598271, p-value <0.001.

SNPs and are therefore were not investigated further for candidate genes.

There is a significant relationship between ss715598271 allele polymorphism (A/C) and studied traits, EPH and EGR. The A allele morph is correlated with higher EGR and higher EPH than the C allele morph at this marker ($p < 0.02$) (Figure 6a, Figure 6b). The relationship between allele and trait suggests that ss715598271 is highly linked with causative SNPs.

Candidate genes, *Glyma.07G055800.1* and *Glyma.07G055900.1*, are found to be associated with EPH and EGR. *Glyma.07G055800.1* is predicted to be a transmembrane protein containing DoH and Cytocrhome b-561/ferric reductase transmembrane domains(94). *Cytochrome b561/ferric reductase transmembrane with DOMON related domain* protein (*CYB561*) is the closest homolog in *Arabidopsis*. This protein is responsible for catalyzing transmembrane electron transfer by ascorbate, which is a key metabolite in growth and development of plants (95). *Glyma.07G055900.1* is predicted to be a Tetratricopeptide repeat (TPR)-like super family protein, playing a role in protein binding and translation initiation (94). The closest homolog in *Arabidopsis* is *Reduced Chloroplast Coverage 1* (*REC1*), which is associated with regulating the compartment size of chloroplasts (96).

**Table 3** Significant SNP markers associated with EGR and EPH. The q-value given is the Chromosome-wide FDR adjusted p-value. Significance in CB (Chromosome-wide Bonferroni threshold), GB (Genome-wide Bonferroni threshold), and GFDR (Genome-wide FDR adjustment) with $p < 0.05$ are indicated with a (*). The position is physical position on the chromosome (Chr.) in BP.

| SNP | Chr. | Position | r2 | Location | Associated Gene | Associated QTL | p-value | q-value | CB | GB | GFDR |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **EGR** | | | | | | | | | | | |
| ss715598271 | 7 | 4924020 | 0.1085 | Intron | Glyma.07G055900.1 | Plant Height 19-5 (Wang et al., 2004) | 9.14E-07 | 0.00143 | * | * | * |
| ss715614175 | 13 | 19487316 | 0.071 | Intergenic | - | Plant Height 26-11(Sun et al. 2006) | 0.000065 | 0.04206 | - | - | - |
| ss715615103 | 13 | 31173270 | 0.0911 | Intergenic | - | | 6.12E-06 | 0.00605 | * | - | - |
| ss715616082 | 13 | 39280839 | 0.0944 | 5UTR | Glyma.13G292800.1 | | 6.23E-06 | 0.00605 | * | - | - |
| **EPH** | | | | | | | | | | | |
| ss715579500 | 1 | 45269059 | 0.0792 | Intergenic | - | | 0.0000226 | 0.02802 | * | - | - |
| ss715598269 | 7 | 4915929 | 0.0682 | Intron | Glyma.07G055800.1 | Plant Height 19-5 (Wang et al., 2004) | 0.000104 | 0.02714 | - | - | - |
| ss715598270 | 7 | 4918294 | 0.0834 | 3UTR | Glyma.07G055800.1 | Plant Height 19-5 (Wang et al., 2004) | 0.0000164 | 0.01002 | * | - | - |
| ss715598271 | 7 | 4924020 | 0.1242 | Intron | Glyma.07G055900.1 | Plant Height 19-5 (Wang et al., 2004) | 1.62E-07 | 0.00025 | * | * | * |
| ss715598272 | 7 | 4928272 | 0.0817 | Intron | Glyma.07G055900.1 | Plant Height 19-5 (Wang et al., 2004) | 0.0000192 | 0.01002 | * | - | - |
| ss715598304 | 7 | 5214440 | 0.0816 | Intergenic | - | Plant Height 19-5 (Wang et al., 2004), Plant Height 3-3 (Mansur et al., 2993), Plant Height 25-6 (Guzman et al. 2007) | 0.0000411 | 0.01609 | - | - | - |
| ss715598895 | 7 | 8788505 | 0.0668 | Intron | Glyma.07G094100 .1 | | 0.000104 | 0.02714 | - | - | - |
| ss715598145 | 7 | 42926704 | 0.0628 | CDS | Glyma.07G251700.1 | | 0.000186 | 0.04161 | - | - | - |
| ss715614175 | 13 | 19487316 | 0.09 | Intergenic | - | Plant Height 26-11(Sun et al. 2006) | 7.24E-06 | 0.00703 | * | - | - |
| ss715615103 | 13 | 31173270 | 0.0893 | Intergenic | - | | 7.18E-06 | 0.00703 | * | - | - |
| ss715616082 | 13 | 39280839 | 0.0874 | 5UTR | Glyma.13G292800.1 | | 0.0000122 | 0.00789 | * | - | - |
| ss715620138 | 14 | 9595999 | 0.0873 | Intergenic | - | | 8.84E-06 | 0.01383 | * | - | - |

## 3. Discussion

Significant correlations were found between all traits studied, suggesting that EPH, EGR, inter-node length, and node count are all related phenotypic traits (Table 2). The positive relationship between inter-node length and node count, inter-node length and EPH, and internode-length and EGR, demonstrates the combined effect of elongation of the inter-nodes along with the addition of new nodes in early growth of the soybean. Each node of the wild soybean is accompanied by one branching compound leaf. Therefore, the observed increase in node count with EGR and EPH relates directly to foliage increase and acquisition of available light, having a cumulative effect on growth. The total yield has previously been shown to be positively correlated with the branching of a soybean (71). A positive correlation between all early vigor traits tested demonstrates that the traits chosen are reliable indicators of early vigor and eventual success of a plant.

The results from STRUCUTRE analysis, PCA, and Newick tree construction suggest a population structure of k=3 with high amounts of admixture between the populations (Figure 4). The genotypic clustering is predominantly governed by the geographic location of these ecotypes, grouping into China and Russia, Japan, and South Korea. A high level of admixture has been seen in previous studies of wild soybean populations, specifically in regards to South Korea and Japan populations, which is consistent with this study (28). Our results, however, suggest that China and Russia cannot be clearly defined as separate populations, most likely due to the fact that their locations are contiguous aside from country borderlines or possible materials exchanges. In addition, environmental effects are

shown to have a much higher influence on genetic grouping of wild soybeans than the geographic locations themselves, which explains the clustering of China and Russia together (28).

We discovered a total of 12 significant SNPs for EPH with 4 shared significant SNPs for EPH and EGR. The most compelling results are found on chromosome 7, in which a contiguous set of markers, ss715598271, ss715598270, and ss715598270, are highly significant, revealing the importance of this genomic location to EGR and EPH. This locus is also collocated with a known QTL, Plant Height 19-5 (90), and is adjacent to two QTLs Plant Height 2-2 and Plant Height 2-4 (91). Although these consistently mapping QTLs were previously identified, variation in this region might be unique to *G. soja* as previously described in the finding of a *G. soja*-unique salt-tolerant gene *Gm*CHX1(5). This location is further verified with LD analysis, showing that significant linkage disequilibrium is evident in this locus. High linkage disequilibrium can suggest that this area has been selected for in its natural environment, which is expected of plant height and growth-related loci. We were able to determine an allelic correlation with EPH and EGR at marker ss715598271. The A allele morph is correlated with higher EGR and higher EPH than the C allele morph at this marker. This marker is unlikely the causative SNP, but linked to one unmapped in *G. soja*. This is supported by surrounding marker significance and high LD in the region.

Two candidate genes were identified as significantly related to EPH and EGR, *Glyma.07G055800.1* and *Glyma.07G055900.1* on chromosome 7. Significant markers are

located within the introns of these genes and the untranslated 3' region of *Glyma.07G055800.1*. *Glyma.07G055800.1* codes for a Cytochrome b561 transmembrane protein, or *CYB561* in *Arabidopsis*, a transmembrane protein involved in electron transport(97). *Glyma.07G055900.1*, codes a Tetratricopeptide repeat like super family protein, homologous to *REC1* in *Arabidopsis*. The TPR domain protein has been shown to be involved in plant height related phenotypes in Maize and *Arabidopsis* (98, 99) and is a known motif in plant hormone signaling such as auxin, gibberellin, cytokinin responses (100, 101). It is not unlikely that multiple individuals of this protein family share a similar influence on traits in various locations in the genome. It has been proposed that *REC1* is involved in establishing and maintaining chloroplast coverage in *Arabidopsis*, and could be manipulated in order to influence energy intake and yield (96). G. *soja* type *REC1*, *Glyma.07G055900.1*, is likely involved in similar chloroplast coverage and plant hormone signaling pathways to those found in Maize and Arabidopsis, with a direct relationship to photosynthesis efficiency and growth rate, and therefore is an ideal gene to further investigate for crop improvement.

This study identifies two candidate genes, *Glyma.07G055800.1* and *Glyma.07G055900.1*, related to early vigor traits, EGR and EPH. These traits are highly beneficial to agriculture, and have been artificially selected for in breeding practices in the cultivated soybean, *G. max*. However, the extent of adaptation by selection in the cultivated soybean population is limited by the gene pool of the initial landraces. We explored the genetic architecture of these traits in the more diverse wild soybean *G. soja*, to further improve our cultivation

practices beyond the limited genotype of *G. max*. Future direction includes validation of these genes' associations to early vigor traits in cultivated soybeans by gene transfer.

## 4. Materials and Methods

### 4.1 Plant materials and phenotyping

In total, 225 *G. soja* accessions that were obtained by the USDA Soybean Germplasm Collection were used for measurements and analysis. The original geographic distribution of these accessions includes areas in China, Japan, Russia, and South Korea (Figure 7). All seeds were manually scarified and germinated on filter paper for three days, after which 4-5 seedlings per genotype were transplanted into MirocalGro soil in separate cells of a 3 x 5 growing tray. Optimal growing conditions were kept constant at 27°C and 12h light/day in the greenhouse at University of North Carolina at Charlotte. The plants were watered regularly to keep soil moist.

Four phenotypic traits were recorded or measured with a caliper to the nearest 0.1 mm on each accession: node count, internode length, EGR, and EPH. EPH was measured at 20 days after germination, and node count, EGR, and internode length were obtained as the average of three recordings take at days 7,14, and 20 after germination. For each accession, measurements from two or three seedlings, quality filtered by coefficient of variance and noted damage during growth, were averaged for each trait used in the analysis.

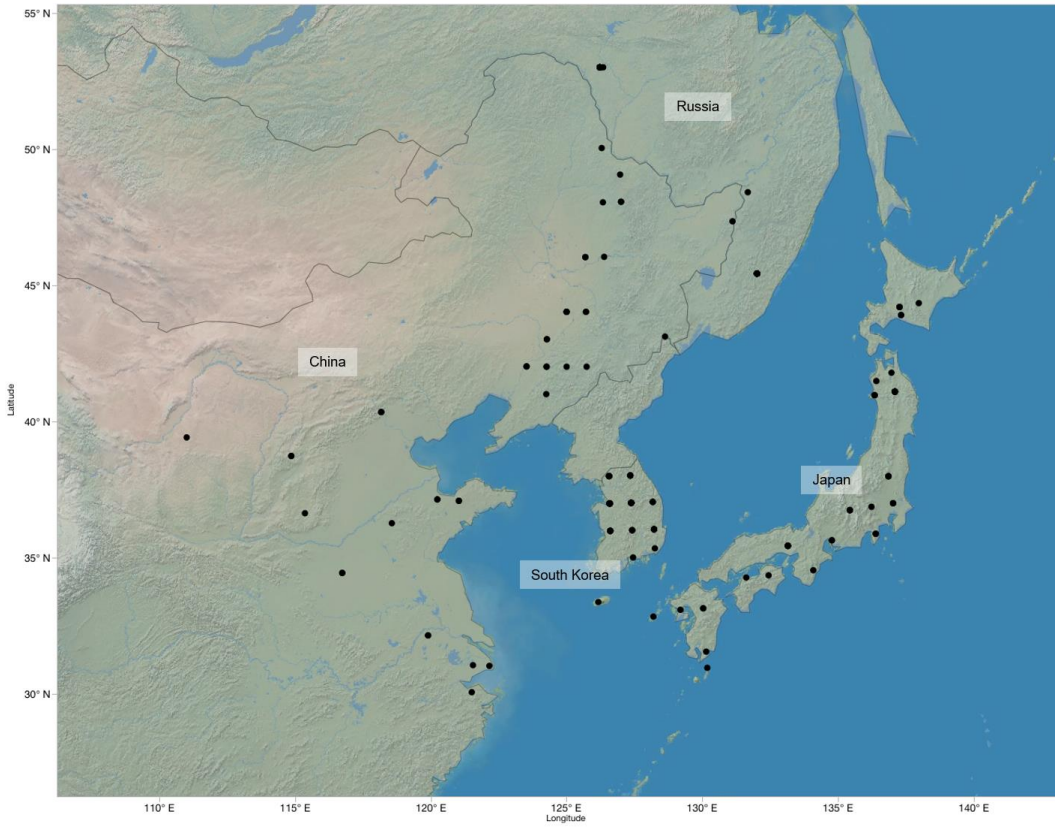



**Figure 7.** Geographic distribution of 225 *G. soja* accession. Each point marks a location

## 4.2 Genotypic data

Previously identified markers, single nucleotide polymorphisms (SNPs), for all 225 *G. soja* accessions were obtained from SoyBase (http://soybase.org/snps/*)(94).* All markers were originally determined by the use of the Illumina Infinium SoySNP50k iSelect BeadChip, with 52,041 total verified SNP markers(40). All markers with a minor allele frequency (MAF) < 0.05 or missing rate of >10% were filtered out of the analyzed data, leaving 31,726 SNPs. All cleaned data were imputed using BEAGLE (v 3.3.1) (102-104).

### 4.3 Phenotype analysis

Pair-wise associations among the four traits were evaluated with Spearman's correlation analysis, due to non-normal distributions, to determine the relationship between traits. Specifically, the relationship between height and growth traits with the number and frequency of nodes, and therefore relative foliage quantity. Phenotype data was not normalized due to sufficient sample size for GWAS, and to avoid error, such as false positives (105).

### 4.4 Analysis of population structure

A PCA was conducted using and GAPIT package (106). A genetic distance (1-IBS) matrix, calculated in TASSEL (107), was used to build a Neighbor Joining (NJ) tree and visualized by MEGA 7 (108). STRUCTURE software was used to validate the determined population structure with 10,000 randomly selected markers with parameter sets K=1 to 10, 10,000 Burnin period length, and 50,000 MCMC reps after Burnin (109).

### 4.5 Genome-wide association study

A Mixed Linear Model (MLM) in TASSEL (107) was preformed to analyze the associations of EGR and EPH with SNPs for all 225 accessions. A principle component (PC) value of three, selected by analysis of quantile-quantile plots at various PC values, and a scaled IBS Kinship matrix was used to control for the population structure. Multiple significance threshold tests were used in order to substantiate the determination of significant SNPs. The significance threshold was determined by chromosome-wide false discovery rate (FDR) resulting in an adjusted p-value (q-value) for each marker (110).

Genome-wide FDR, along with genome-wide and chromosome-wide Bonferroni adjustments (111), were also used to validate the significant SNPs at $p < 0.05$ or $q < 0.05$ cutoff.

### 4.6 Investigation of candidate loci

All genes within 50 kb of significant SNPs were evaluated for potential association with each phenotype. The annotated soybean reference genome, Wm82.a2.v1 (SoyBase, http://soybase.org), was used to determine these genes along with further investigation into the function using Phytozome (112), TAIR (113), and BLAST2GO (114). Known QTL's for each phenotype from SoyBase (94) in each candidate loci were considered for validation of candidate genes. SNP allele to phonotype comparison is done by non-parametric Median tests. Pairwise linkage disequilibrium (LD) was calculated with TASSEL and visualized with the LDheatmap R package (115)

# CHAPTER 3: WILD SOYBEAN CONFERS NOVEL RESISTANCE TO THE BIGGEST SOYBEAN PATHOGEN, SOYBEAN CYST NEMATODE

## 1. Introduction

The soybean (*Glycine max* (L.) Merr.), is an important legume crop that supplies more than half of the word's vegetable fats, oils, and protein meal (116). Soybean cyst nematode *(Heterodera glycines*, SCN) is the most destructive soybean pest, and is a growing problem (117). The distribution of SCN infected regions has shown rapid growth since the initial infection in 1954 in North Carolina, due to natural spread of pathogens and the lack of resistant cultivars (118).

SCN resistant crops in the United States are dependent on few sources of resistance. There are three sources of SCN resistant *G. max* used for breeding commercial varieties in the United States: PI88788, Peking (PI548402), and PI437654. The majority, over 95%, of which are sourced from PI88788 (119). In 2005, 94% of the SCN resistant cultivars used in Illinois sourced from PI88788 were no longer resistant to the majority of SCN populations found in the soil, including SCN race 5 (9). In fact, SCN resistant cultivar, Peking, is the main source of resistance to SCN race 5 in the United States, and only makes up less than 5% of the cultivars used. The least used cultivar, PI437654, demonstrates broad resistance and is able to resist SCN race 5, although it is rarely used due to risks associated with broad resistance and complications with the genetic background and linkage to less

favorable traits (120). Nematodes that are adapted to PI437654 type resistance are able to overcome all other sources of resistance, and therefore cultivars sourced from this accession should be used sparingly (121, 122). Nematodes have demonstrated the ability to adapt and overcome all resistance found in the *G. max* gene pool (9, 121-123), creating a high demand for new sources of SCN resistance.

Research on SCN resistance dates back to 1960, with the discovery of the first *rhg* loci (Resistant to *Heterodera glycines*) (124). It was initially assumed that the mechanism of SCN resistance in the cultivated soybean was comprised of Leucine-rich-repeat-Kinase (LRR-Kinase) genes, which mapped closely to *rhg* loci and were known to confer resistance in other crop species (125, 126). It was later found that resistance by the two strongest loci, *rhg1* and *Rhg4*, was independent of LRR-Kinase genes (127, 128), and instead is conferred by a novel resistance strategy. A complex mechanism of resistance in the cultivated soybean consists of copy number variation of causative genes at the *rhg1* and *Rhg4* loci, and the genetic composition of these genes and associated promoters (34, 129-133). Two main types of resistance are found within the cultivated soybean, Peking-type and PI88788-type, with Peking type being the only type resistant to SCN race 5(119). Peking-type resistance is conferred by low-copy SNAP (soluble NSF attachment protein) gene (*Glyma.18G022500*, alternative ID: *Glyma18g02590*) at *rhg1-a* in combination with resistant-type SHMT (serine hydroxymethyltransferase) gene (*Glyma.08G108900*, alternative ID: *Glyma08g11490*) at *Rhg4 (131)*. Broad resistance in PI437654 is derived from Peking-type, with a high-copy *SHMT* allele (133). PI88788 type resistance is conferred by a high-copy number of three genes, including *SNAP*, at rhg1-b, with no

presence of resistant-type *Rhg4* (34, 129). Duplication and selection at resistant *rhg1* occurred in wild soybean, but the resistant *Rhg4* allele emerged after domestication (32, 134). It is likely that the wild soybean population has an ancestral and separately derived method of SCN resistance, not found in the cultivated gene pool.

The wild counterpart to the cultivated soybean, *G. soja* (Siebold & Zucc.), is analyzed for SCN resistance novel to the crop gene pool. The domestication bottleneck resulted in an 81% loss of rare alleles, 60% gene allele frequency change, and nucleotide diversity ($\pi$) was almost halved. After domestication, selection toward the elite cultivars resulted in an additional loss of 23% ($\pi$) nucleotide diversity, and a 21% loss of rare alleles (4). Due to the dilution of the cultivated gene pool, half of the resistance-related sequences in *G. soja* are not found in landraces or the domesticated soybean (18). Selection within *G. max* alone has been shown to have a significant effect on diversity across the entire genome (69). SCN is likely native to China and still prevalent throughout the native range of wild soybean (135, 136). Genetic isolation of environmental niches suggests that there is strong selection of environmentally tailored adaptations, such as nematode resistance, within the wild soybean gene pool (79, 137). Novel traits such as pest and disease resistance from wild relatives have been incorporated into major crops (138-147), but progress of this sort is lacking in soybeans (3, 6, 37, 79). *G. soja* has been shown to exhibit native resistance to SCN populations currently inflicting the United States crops (9, 78), and only a few studies have reported candidate genes for SCN resistance in *G. soja* (7, 9, 78). Further research on *G. soja* ecotype, S54, has revealed a mechanism to confer resistance that is not related to the mechanism in *G. max* (7, 47). Since resistance in wild soybean is novel, it can be used

to benefit cultivated soybeans, which are losing effectiveness. SCN resistance is complex, and additional understanding of the mechanism in wild soybean is needed to benefit breeding.

In the present study, we dissect the mechanism of resistance found in wild soybean ecotype, PI578345 (named NRS100100 for nematode resistant *soja*, lab identifier 100). This ecotype is significantly resistant (Female Index = 3.3%) to SCN HG Type 2.5.7 (SCN race 5), which is prevalent in the south-eastern United States. To reveal the novel mechanism of resistance, we compare NRS100 to SCN race 5 resistant cultivar, Peking. A direct comparison between resistance in wild *G. soja* and cultivated *G. max* has not been reported previously. We use RNA-seq transcriptomes of SCN treated and control conditions across four genotypes: Peking (PI548402), Williams 82 (PI509044), S-soja (susceptible *soja*, PI468396B*)*, and NRS100 to determine differential regulation between genotypes. Nematode susceptible and resistant genotypes from both cultivated and wild soybeans are used in this study to eliminate species-specific and stress-induced responses. Novel sources of resistance are found within NRS100 that can be used for applications in the cultivated crop.

## 2. Results

### 2.1 NRS100 is highly resistant to SCN

Our previous study identified a *G. soja* accession S54 showing resistance to SCN race 5 after a large-scale screening of 234 accessions of *G. soja* (78). This screening study was

able to uncover *G. soja* ecotypes highly resistant to SCN race 5. A bar plot shows the resistance responsiveness of all accession to race 5 SCN, indicating the FI difference between NRS100 and S-*soja* (S1 Fig). S-*soja,* originally collected from China, was susceptible with a Female Index (FI) of 149%. NRS100, originating from Russia, was highly resistant with a FI of 3.3%, which was the most resistant of all accessions studied.

### 2.2 *rhg1* and *Rhg4* expression patterns in wild soybean

Analysis of known SCN resistance mechanisms was analyzed to determine if NRS100 uses *rhg1*, *Rhg4*, or SNAP genes to confer resistance. Resistance conferring loci, *rhg1* and *Rhg4* are known to be used in Peking-type SCN resistance (130, 132, 148, 149). The *rhg1* locus differs between Peking-type (*rhg1-a*) and PI88788-type (*rhg1-b*) SCN resistance, but shares the AAT (amino acid transporter) gene (*Glyma.18G022400*, alternative ID: *Glyma18g02580*) and SNAP gene. Peking-type resistance is conferred by only one gene at *rhg1* and *Rhg4* each, SNAP and SHMT, respectively (131, 133). Relative FPKM (Fragments Per Kilobase of transcript per Million mapped reads) of SHMT at *Rhg4*, AAT at *rhg1,* and SNAP at *rhg1* for all four genotypes at control and SCN treatment conditions reveal a different expression profile for Peking at these genes than any of the other genotypes. Peking exhibits high expression of these genes, compared to other genotypes, both at control and treatment conditions, where all other genotypes do not (Figure 8). Expression of *rhg1*-b genes in PI88788-type resistance are not significant in NRS100. A PCA of FPKM at the resistance conferring genes, SHMT and SNAP, confirms that Peking differs from all other genotypes in at the Peking-type resistance conferring genes (Figure 9). In contrast, NRS100 does not exhibit any significant difference in RNA-seq based

expression at *rhg1* or *Rhg4* from the SCN susceptible wild, S-*soja* soybean or cultivar, Williams 82.

Analysis of the transcript sequence of SNAP at *rhg1* and SHMT at *Rhg4* revealed Peking specific variations not found in the other genotypes. In SNAP, Peking has two non-synonymous SNPs, one at 1,634,660 bp on exon 6, resulting in an amino acid change from Aspartic acid to Glutamic acid. The other at 1,645,409 bp on exon 9, resulting in an amino acid change from Threonine to Asparagine (S2 Fig). In addition, a *G. soja* specific synonymous variation exists on at 1,641,767 bp at exon 2. Another synonymous variation

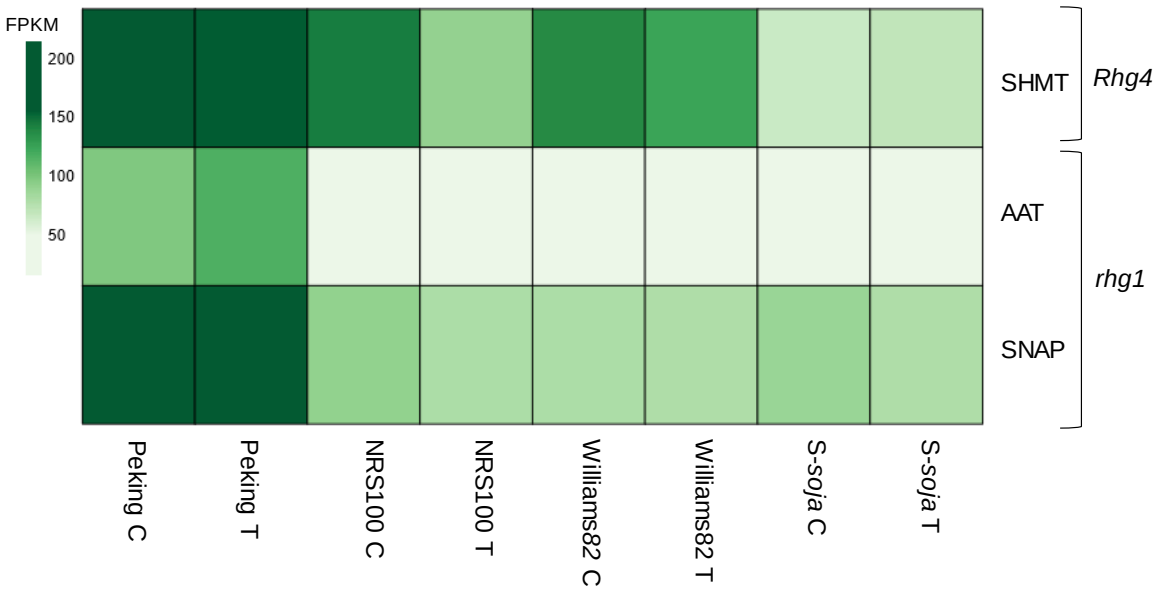



**Figure 8.** Relative expression of *rhg1* genes AAT (*Glyma.18G022400*) and SNAP (*Glyma.18G022500*), and *Rhg4* gene SHMT (*Glyma.08G108900*). No clustering, scaling, or imputed values. Sequential color scheme shows FPKM values.

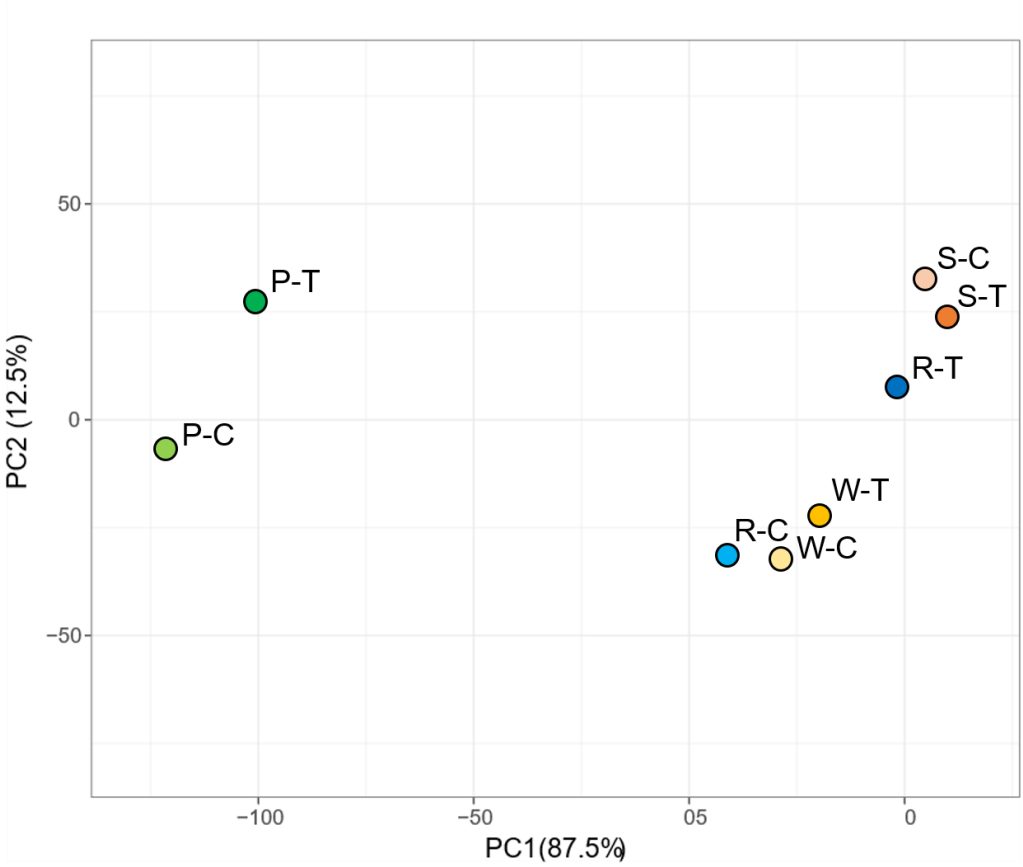



**Figure 9.** Probabilistic principal component analysis (PCA) plot of expression of genes SNAP (Glyma.18G022500) and SHMT (Glyma.08G108900). Individual points represent FPKM of treatment groups P-C (Peking control), P-T (Peking SCN treated), R-C (NRS100 control), R-T (NRS100 SCN treated), W-C (Williams 82 control), W-T (Williams 82 SCN treated), S-C (S-*soja* control), S-T (S-*soja* SCN treated).

exists at 1,641,789 on exon 2, specific to NRS100 (S2 Figure). The only non-synonymous variations in SNAP are found in Peking, suggesting resistant-type SNAP allele is only found in Peking, while NRS100 has a similar susceptible-type SNAP gene similar to Williams 82. A similar trend was seen in the SHMT transcript sequence. A synonymous SNP was found in NRS100 at 8,361,269 bp (S3 Figure). A non-synonymous variation was found in Peking at 8,361,924 bp, resulting in an amino acid change from Asparagine to Tyrosine (S3 Figure), as well as distinct variation in mapping from the reference (Williams

82) in this area, which was not found in NRS100 of S-*soja*. All sequence variations that resulted in amino acid changes were found only in Peking-type SNAP and SHMT sequences, further suggesting that Peking and NRS100-type resistance are not the same.

The resistance conferring gene of *rhg1-a* in Peking-type resistance is a SNAP gene at chromosome 18. Paralogous SNAP genes were also analyzed to determine if resistance in NRS100 was conferred by another member of the SNAP gene family. No significant expression or fold change is found on SNAP 2, 9, 11, 14 by RT-qPCR or RNA-sequencing, as if found in resistant soybeans using SNAP for resistance (150). Fold change <0.4 for all NRS100 SNAP genes (S4 Figure). It was determined that NRS100 does not have significant differential expression at SHMT, SNAP 18 or a SNAP paralog to maintain SCN resistance.

### 2.3 Previously identified SCN resistance in wild soybean does not explain NRS100 resistance

Previously, SCN resistance related QTLs were mapped for *G. soja* accession PI468916, and fine mapping revealed two candidate loci, *cqSCN-006* on chromosome 15 and *cqSCN-007* on chromosome 18 (86, 87, 151). None of the candidate genes on previously described resistance locus *cqSCN-006* in SCN race 5 resistant *G. soja* PI468916 were found significantly upregulated in NRS100, including the *y*-SNAP gene (*Glyma.15g191200*). However, one candidate gene in in resistance locus *cqSCN-007* is significantly upregulated in NRS100 exclusively, Apetala 2 (AP2) transcription factor (*Glyma.18g244600*) (84). AP2 is known to regulate multiple developmental pathways and play a role in abiotic and biotic stress response (152).

A previous RNA-sequencing study of SCN race 5 resistant *G. soja*, S54, revealed potential pathways of resistance in this genotype (7, 47). Expression patterns of S54 and NRS100 were compared to determine if the mechanism of resistance was the same. Similarities between these genotypes were limited, and are noted. Previously identified Leucine-rich repeat receptor-like protein kinases (LRR-RLKs) DEGs in *G. soja* SCN response, Brassinosteroid insensitive 1 (BRI1)-associated kinases (BAK1) (*Glyma.05G119500*, *Glyma.05G119600*) genes and BAK-1interacting receptor kinase 1 (SOB1R1) genes *(Glyma.04G190400, Glyma.06G175100)* are induced in response to SCN, but not exclusively to NRS100 (47). Previously studied (chitin elicitor receptor kinase 1) CERK1 genes (*Glyma.02G270700, Glyma.15G111300*) showed opposite reactions, significantly down and upregulated respectively, just like in S54 SCN race 5 response. The response was specific to NRS100. Two lectin receptor kinase (*Glyma.07G135400*) was also exclusively induced in NRS100, which was also found in previously studied S54 (47). In addition, significant downregulation of NBS-LRR gene (*Glyma.16G209000*) and upregulation of (*Glyma.17G180000*) was found exclusively in NRS100, just like in S54. Significant upregulated DEG's associated with calmodulin binding, *Glyma.05G237200*, *Glyma.07G093900*, *Glyma.08G044400*, were exclusively upregulated in NRS100 and S54. Calcium transport genes, autoinhibited $Ca^{2+}$-ATPase 9 (ACA9, Glyma.07G004300) and protein calcium exchanger 7 (CAX7, Glyma.19G066700) were exclusively upregulated in both NRS100 and S54. MAPKKKs *Glyma.17G245300* and *Glyma.05G094400* were significantly up and down regulated in both S54 and NRS100 exclusively, however, the strongest DEG in the MAPK cascade found in S54 (*Glyma.15G048500*) was not found to exhibit the same response in NRS100. Jasmonic acid (JA) pathway gene,

oxophytodienoate-reductase 3 (DDE1/OPR3, *Glyma.13G109800)* showed a similar induced effect in S54 and NRS100 exclusively, however this was the only one out of the JA pathway genes to show this similarity between the two ecotypes. Highly induced WRKY40 transcription factors shared between NRS100 and S54 are *Glyma.13G370100* and *Glyma.17G222500,* although the fold change is not as high in NRS100 in either of them. No similarities were found in upregulated chitinase genes in S54, however, the downregulated Chitinase-like protein (CLT) genes *Glyma.09G038500* and *Glyma.15G143600* had similar expression in NRS100 in response to SCN infection (47). Overall, similarities between expression in NRS100 and S54 are limited, suggesting a different mechanism of resistance.

### 2.4 Induced expression in response to SCN infection differs in NRS100 from all other genotypes

Previously identified mechanisms of SCN resistance could not explain the resistance mechanism used in NRS100. Therefore, RNA-sequencing expression results from the entire genome of NRS100 in comparison with other genotypes were considered in order to determine the mechanism used in NRS100. Illumina sequencing generated 17.9-29.7 million raw reads per library, in 24 libraries total. Filtering by quality of reads resulted in 93.8-95% usable reads for alignment. The mapping rate of quality-controlled reads ranged from 84.9-92.8% of all 24 libraries to *G. max* Wm82.a2.v1 (S1 Table). On average 56, 710 expressed genes were found in root tissue from each genotype. The amount of significantly ($q < 0.01$) differentially expressed genes between control and treatment groups for each genotype were 4,641, 2,911, 498, and 663 in NRS100, S-*soja*, Peking, and Williams 82 respectively.

In order to determine the genes associated with induced resistance response to SCN in NRS100, DEGs of four genotypes between control and SCN treatment were compared. A count of the significant DEGs of all four genotypes, NRS100, S-*soja*, Peking, Williams 82, in response to SCN infection revels the higher quantity of expressed genes in NRS100 compared to other genotypes, with 1602 and 3039 genes significantly up and down regulated respectively (Figure 10). The SCN susceptible wild soybean (S-*soja*) has the second most substantial response to SCN infection. Both cultivars, Peking and William 82, exhibit less extensive and more tailored response to SCN infection, with Peking exhibiting a lower number of down regulated genes. This trend is further supported by the expression profiles of highly-significantly expressed genes, with a cutoff $Log_2FC$ no less than +/-2 for NRS100 (Figure 11).

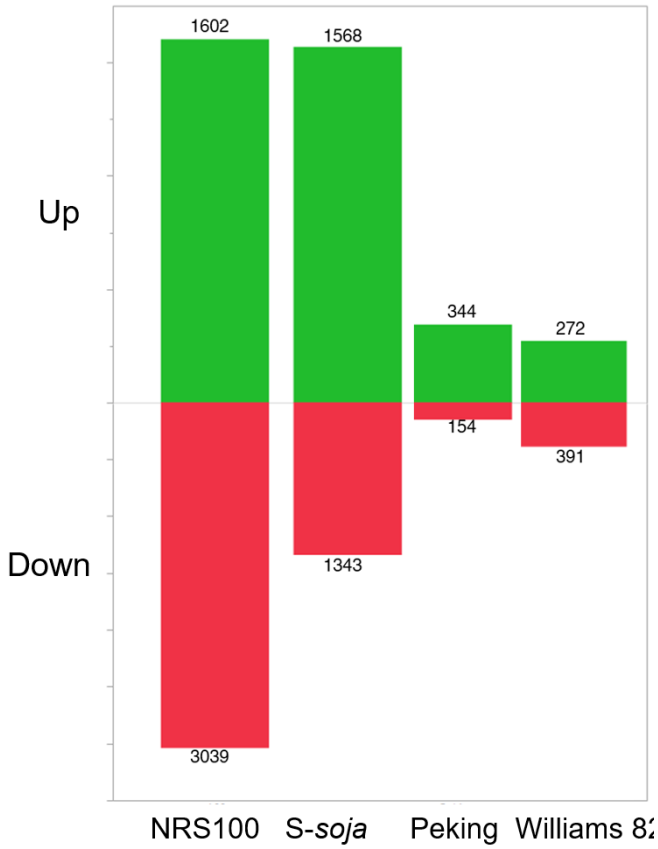



**Figure 10**. Significantly up and down regulated genes by genotype. DEGs considered significant if q-value<0.01, $log_2$foldchange no less than +/-0.6.

NRS100 is seen to have a different set of genes induced in response to SCN infection, (green and red bars (Figure 11), when compared to other genotypes. By contrast, the induced response to SCN infection in Peking (blue and orange bars (Figure 11), is controlled by smaller set of genes, separate from those induced in NRS100. Shared responses by all genotypes, and species-specific shared responses are also seen, indicating a shared stress-response mechanism. GO analysis of resistance specific, shared response, and stress response genes was done after visualization of all DEG's in a venn-diagram (153) (S5 Figure).

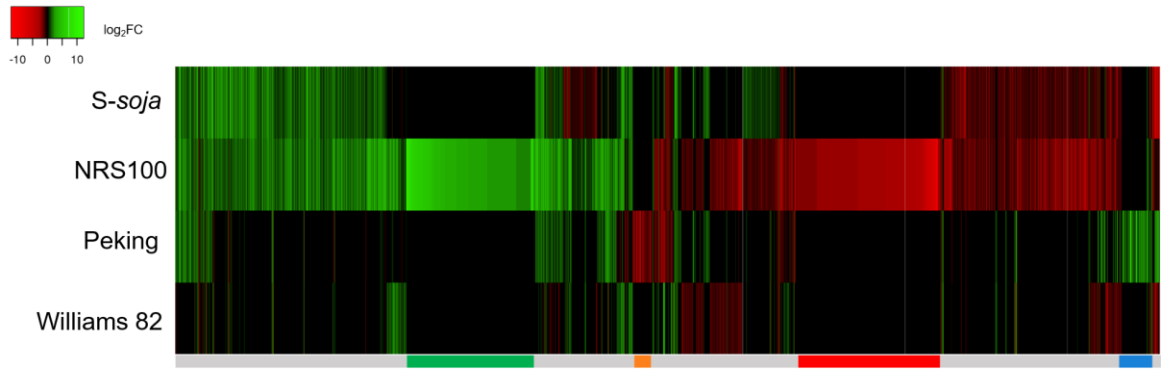



**Figure 11.** Heatmap of all DEGs by $log_2FC$ (fold change) of induced response to SCN treatment of all genotypes clustered by. Centroid clustering on genes. Cutoff $Log_2FC$ no less than +/-2 for NRS, including 2076 genes. Green bar indicates NRS100 specific up regulated genes in response to SCN stress. Red bar indicates NRS100 specific down regulated genes in response to SCN stress. Blue bar indicates Peking specific up regulated genes in response to SCN stress. Orange bar indicates Peking specific down regulated genes in response to SCN stress.

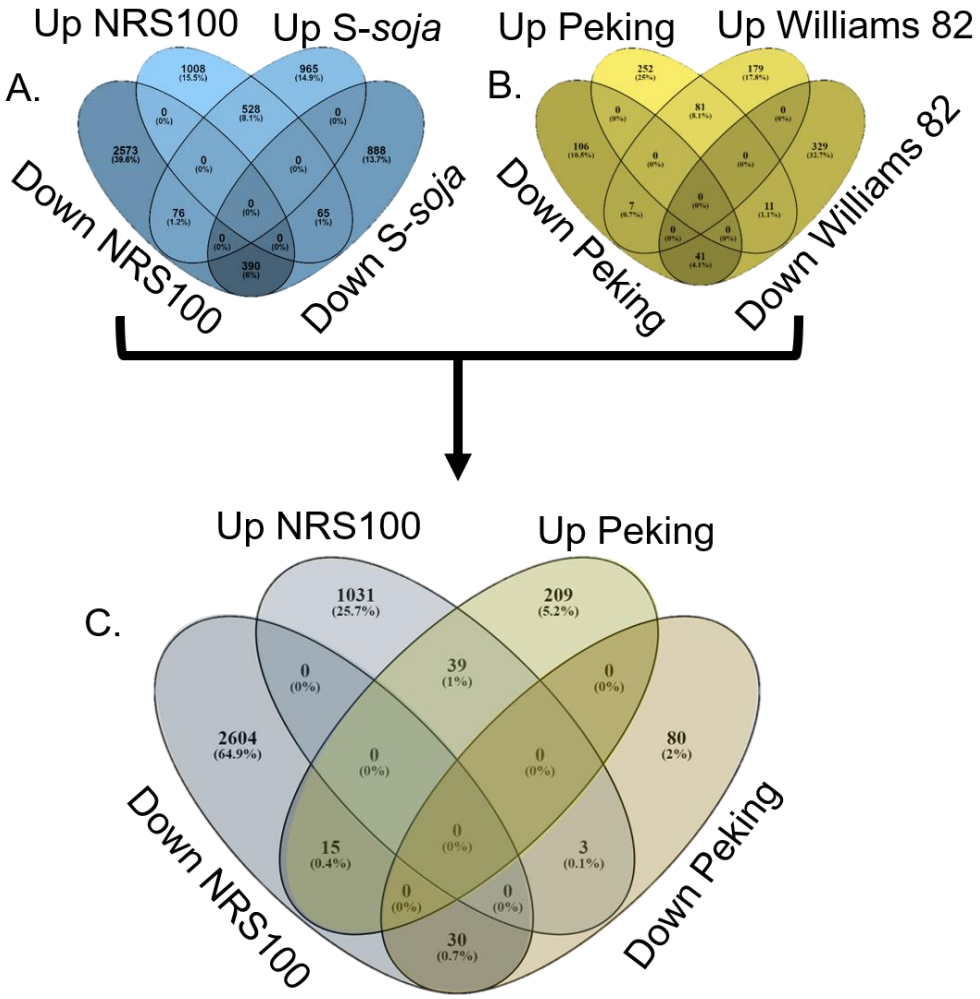



**Figure 12.** A. Comparison of DEGs for SCN susceptible and SCN resistant *G. soja* genotypes. B. Comparison of DEGs for SCN susceptible and SCN resistant *G. max* genotypes. Shared responses from susceptible and resistant genotypes from the same species were eliminated to produce a comparison of resistance-specific associated genes. C. Comparison of resistance-specific genes between Peking and NRS.

To elucidate the expression profile associated with resistance and eliminate generalized stress responses, species-specific induced responses were compared (Figure 12A, 12B). Shared induced responses by genotype were eliminated to create a subset of resistance-specific induced DEGs (Figure 12C). Again, resistance-specific expression in NRS100 is at a much larger scale than in Peking, consisting of expression profile more than ten-fold

to Peking. Comparison of resistance-specific expression profiles of Peking and NRS100 reveal a shared response consisting of 69 genes, contributing to only 1.8% of the overall resistance-specific profile of NRS100. Gene ontology (GO) functions associated with NRS100 resistance include up regulation of systemic acquired resistance mechanisms, salicylic acid signaling pathway, and regulation of defense response mechanisms (S5 Figure, S2 Table, Table, S3 Table) and down regulation of secondary cell wall biogenesis and light harvesting by photosynthesis (S4 Table, S3 Table).

### 2.5 Constitutive SCN resistance in NRS100

To further analyze the potential mechanism of SCN resistance in NRS100, the constitutive expression profile specific to NRS100 was considered due to its important role in plant-pathogen protection (154-158). To determine the constitutively expressed genes associated with SCN resistance in NRS100, the basal regulation of NRS100 was compared to S-*soja* and Peking at control conditions. This eliminates specific-specific regulation shared with wild soybeans and resistance-specific regulation shared with cultivated crops. The significantly induced expression profile from SCN infection in NRS100 was also eliminated from the basally regulated genes. This method ensures that only the genes highly upregulated in the control condition in NRS100 specifically, that also did not significantly alter their expression after SCN infection, were captured. Comparison of all expression profiles results in a set of 1656 genes that are constitutively expressed in NRS100, specific to resistance to SCN (S6 Fig). The overrepresented GO functions of these genes include pathways associated with photoprotection, defense response, and respiratory burst defense response, indicative presence of polyamine derived hydrogen peroxide (S7

Figure) (159, 160). Analysis of combined induced and constitutive NRS100 specific upregulation revealed the importance of glutamine metabolic process (GO:0006541), glutathione metabolic process (GO:0006749), and chitin binding (GO:0008061) molecular functions (S8 Figure). Associated biological processes include defense and stress response processes (GO:0006950, GO:0006952, GO:0006979) and response to biotic stimulus (GO:0009607, GO:0043207) (S9 Figure). The cellular component of these processes were allocated to areas including the cell periphery (GO:0071944), plasma membrane (GO:0005886), extracellular region (GO:0005576), and cell wall (GO:0009505) (S10 Figure).

**2.6 Expression patterns in NRS100-specific SCN resistance pathways**

Based on the above findings, putative metabolic pathways conferring resistance were selected based on exclusive constitutive and induced responses, importance in GO enrichment, and previous understanding of plant-pest interactions. The Jasmonic acid (JA) signaling pathway is a well-studied pathway, associated with plant-pathogen interactions, and known to have antagonistic effects on the salicylic acid (SA) signaling pathway, and therefore was analyzed further to determine if signal transduction plays a role in SCN resistance in NRS100 (161-165). Within the JA signaling pathway, NRS100 specific downregulation occurred at JAR1 (jasmonate resistant 1) genes and COI1 (coronatine-insensitive protein 1) gene. JAR1 genes *Glyma.03G256200* and *Glyma.16G026900* have NRS100 exclusive induced down regulation in response to SCN. Downstream of JAR1, COI1 gene, *Glyma.18G030200*, was downregulated in NRS100 exclusively in response to nematodes. No other genotypes had a significant induced response at these genes.

Regulation at known genes in JAZ (jasmonate ZIM domain-containing protein) and MYC2 transcription factor (*Glyma.08G271900*) were not induced in NRS100 but were highly expressed in control conditions, or constitutively expressed significantly higher than other genotypes (S11 Figure). Expression at these significant genes in the JA signaling pathway were confirmed in the RT-qPCR results with representatives for JAR1 and JAZ due to modulation of these proteins by multiple genes (S5 Table).

The polyamine biosynthesis pathway, specifically at spermidine synthesis, was analyzed in further detail to determine if NRS100 uses a polyamine related defense response known to play a role in resistance to other plant pathogens (166, 167). Upstream of spermidine synthase, ODC1 (ornithine decarboxylase, Gene ID: *Glyma.04G020200)* is significantly upregulated in response to nematode infection in both Peking and NRS100. There are two paralogous spermidine synthase genes, SPDS17 (*Glyma.17G091123*) and SPDS5 (*Glyma.05G036300*). SPDS5 is significantly upregulated in both NRS100 and Peking. However, SPDS17 is significantly upregulated in response to nematode infection in NRS100 only. These expression results are confirmed by both RNA-sequencing and RT-qPCR results. Downstream of spermidine synthesis, expression of PR-1 (pathogenesis-related protein 1) genes is highly constitutively expressed in NRS100 at *Glyma.15G062300*, and non-specifically significantly induced at *Glyma.15G062400* and *Glyma.15G062500*. All three PR-1 genes are significantly induced in response to SCN infection in S-*soja*, but not in either cultivar. Species-specific response at PR-1 is found, with NRS100 specific constitutive expression of *Glyma.15G062300* and overall higher level of expression when induced in NRS100. The pattern of expression visualization was

adapted from Pathview created figures (S11 Figure). The polyamine related defense response includes direct and indirect resistance mechanisms, where indirect involves the production of an unknown secondary metabolite.

### 2.7 Validation of expression by RT-qPCR

RNA-sequencing expression of 20 genes (S5 Table), including randomly selected and selected from important pathways detailed in pathview results (S11 Figure), was validated using RT-qPCR for all genotypes and condition replicates. RT-qPCR results of the 20 selected genes correlated with RNA-sequencing results. An expected indirect correlation (R=0.621) of RT-qPCR derived ΔCT to RNA-sequencing derived FPKM of all biological replicates and averaged technical replicates was found (Figure 13). In addition, calculated

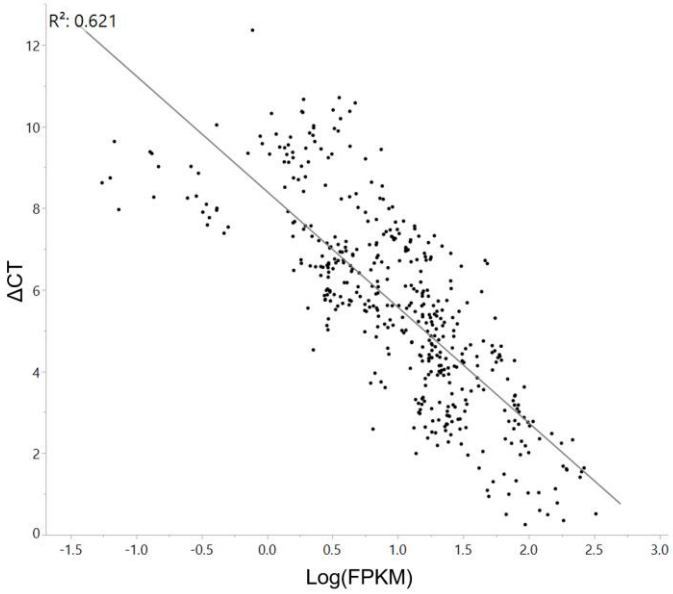



**Figure 13.** Correlation ($R^2$=0.621) between qPCR expression values (ΔCT) and correlated RNA-sequencing expression values (FPKM) for 20 genes (supplementary table 6). Individual points represent relative expression of individual replicates calculated by RT-qPCR analysis $2^{-\Delta\Delta CT}$ method and calculated FPKM expression of individual replicates from RNA-sequencing analysis.

fold change for from RT-qPCR correlates directly (R=0.762) with RNA-sequencing derived fold change (S12 Figure). Results confirm that assumptions made from RNA-sequencing results are valid. The selected genes were used not only to validate RNA-sequencing results but to better understand the expression of genes of interest, including those within *rhg1*, *Rhg4*, SNAP paralogs, the Jasmonic Acid (JA) signaling pathway, and the Polyamine defense response.

## 3. Discussion

SCN is the most destructive pest to soybean, and is rapidly evolving increased virulence against resistant crops (117, 119, 123, 168). Current efforts to understand SCN resistance is focused on PI88788 and Peking type resistance, which use *rhg1-b* or a combination of *rhg1-a* and *Rhg4* (34, 129-133, 169). Little effort has been placed on sources of resistance outside *rhg1* or *Rhg4*, but is necessary to combat the future SCN population. Due to evidence of selection at resistant *rhg1* in wild soybean, and emergence of resistant *Rhg4* allele after domestication it is likely that the wild soybean population has a separately derived method of SCN resistance (32, 134). However, research to understand the mechanisms of resistance in wild G. *soja* is lacking, and only a handful of studies describe potential mechanisms (7, 47, 78, 83, 86-88, 151).

### 3.1 NRS100 does not use *rhg1*, *Rhg4*, or SNAP family to confer resistance

Our results show that NRS100 uses a resistance mechanism independent of Peking-type or PI88788-type resistance. Genes at *rhg1-b*, *rhg1-a*, and *Rhg4*, are not induced in response to SCN infection or highly expressed at basal levels in NRS100, and the overall expression

pattern at resistance conferring genes in Peking is no different in NRS100 than the susceptible genotypes (Figure 8, Figure 9). Of the two main cultivated sources of resistance, Peking-type resistance is the only one that is resistant to SCN race 5, which uses SNAP at *rhg1-a* and SHMT at *Rhg4* to confer resistance(119, 131). In addition, analysis of the molecular sequence of exons in SNAP and SHMT reveal that non-synonymous mutations found in Peking are not found in any other genotype in our study, including NRS100. This confirms that the Peking-type variation of these resistance conferring genes are not found in NRS100. Resistance to SCN is conferred by a complex mechanism in the cultivated soybean, that consists of copy number variation of causative genes at the *rhg1* and *Rhg4* loci, and the genetic composition of these genes and associated promoters (34, 129-133). Our results suggest NRS100 does not make use of *rhg1* and *Rhg4* based on their constitutive and induced transcription, and the molecular sequence of the coding region, but the copy number of these genes and the composition of their promotors is yet to be determined.

It has been shown that the product SNAP in *rhg1* likely associates with proteins within the SCN secretion at the point of infection. SNAP gene products interact with SCN protein products of Hg-SLP-1, and potentially HgBioB, to confer resistance (169-171). Due to the importance of SNAP in SCN resistance, the SNAP family was analyzed further to determine if another member conferred resistance by a similar mechanism. Previous characterization of the SNAP family revealed that SNAP 11 contributes to additive resistance to SCN in Peking-type resistance, however our results suggest that none of the SNAP genes in the family (SNAP2, SNAP9, SNAP11, SNAP14, and r*hg1* SNAP18) are

significant in SCN resistance in NRS100. In addition, NSF gene, *Glyma.07g195900,* previously shown to be associated with resistant SNAP at *rhg1,* is not significantly induced or highly constitutively expressed in NRS100 (150, 169). Thus, suggesting that a novel mechanism outside of the previously identified *rhg1*, *Rhg4*, and SNAP family conferred resistance exists in NRS100.

### 3.2 Mechanisms of SCN resistance in wild soybean

The only gene found in common with *G. soja* PI468916 resistance by *cqSCN-006* and *cqSCN-007,* was an AP2 transcription factor. This transcription factor is related to flower development in Arabidopsis (*Arabidopsis thaliana)* and stress response (151, 152). AP2 transcription factors can be activated by either JA or SA pathways, which may indicate that signaling pathways are important in resistance in wild soybeans (172-175).

Similarities between NRS100 and previously reported important genes in S54 resistance are limited to RLK-LRRs, Non-LRR domain RLKs, calmodulin binding genes, calcium transport genes, WRKY40, a JA pathway associated gene, MAPKKKs, and downregulated CLTs. Although similarities are seen between NRS100 and S54, the expression profile found in NRS100 does not include the most important genes in the S54 proposed resistance mechanism (47). In addition, none of the candidate genes found in S54 GWAS correlate with DEG's of NRS100, therefore it is likely that these do not use the same method of resistance to SCN race 5 (78).

### 3.3 Novel mechanism of resistance in NRS100

NRS100 displays a novel expression profile before and after SCN inoculation, indicating a new mechanism of resistance. The amount of DEGs in response to SCN is much greater in NRS100, with twice as many DEGs than *S-soja* and almost 10-fold the amount found in Peking or Williams 82. This strong response by NRS100 suggests that resistance is controlled by many genes, and is a less tailored response than that found in the artificially selected Peking-type resistance. The number of genes downregulated in NRS100 exclusively was more than twice that upregulated, demonstrating a significant trade off, consisting of downregulation of energy and nutrient uptake. NRS100 and Peking have a limited shared resistance specific response to SCN, consisting of 39 up and 30 downregulated genes, an insignificant amount compared to the 1031 up and 2604 downregulated genes specific to NRS100. There is a clear induced signal in NRS100 that is non-existent in the SCN resistant cultivar, Peking, or either susceptible genotype, which has not been directly demonstrated in any other *G. soja* ecotype.

In addition to a clear induced effect, constitutively expressed genes, or genes that are highly upregulated in control conditions of NRS100 compared to the other genotypes, suggest a resistance mechanism that is already turned on before SCN infect the roots. Constitutively expressed genes are known to play a significant role in plant-pathogen resistance, which can promote toxin production, increase the sensitivity to pathogen stimuli, or support an inducible immunity (154-158). Intracellular signal transduction is highly upregulated in control NRS100 roots, suggesting a higher sensitivity to pathogen trigger via signal

transducers. Ethylene biosynthesis, another signaling molecule, is also highly expressed, which is associated with induced resistance to root knot nematode and SCN, but conversely also found to increase susceptibility to SCN in a separate study (176-181) . To explain the unique resistance found in NRS100, we hypothesize two modes of resistance in NRS100, 1) induced immunity by suppression of JA signaling pathway, and 2) spermidine synthesis mediated induced defense against SCN growth.

### 3.4 JA suppression in induced immunity

Signaling pathways are widely studied for their role in parasite and pathogen defense of plants. Signaling pathways, including JA, SA, and ethylene act as detection system to decipher pest and pathogen signals in order to allow a plant to mount a defense (165, 182). High levels of jasmonate are associated with increase wound signaling to insect pests by inducing an effective defense against pests, and induced systemic defense in root knot nematode by PR1 (162, 180, 183, 184). However, JA signaling is antagonistic with SA signaling (182, 185). It has been shown in tomato (*Solanum lycopersicum)* that JA signaling actually has a negative impact on resistance to aphids and root knot nematode, and JA-deficient mutants exhibit elevated resistance (186, 187). In the interaction with aphids, susceptibility can be restored by suppressing the SA signaling pathway, demonstrating the direct interplay between the JA and SA signaling pathways (185). A separate example exists in rice where modification of JA signaling increases SA levels and in turn leads to increased resistance to multiple insect herbivory (188). It is likely that JA is important for broad resistance and wounding response, but can be costly due to its negative interaction with SA signaling which is needed for constitutive and inducible

resistance mechanisms (164, 189-191). In addition, suppression of the JA signaling pathway can promote plant growth via activation of PIF regulated gibberellin signaling (192). It is also possible that JA signaling is suppressed by a deactivation of NRS100 defense mechanisms by a decoy produced by the nematode, similar to that found in the tobacco-*Helicoverpa zea* larvae interaction for example (164, 193). However, this is unlikely considering this suppression is only found in NRS100 and not the SCN susceptible genotypes. With suppression of JA, SA can accumulate. SA accumulation has been found to be important in specialized defense response in Arabidopsis (194).

In NRS100, JA signaling is suppressed at JAR1 and COI1, important in regulators in the pathway (184, 195, 196). This suppression reduces the ability of COI1 and JAZ to interact physically, and does not allow a release of JA mediated transcription factors (184). Upregulation of JAZ and MYC2 act as repressors to the JA pathway (197, 198). Suppression of JA allows SA signaling to induce a defense response, avoiding potential conflicts with SCN produced decoys to mediate traditional JA signaling in SCN defense response (Figure 14). Interestingly, the downregulation of JAR1 and COI1 is induced in NRS100 only, but the induced response at JAZ is found in both Peking and NRS100 but regulated by separate genes. This suggests that JA signaling is not only regulated differently at key steps (COI1 and JAR1) in NRS100 but is regulated by different genes at points of similar expression of with Peking (JAZ). It is likely that many different JAZ products exist, especially in the case of regulation by separate genes, which is expected to occur in NRS100 (164, 191).

### 3.5 Spermidine synthesis mediated induced defense against SCN growth

We hypothesize that spermidine synthesis and secondary metabolite production induce defense against SCN growth and syncytium formation in NRS100. Polyamine metabolism, including spermidine synthesis, is often altered in abiotic and biotic stress responses in plants (199, 200). Spermidine synthesis upregulation can result in accumulation of free spermidine or lead to the production of polyamine derived molecules. This can trigger $H_20_2$ regulated hypersensitive response, cooperative effects with SA signaling to induce PR proteins or other modes of pathogen resistance, lead to production of secondary metabolites like nicotine, and reinforce/maintain structural integrity of cell walls to hinder pathogen invasion (159, 160, 166, 167, 201-205). In addition, higher expression of spermine

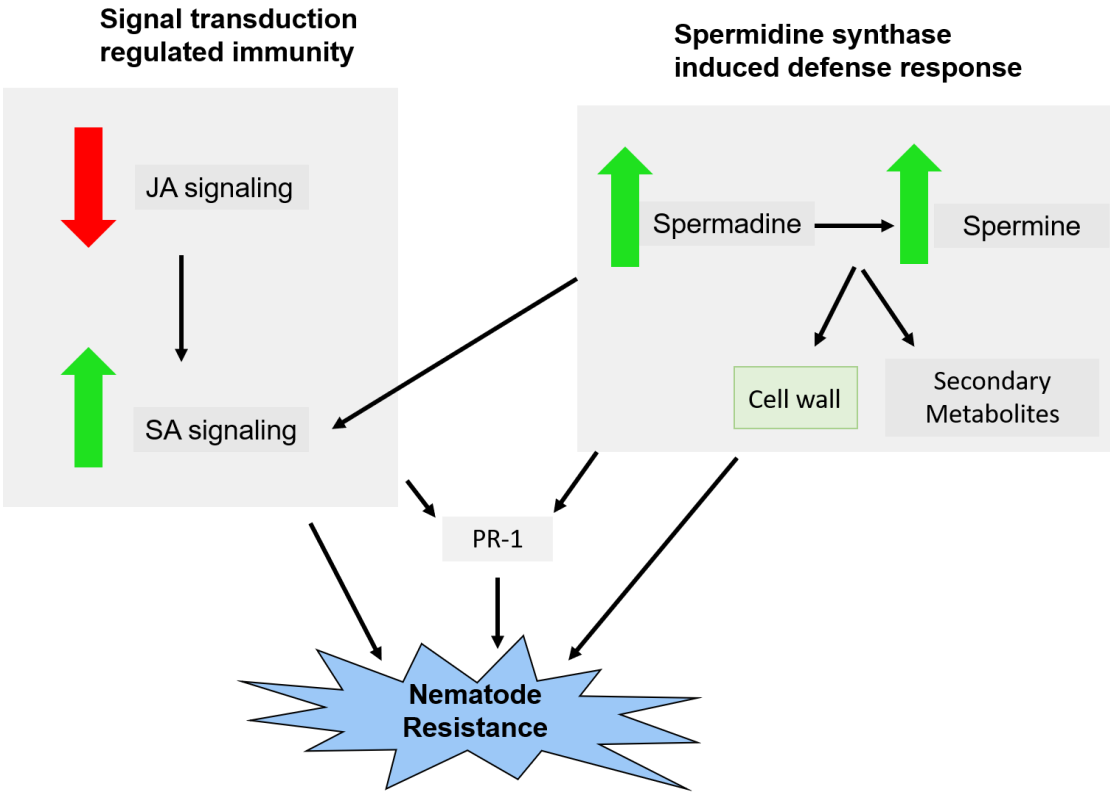



**Figure 14.** Pathway of resistance in NRS100 includes upregulation of SA signaling by JA suppression, induced PR-1 genes, accumulation of polyamines (spermidine and spermine), reinforcement of the cell wall, and production of secondary metabolites.

synthesis, directly following spermidine synthesis, has been shown to induce pathogen resistance (160, 166). NRS100 exhibits non-exclusive upregulation of spermine synthesis, downstream of exclusively upregulated SPDS. Therefore, accumulation of spermine and spermidine in response to SCN is likely. Conversely, it has been found that upregulation of SPDS can be induced by effector proteins secreted by nematodes, and therefore leads to a compatible plant-pathogen interaction (206). Spermidine synthesis gene, SPDS17, is found to be significantly upregulated in NRS100, a response that is specific to the genotype. Interestingly, a separate SPDS gene on chromosome 15 is also highly upregulated in both Peking and NRS100, but not to the same extent as SPDS17 is in NRS100. In addition, upregulation of precursors, at putrescine synthesis (ODC1) is upregulated in both NRS100 and Peking but at a much higher level in NRS100. Overexpression of ODC1 has been shown to lead to elevated levels of putrescine but not spermidine, meaning that the upregulation of spermidine synthesis in NRS100 activates an independent response and production of spermidine, regardless of the shared ODC1 upregulation with Peking (207, 208).

Although polyamine synthesis can be related to both compatible and incompatible interactions with pathogens, we propose that SPDS17 regulated spermidine synthesis in NRS100 has a highly incompatible interaction with SCN (199). The outcome impedes the nematodes' ability to expand a feeding site, leading to decreased nutrition and death.

## 4. Conclusion

These results identify an expression profile unique to wild soybean SCN resistance and identify candidate mechanisms significantly associated with SCN resistance by dissecting the expression profile of a newly found SCN resistant ecotype, NRS100. The method makes use of both wild and cultivated soybeans in order to identify novel resistance mechanisms. The identification of candidate pathways involved in SCN resistance will advance the long-term goal to develop SCN resistant soybean cultivars, which has crucial significance to agriculture and environmental sustainability. Novel genes involved in SCN resistance can be incorporated into the cultivated crop to advance crop improvement methods, and provide a model for further improvement using crop wild relatives.

## 5. Materials and Methods

### 5.1 Screening for resistance

SCN race 5 were reared on soybean cv. Williams 82 plants in the greenhouse under optimal conditions at 27°C and 16h light for over 30 generations. Female SCN cysts were harvested from the roots using a nested sieve collection (850 and 250 μm) method, and females released from cysts by pressure into a 25 μm sieve (47). Released eggs were purified by sucrose flotation (209) and allowed to hatch on incubation trays over water.

For screening, second stage juveniles (J2) were collected from the water and diluted to a concentration of 2500 eggs/ml for inoculation. A set of 234 previously screened wild

soybean accessions from USDA Soybean Germplasm collection were inoculated along with indicator lines (Peking, PI88788, PI90763, PI437654, PI209332, PI89772, and PI548316) to verify HG Type 2.5.7 After 35 days, cysts were collected from roots and counted under a stereomicroscope to determine resistance level by FI (Female Index) calculation as a percentage *(FI=(number of females on given individual/average number of females on susceptible control)/100)* (78, 210) Significantly resistant ecotypes from the set of 234 previously screened wild soybean accessions were revealed (S1 Figure) (7, 78). The accessions natural ranges include China, Russia, South Korea and Japan. *G. soja* ecotype, NRS100, was found to be significantly resistant to SCN race 5 with a FI of 3.3%. A bar plot is used to visualize of the resistance level of the 234 genotypes studied, indicating NRS100 and S-*soja* specifically (JMP® Pro Version 13. SAS Institute Inc., Cary, NC, 1989-2019).

**5.2 SCN stress experiment**

A controlled SCN stress experiment was performed on SCN-resistant wild soybean accession (NRS100), SCN-susceptible wild soybean accession (S-*soja*), SCN-resistant cultivated soybean accession (Peking), and SCN-susceptible cultivated soybean accession (Williams 82). Germinated soybeans were planted in cone planters (Stuewe & Sons, Tangent, Oregon, USA) in replicates of 12 per condition, treatment and control for each accession in sterilized sand with nutrients. After 2 days post transplanting, each plant was inoculated with hatched J2 of SCN race 5 (1500/plant) suspended in 0.09% agarose, or a blank (1 ml 0.09% agarose) for control groups. Optimal growing conditions were kept constant at 27°C and 15h light at 50% relative humidity in an environmental chamber.

Roots were collected 6 days after SCN inoculation and fragmented in liquid nitrogen in preparation for RNA isolation. Each biological replicate contained pooled roots from three individual plants, providing four replicates of each group in total.

### 5.3 RNA sequencing and library construction

Root tissues was pooled by replicates for total RNA extraction. Total RNA extraction was preformed using the RNeasy mini total RNA isolation kit (Qiagen, Valencia, CA, USA), and RNA integrity, purity, and concentrations were assessed using an Agilent 2100 Bioanalyzer with an RNA 6000 Nano Chip (Agilent Technologies, Palo Alto, CA, USA). Messenger RNA (mRNA) was purified from the total RNA with oligo-dT beads provided in the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England BioLabs, Beverly, MA, USA). Complementary DNA (cDNA) libraries for Illumina sequencing were constructed using the NEBNext Ultra Directional RNA Library Prep Kit (NEB, Beverly, MA, USA) and NEBNext Multiplex Oligos for Illumina (NEB, Beverly, MA, USA) using the manufacturer-specified protocol. The mRNA was chemically fragmented and primed with random oligos for first-strand cDNA synthesis. The double-stranded cDNA was then purified, end repaired and "a-tailed" for adapter ligation. Following ligation, the samples were selected and sample-specific indexed. The final quantified libraries were pooled in equimolar amounts for sequencing on an Illumina HiSeq. 2500 utilizing a 125-bp read length with v4 sequencing chemistry (Illumina, San Diego, CA, USA).

Technical and biological replicates of total RNA sequences of controlled and treatment group root tissues were obtained using the Illumina 1.9 Hi-seq platform (NEB, Beverly,

MA, USA). Each replicate was on two separate lanes and merged, generating 24 total libraries in triplicates per sample group (control 1-3, treatment 1-3 per genotype). Raw fastq reads were checked for quality control with FastQC (version 0.11.6) and filtered for quality, and trimmed (quality score <30, and adapters) using FASTx toolkit (version 0.0.13). Reads were mapped against reference genome *G. max* Wm82.a2.v1(211) with Tophat (version 2.1.1) and Bowtie2 (version 2.2.9) for all four genotypes, control and treatment. Assembly of transcripts of each replicate to the reference was done using Cufflinks (Version 2.2.1), a final assembly of each replicate was obtained, and differently expressed genes (DEGs) as the final step in the Cufflinks pipeline (212-215).

**5.4 Comparative transcriptomics**

In order to investigate the global gene expression changes, we compare RNA seq-based transcriptome of the NRS100, S-*soja*, Peking, and Williams 82, under both control and nematode-treated conditions. In order to determine induced response, differentially expressed profiles were found between control and treatment groups for each genotype. In order to determine constitutive expression, differentially expression profiles were compared between control conditions of each genotype.

Resulting transcripts from mapping, assembly by Tophat (version 2.1.1) and Bowtie2 (version 2.2.9) were analyzed for differential expression using Cufflinks (Version 2.2.1). Genes with FDR significance ($q < 0.01$) were considered Differentially Expressed Genes DEGs. Putative metabolic pathways assigned for SCN-resistant genes with the KEGG pathway database, SoyBase Gene Model Data Mining and Analysis Tool (94, 216),

(http://www.soybase.org) PlantRegMap (217), and visualization of pathways in Pathview 1.22.3 (218, 219). Integrative Genomics Viewer 2.5.3 (220) was used for sequence analysis against susceptible type Williams 82.

### 5.5 Quantitative reverse transcription PCR (RT-qPCR)

Intron-spanning primers designed to bracket 80-150 bp of 20 selected genes (S5 Table). RT-qPCR performed on an ABI 7500 Fast real-time PCR system (Applied Biosystems, Foster City, CA, USA) using PerfeCTa™ SYBR® Green FastMix™ (Quanta Biosciences, Gaithersberg, MD, USA) with biological and technical triplicates. Relative expression quantified by comparison of ΔCT and FPKM from RNA sequencing results, in addition to the $2^{-}$ ΔΔCT method (221).

# CHAPTER 4: COMBINATION STRESS PROFILING IN SOYBEAN CYST NEMATODE RESISTANT WILD SOYBEAN

## 1. Introduction

The wild soybean, *Glycine soja* (Siebold & Zucc.), is the ancestor of the cultivated soybean, *Glycine max* (*Glycine max* (L.) Merr.). Soybean is the sixth most produced crop, and the leading legume crop in production and economic importance (222). The largest threat to this staple crop is soybean cyst nematode (SCN, *Heterodera glycines*), resulting in $1.2 billion in yield loss in the United States annually (223). SCN is a rapidly growing pest and has the demonstrated ability to adapt and overcome all resistance sources in the cultivar gene pool (121, 122). Therefore, it is necessary to refresh the cultivated gene pool with the use of its crop wild relative. *G. soja* is more genetically diverse than its crop descendant, *G. max* (4, 18, 69, 79), and has been shown to exhibit natural resistance to SCN populations currently inflicting the United States crops (9, 78).

Abiotic stress conditions, such as drought, have been shown to weaken plant defense mechanisms against pathogens (224-229). Climate prediction models anticipate an increase of temperature and extended droughts, which are proposed to amplify the spread of plant pathogens (230). Drought is the most devastating abiotic stress to soybeans (231). Drought impacts soybean directly at the root and rhizobium, which translates to overall

modifications in the plant's proteome, metabolome, transcriptome, and phenome, leading to a decline plant growth rate and yield loss up to 50% (232-236). Short-term droughts have a significant effect on yield variability in soybean, with the highest susceptibility to drought during the reproductive stage. Long-term drought sensitivity has increased in soybean crops since 1958, predominantly in central and southeastern United States (237, 238). Interestingly, this broadly correlates with the locations of SCN infestation in the United States, which has spread rapidly since the initial inoculation in 1954 in North Carolina (35). SCN is not limited to the United States, and is found throughout Canada, South America, Russia, China, Japan, Korea, and Egypt (239-241). SCN is the most destructive biotic stress of soybean crops, and causes more yield loss than any other pest (117, 223).

It has been demonstrated that a response to combination stresses cannot be extrapolated from each individual stress alone, and that a combination stress will show a non-additive, unique expression profile (227, 242). Combination of abiotic stresses, heat and drought, have been shown to elevate the level of expression of steady-state genes which belong to unique pathways not involved in the individual stress response (243). An independent abiotic stress combination expression profile is consistently observed; however, the level of transcript overlap for abiotic stress combination and the individual stress varies depending on the type of abiotic stress (243-245). This relationship is even further complicated when determining the interaction between an abiotic and biotic stress (246). The activation of one stress response pathway can have a synergistic or antagonistic effect on the stress response pathway associated with a secondary stress, an interaction that is

largely regulated by hormone signaling pathways (247). In a synergistic relationship, the cross-talk between pathways can induce cross-tolerance, such as the positive effect of UV or ozone treatment on pathogen tolerance by stimulation of salicylic acid (SA) accumulation (248-250). In an antagonistic relationship, the stimulation of one pathway can silence or reduce the sensitivity of another (225, 228, 242). Previously studied pathogen and environmental stress interactions have demonstrated that the outcome is unpredictable and specific to the stress and plants studied (229, 242, 246, 251-253). Simultaneous drought and pathogen stress can lead to drought induced pathogen resistance, pathogen induced drought tolerance, increased pathogen susceptibility by drought induced weakening, induced tolerance to combination stress by unique strategy, or increased susceptibility to both stresses due to weakening effect. Specific outcomes are determined by physiological changes mediated by abscisic acid (ABA), SA, jasmonic acid (JA), ethylene signaling, transcription factors, pathogenesis related (PR) genes, and modifications of reactive oxygen species (ROS) concentrations(229, 254, 255).

With a limited ability to predict the resulting interaction of two stresses, further investigation into important and common stress interactions is needed. A combination of stresses should be addressed as a new stress that cannot be resolved by adding two individual stress responses together. The failure of genetically modified plants for improved tolerance to perform in the field is likely due to the addition of mixed stress in the natural environment (242). A deeper understanding of the effects of a combination of stresses is required to facilitate the production of effective transgenic crops (256, 257). The natural environment does not have only one stress at a time, and therefore to better study

abiotic/biotic stress resistance/tolerance, it is necessary to understand a combined stress effect in addition to each stress individually.

In order to model the natural environment, we have chosen SCN and drought stress to study in combination due to overlapping incidences of both in the United States, agronomic importance, and predicted escalation (230). To our knowledge, gene regulation of SCN and drought stress has not been studied in combination. We make use of a well described SCN resistant *G. soja* ecotype, s54 (47, 78), in order to study the unique response when combined with drought stress.

## 2. Results

### 2.1 Differential expression profiles

Illumina sequencing generated 12 libraries of total RNA from root tissue, 3 for each condition: drought SCN, drought control, SCN, and control. After quality control and trimming adapters, 20.39 – 27.02 million filtered reads per library were generated, with a mapping rate of 76.1% - 91.3% of filtered reads aligned to *G. max* Wm82.a2v1.

Differential expression analysis of each condition to the control condition revealed differentially expressed genes (DEGs) of root tissue in response to drought, SCN, and a combination of both. A total of 57,275 mapped genes were found in each treatment condition. A higher quantity of significant DEGs were found in the combination treatment, with almost twice as many significantly regulated DEGs in the combination treatment than the SCN treatment alone. In addition, the combination stress had a strong unique signal,

with 44% of DEGs from all treatments specific to combination stress. Almost half of the combination DEGs were unique to the combination, while 47% were shared with the SCN treatment (Table 4, Figure 15).

**Table 4.** Differentially expressed genes (DEGs) present in each condition and the proportion of shared and unique DEGs for each condition.

| Condition | DEGs | Unique to condition | | Shared with SCN | | Shared with Drought | |
|---|---|---|---|---|---|---|---|
| **SCN** | 4299 | 581 | **13.51%** | | | 36 | **0.84%** |
| **Drought** | 496 | 177 | **35.69%** | 36 | **7.26%** | | |
| **SCN + Drought** | 7358 | 3588 | **48.76%** | 3487 | **47.39%** | 88 | **1.20%** |

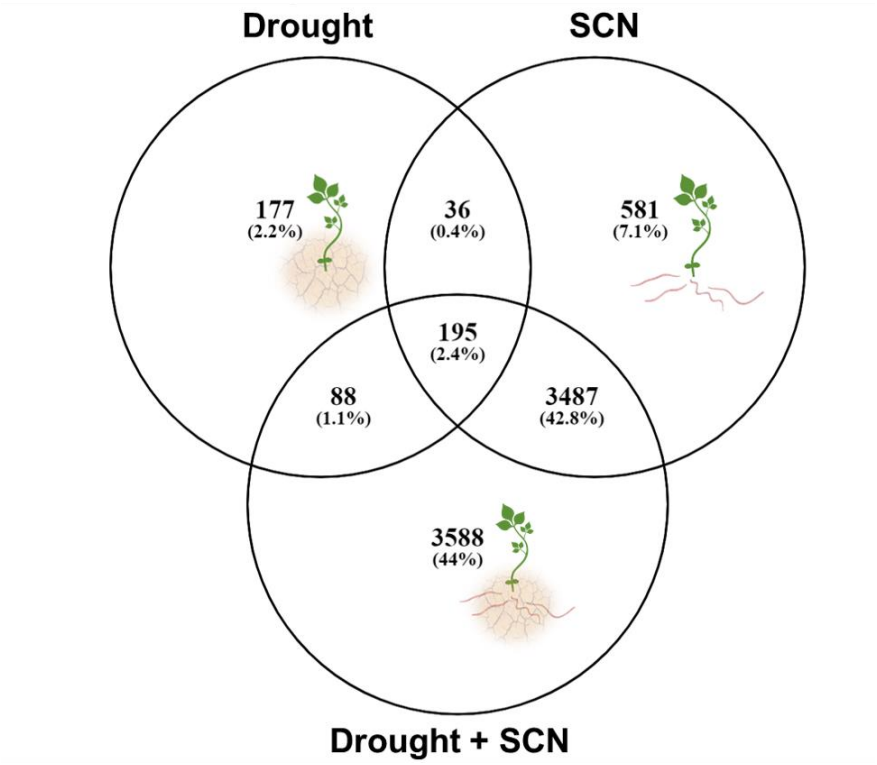



**Figure 15.** Significantly differentially expressed genes (DEGs) unique and overlapping in each condition.

Separation of the up and down significant DEGs reveals a similar pattern. SCN and Drought combination stress has the most up and down regulated genes. Interestingly, the combination stress has more downregulated genes than upregulated, while each stress individually has more upregulated genes than down regulated (Figure 16). Overall, drought stress has the lowest number of DEGs, both up and down regulated, and shares the least number of genes with the combination stress and SCN stress alone.

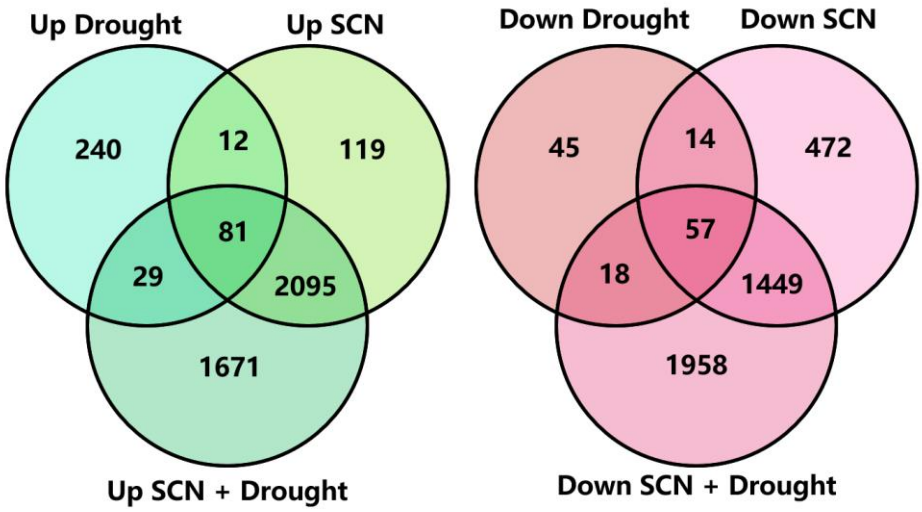



**Figure 16.** Significantly up and down regulated differentially expressed genes (DEGs) unique and overlapping in each condition.

## 2.2 Functional analysis of differentially expressed profiles

In order to determine the function of DEGs in the shared and unique expression profiles, significantly enriched overrepresented GO functions for each unique set of DEGs per treatment and for each unique shared profile were analyzed (Figure 17, Figure 18, Figure 19, Figure 20). Upregulated expression profiles reveal a shared importance of systemic acquired resistance (SAR, GO:0009862, GO:0009627, GO:0010112) , salicylic acid (SA)

biosynthesis and signaling (GO:0009751, GO:0080142, GO:0009697), ethylene biosynthesis and signaling (GO:0009693, GO:0009723, GO:0009873) , Jasmonic acid (JA) biosynthesis and signaling (GO:0009867, GO:0009864, GO:0009753, GO:0009695), abscisic acid (ABA) signaling (GO:0009738), and signal transduction (GO:0035556, GO:0007165) for SCN and combination stress (Figure 17C, S8 Table)). Similarly, SAR, SA, JA, ABA and ethylene signaling were also highly important in the combination stress alone (Figure 17B, S7 Table).). In addition, positive regulation of autophagy (GO:0010508) is overrepresented only in the combination stress (Figure 17B, S7 Table). Drought treatment alone induced upregulation unique responses, including xanthophyll catabolic process (GO:0016124), response to phosphate starvation (GO:0080040), and carotene catabolic process (GO:0016121) (Figure 17A, S6 Table).

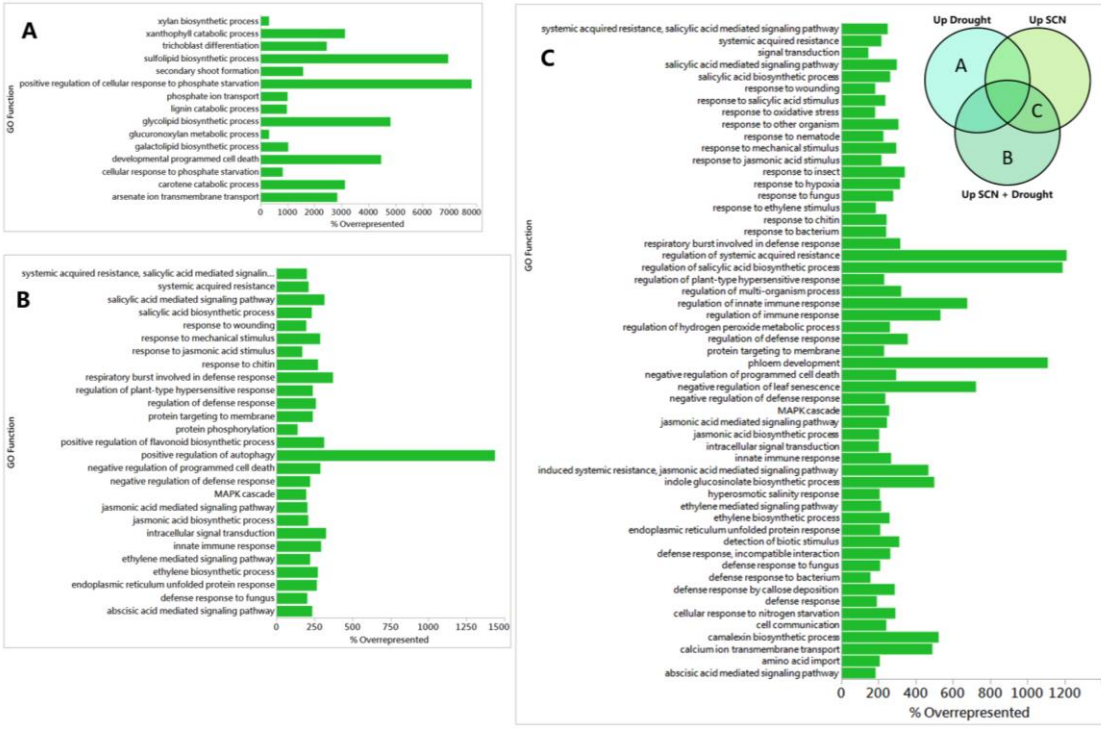



**Figure 17.** Significantly enriched overrepresented GO functions ($p<0.05$) for unique and shared upregulation profiles. Only profiles with significantly enriched functions are shown.

In addition to GO enrichment, broader biologic processes were also analyzed (Figure 18) in the interested of observing trends between treatment profiles. SCN and the combination treatment share important upregulated biological processes, including

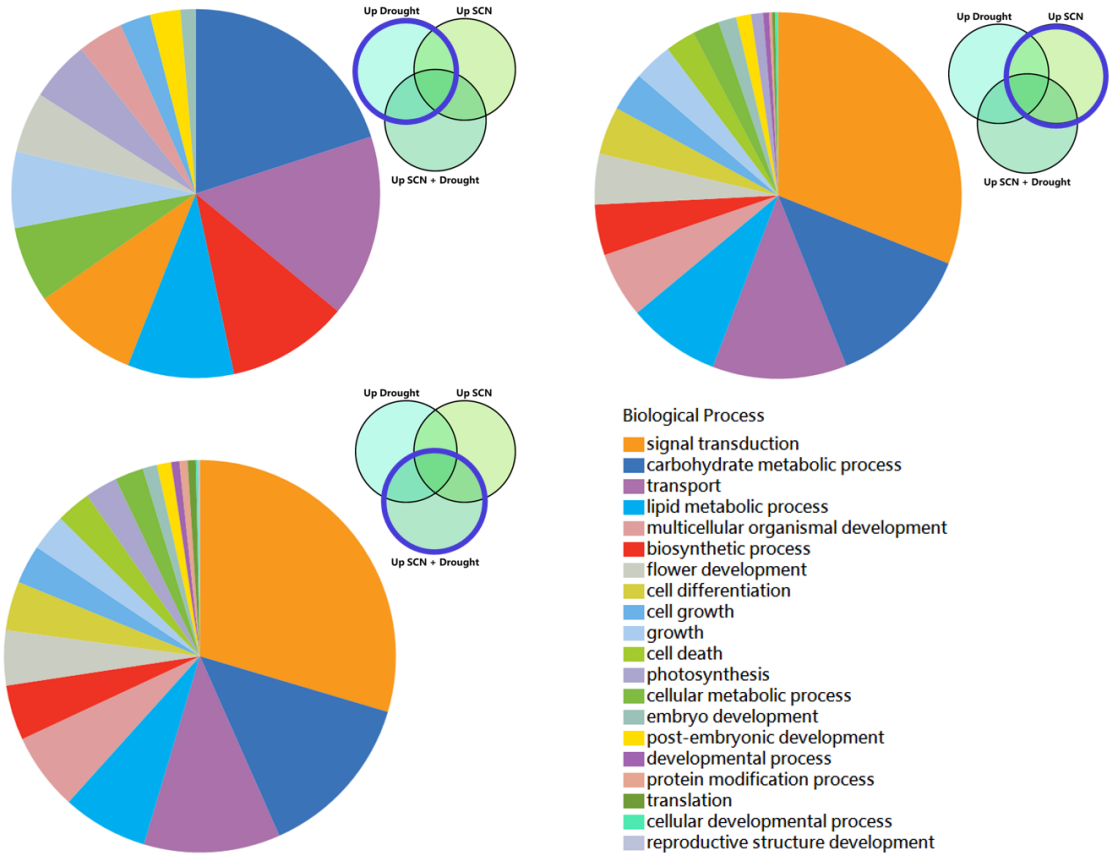



**Figure 18.** Important biological processes in upregulated profiles of drought, SCN, and combination drought and SCN treatment.

signal transduction (GO:0007165) carbohydrate metabolic process (GO:0005975), and transport (GO:0006810), and overall have very similar profiles. The profile of drought treatment differed but shared two of the most important upregulated biological processes with SCN and combination stress, carbohydrate metabolic process and transport, differing by importance of biosynthetic process (GO:0009058).

Downregulated overrepresented enriched GO functions specific to the combination stress, SCN and drought, include translation and ribosomal activity related functions (GO:0006412, GO:0042254, GO:0000462, GO:0006414), rhamnogalacturonan I side chain metabolic process, and brassinosteroid (BR) biosynthetic process (GO:0016132) (Figure 19C, S10 Table). Shared SCN and combination GO functions include downregulation of growth, and secondary and cell wall biogenesis (GO:0009834, GO:0042546) (Figure 19D, S11 Table). Only secondary cell wall biogenesis was significantly overrepresented in SCN alone (Figure 19B, S9 Table). Overrepresented functions in drought alone include cellular response to glucose (GO:0042802) and identical protein binding (GO:0071333) downregulation (Figure 19A, S9 Table). Shared between all treatment groups are downregulation of sugar and lipid transport and binding activity (GO:0008289, GO:0008515, GO:0005504, GO:0051119, GO:0015770, GO:0005319),

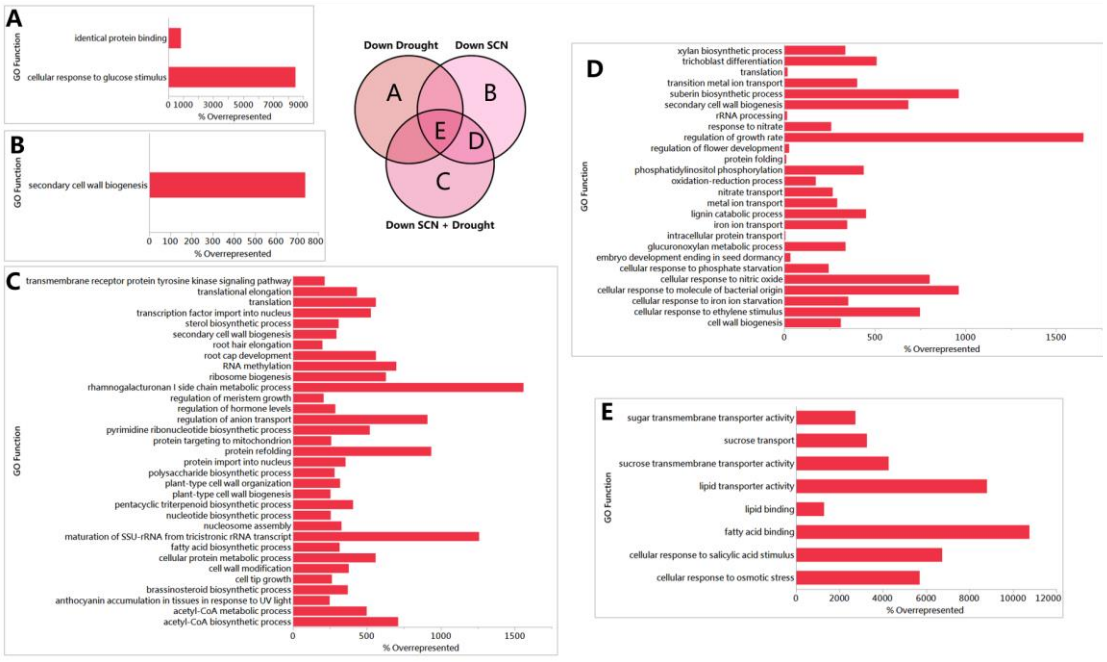



**Figure 19.** Significantly enriched overrepresented GO functions ($p<0.05$) for unique and shared downregulation profiles. Only profiles with significantly enriched functions are shown.

response to osmotic stress (GO:0071470), and response to SA (GO:0071446) (Figure 19E, S9 Table).

The importance of downregulation of carbohydrate metabolic process is common between all three treatments (Figure 20). The most substantial difference in profiles of biological processes is the high proportion of downregulated translation processes in the combination stress, making up the most represented process in this profile only (Figure 20). Pathways

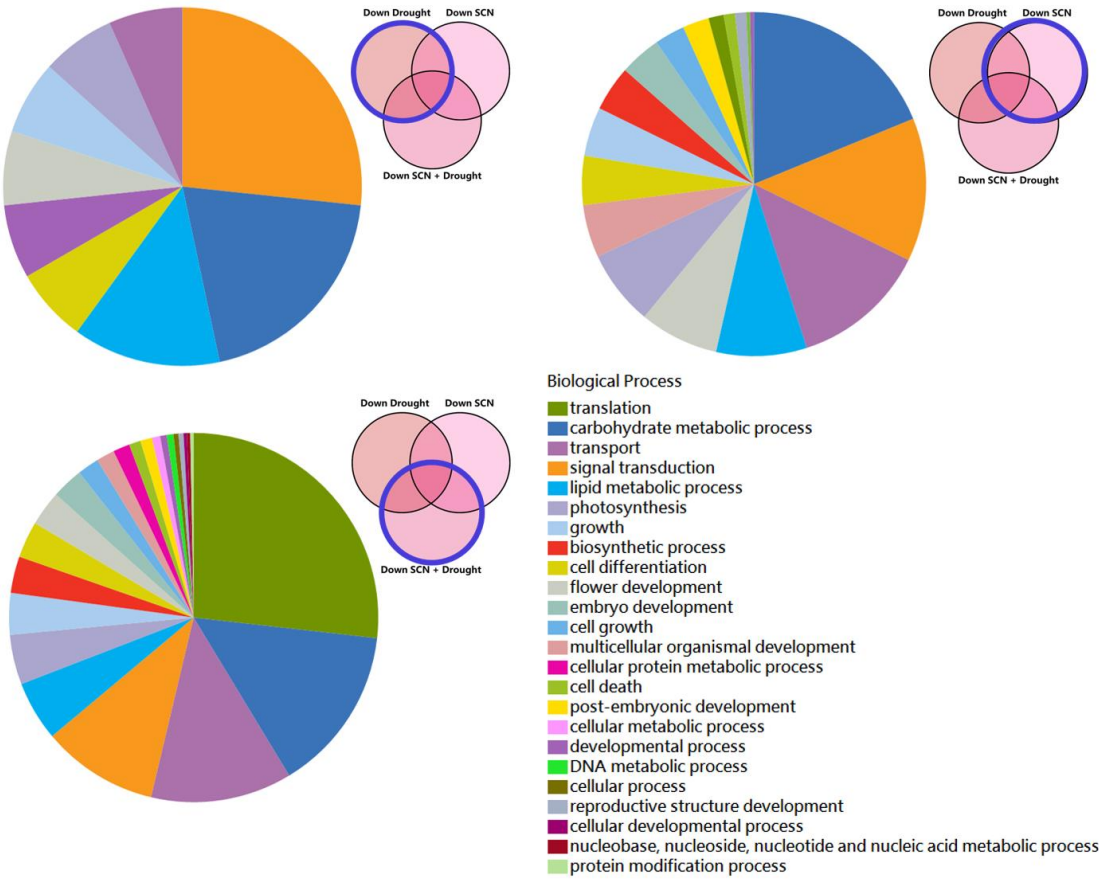



**Figure 20.** Important biological processes in downregulated profiles of drought, SCN, and combination drought and SCN treatment.

of interested were selected for further investigation based on overrepresented enriched functions and known importance in stress response.

## 2.3 Stress and resistance signaling pathways

### 2.3.1 Abscisic acid

The ABA signaling pathway has an extensive influence over plant growth and development and is commonly regulated in response to stress, sometimes inducing tolerance or resistance (258). Significant upregulation of ABA signaling is seen in the combination stress response exclusively at SRK2 (serine/threonine-protein kinase) genes *Glyma.06G160100* and *Glyma.04G205400*, and at PP2C (protein phosphatase 2C) genes *Glyma.11G222600* and *Glyma.18G035000* (Figure 21A). However, this trend is not seen for all genes coding for SRK2 and PP2C. The majority of the ABA pathway is expressed similarly in response to SCN treatment and combination SCN and drought treatment. The response to drought exhibits different expression in ABA signaling and has unique upregulation of AAO3 (abscisic-aldehyde oxidase), CYP707A3 (Abscisic acid 8'-hydroxylase 3) and AOG (abscisate beta-glucosyltransferase-like) at *Glyma.20G167500*, *Glyma.16G133800*, and *Glyma.02G104300* (Figure 21A).

### 2.3.2 Salicylic acid

SA signaling is known to play an important role in stress response and is highly enriched in the GO analysis of differentially expressed genes (155, 165). All NPR-1 coding genes in the SA signaling pathway are only significantly upregulated under combination stress

(Figure 21B), suggesting an important role in the interaction between the individual stresses SCN and drought. Different expression profiles are seen at TGA and bZIP transcription factors in response to SCN and the combination stress. A trend at TGA and bZIP of no expression in SCN treatment and upregulation with the addition of drought is seen at *Glyma.06G107300* and *Glyma.10G092100*, indicating that protein coding genes for the same protein are regulated separately in differing environmental conditions. This same trend is seen at PR1 precursor (pathogenesis-related protein 1-like protein precursor) gene *Glyma.15G062300*, but not any PR1 genes (Figure 21B). No SA related DEGs were found in response to drought.

### 2.3.3 Jasmonic acid

JA signaling pathway was analyzed further, due to its well-studied role in stress response and resistance (165), and overrepresentation in GO enrichment analysis. Expression of JAZ11(*Glyma.11G038600*) was significantly upregulated in all three conditions, but with a compounding effect in the combination stress, with the highest expression exhibited in the combination treatment group (Figure 21C). However, no other JAZ genes exhibit this trend. JAZ genes, *Glyma.15G184900* and *Glyma.13G116100* were only upregulated in the combination treatment, while all other JAZ genes were expressed in both SCN and the combination treatment. Similar expression between SCN treatment and combination stress treatment also occurs at JMT (jasmonate O-methyltransferase-like) and JAR1 (jasmonic acid-amido synthetase). Interestingly, all MYC2 transcription factor genes (*Glyma.08G271900, Glyma.09G204500, Glyma.01G096600, Glyma.16G020500*) are found to be significantly upregulated in the combination stress condition only. Lack of

significant differential expression of MYC2 transcription factor in each individual treatment suggests an important role in the interactive stress (Figure 21C),

### 2.3.4 Brassinosteroid

BR signal transduction is important in plant growth and also plays a role in modulating a plant's response to environmental changes (259). Overall, regulation of BR signaling genes is shown to differ between SCN, drought, and combination SCN and drought treated plants (Figure 21D). Similar upregulation at BAK1 (BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase) genes is seen between SCN and combination stress treatments, but is not expressed in drought treated plants. BSK2 (serine/threonine-protein kinase), BRZ1 (BRASSINAZOLE-RESISTANT 1), and XTH23 (xyloglucan endotransglucosylase/hydrolase protein 23) are only differently expressed in the combination stress treatment, and not found to be up or down regulated in response to individual treatments, suggesting an important role in the interaction between the stresses. Regulation of CYD3-1 (cyclin-D3-1) genes differs by treatment and gene, with significant downregulation in response to drought only at *Glyma.17G167700*, and shared upregulation at *Glyma.01G189700* between SCN and combination responses (Figure 21D).

### 2.3.5 Gibberellin regulation by DELLA

DELLA genes are regulators of the GA (gibberellin) signaling pathways, that can modulate GA expression and its interaction with ABA, JA, and SA signaling (260, 261), and have shown to be activated in response to osmotic stress (262). GA repressor, GAI1, is upregulated in response to both SCN treatment and the combination of SCN and drought,

but not in response to drought treatment alone (Figure 21E). Dissimilarly, DELLA SLR1 (Slender Rice1) is downregulated in both SCN and combination stress; and GA repressor DELLA DWARF8 is downregulated in response to the combination stress exclusively (Figure 21E). No DELLA genes were significantly differentially expressed in response to drought treatment exclusively.

### 2.3.6 Ethylene

Ethylene signaling is vital during plant development and known to respond to biotic and abiotic stress stimuli (263). The ethylene signaling pathway was analyzed at ethylene-responsive transcription factor 1B (ERF), which was the only gene product in the pathway that was significantly differentially expressed by any treatment group (Figure 21F). ERF upregulation is shared at two genes (*Glyma.10G186800, Glyma.20G203700*) in SCN treatment and SCN treatment with the addition of drought. However, upregulation of the ERF gene, *Glyma.10G007000*, is only found in response to SCN treatment and not in the combination stress. No ethylene signaling pathways genes are upregulated in response to drought-only treatment (Figure 21F).

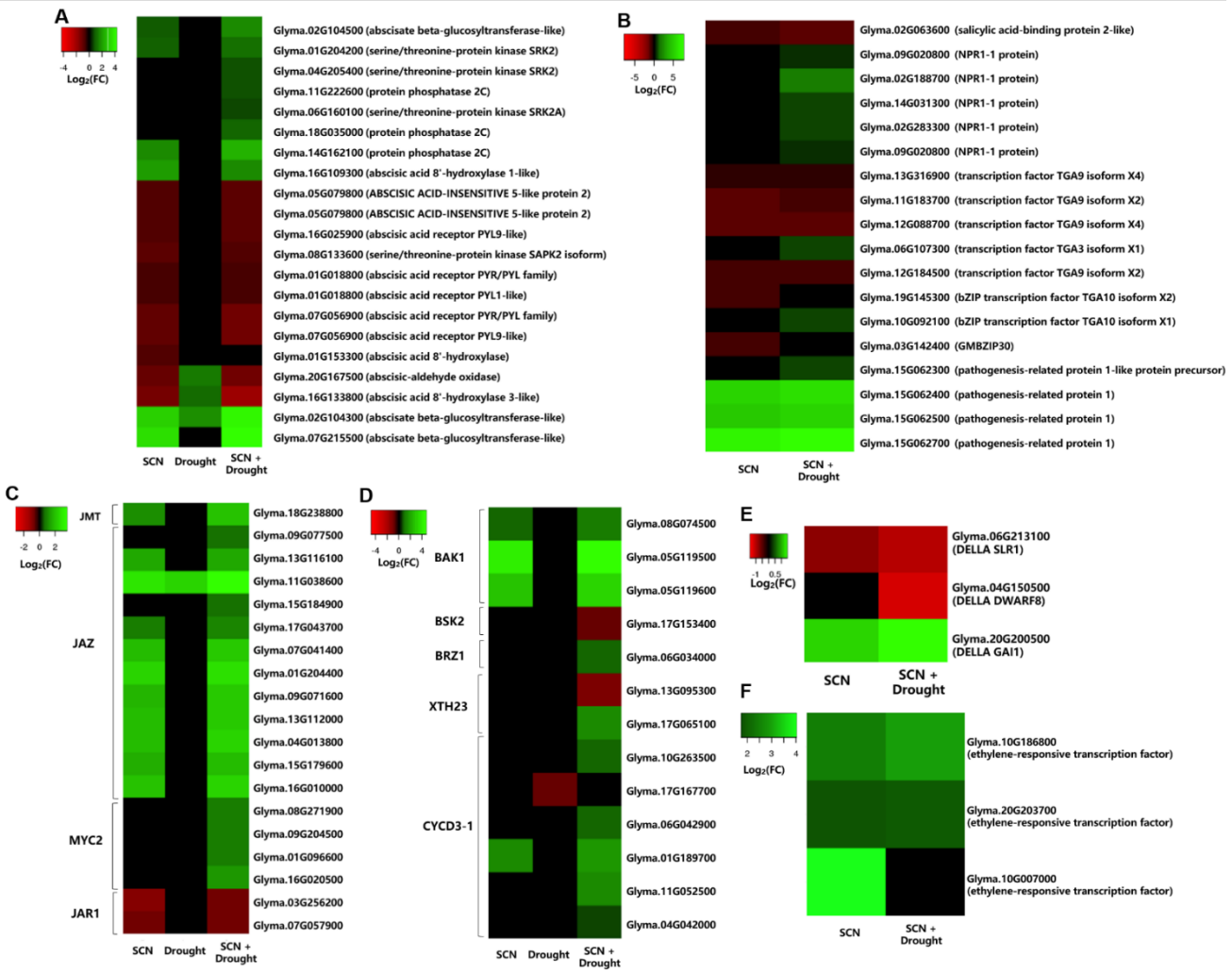



**Figure 21.** Significantly expressed genes in signal transduction pathways: ABA (A), SA (B), JA (C), BR (D), GA at DELLA genes (E), and ethylene (F) signaling pathways in Log2Foldchage (FC).

## 2.4 Differential regulation of polyamines at PAO

Differential expression within the polyamine pathway is analyzed due to known importance in resistance to biotic stress in general and findings from NRS100 SCN resistance (160, 166). Interestingly, Spermidine Synthase (SPDS) genes (*Glyma.05G036300*, *Glyma.05G036300*, *Glyma.06G126700*) and Ornithine decarboxylase (ODC1) gene (*Glyma.04G020200*) are upregulated in the SCN treatment and remain highly expressed throughout drought conditions, but are not highly expressed in drought conditions alone (Figure 22). However, other genes encoding SPDS

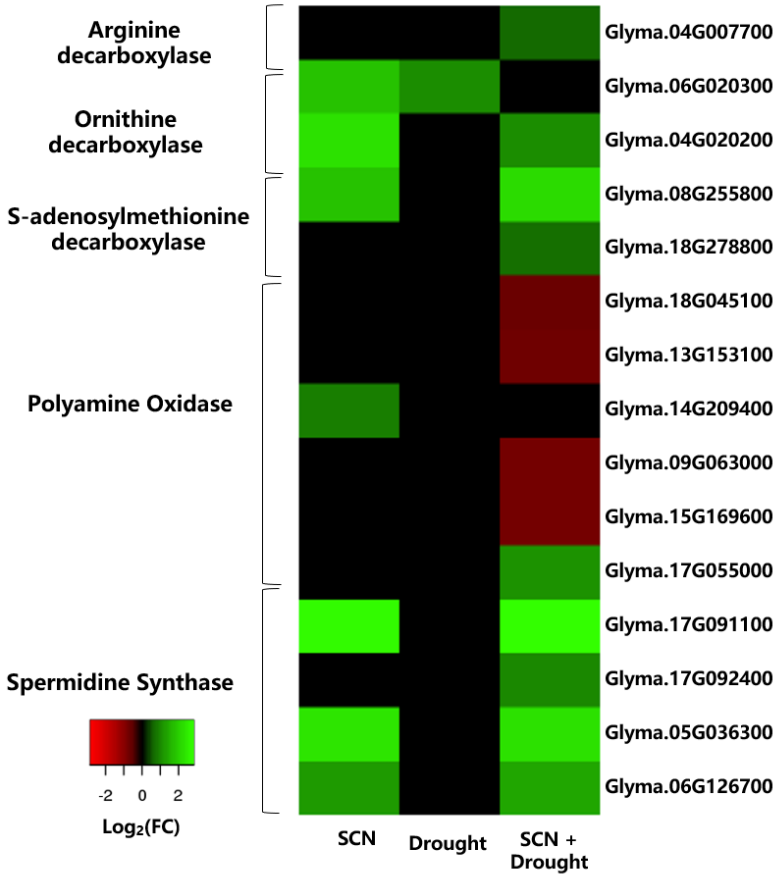



**Figure 22.** Significantly expressed genes in polyamine pathway in $Log_2$Foldchage (FC).

(*Glyma17G092400*) and ODC1 (*Glyma.06G020300*) do not exhibit this trend, in these cases ODC1 is only upregulated in SCN and drought treatment individually and SPDS is only upregulated in the combination stress. The most notable difference in the polyamine pathway between treatment groups is in the regulation of PAO (polyamine oxidase) genes, in which four out of five PAO DEG's are specifically downregulated in response to the combination stress only. The fifth PAO DEG's is exclusively upregulated in response to SCN treatment only (Figure 22). PAO enzymes are responsible for converting spermine to spermidine, and differential regulation at PAO genes could represent an importance of spermine production in combination stress (264).

**2.5 Enhanced regulation of transcription factors in response to combination stress**

Transcription factors MYB and WRKY were analyzed further due to known role in regulation in abiotic, biotic, and combination stress response and resistance (255, 265-268). An overall enhanced upregulation of WRKY genes was found when drought was added to SCN stress (Figure 23A). In addition, specific members of the WRK family were upregulated in response to combination stress that were not differentially expressed in SCN alone, WRKY 56, 49, 37, 39, 5, and probable WRKY33. No WRKY genes were found to be differentially expressed in response to drought only (Figure 23A). Upregulation of MYB 76, 4 and 184 is found in response to combination stress but not SCN alone (Figure 23B). Many MYB DEGs are similarly expressed in SCN and combination stress condition.

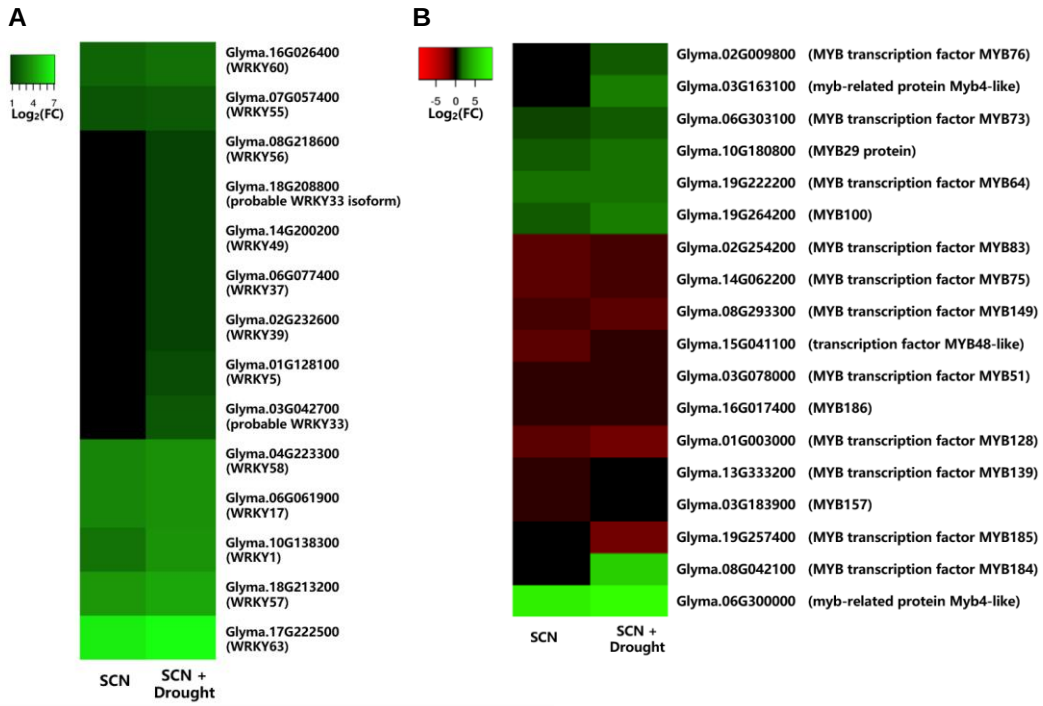



**Figure 23.** Significantly expressed MYB and WRK transcription factors in $Log_2$Foldchage (FC).

### 2.6 Regulation of reactive oxygen species differs between treatments

ROS (reactive oxygen species) are often generated in response to abiotic and biotic stress, mediated by peroxidases (266). We analyzed AOX (alternative oxidase) and Prx (peroxidase) genes (Figure 24). Although frequently associated with plant stress, no APX (ascorbate peroxidase) or CAT (catalase) genes were differentially expressed in any treatment. In both AOX and Prx genes, abiotic drought stress reveals a different expression pattern compared to SCN treatment and the combination stress (Figure 24A, B). In both AOX and Prx DEG profiles, the combination stress response includes expression of a greater number of genes than either of the individual conditions.

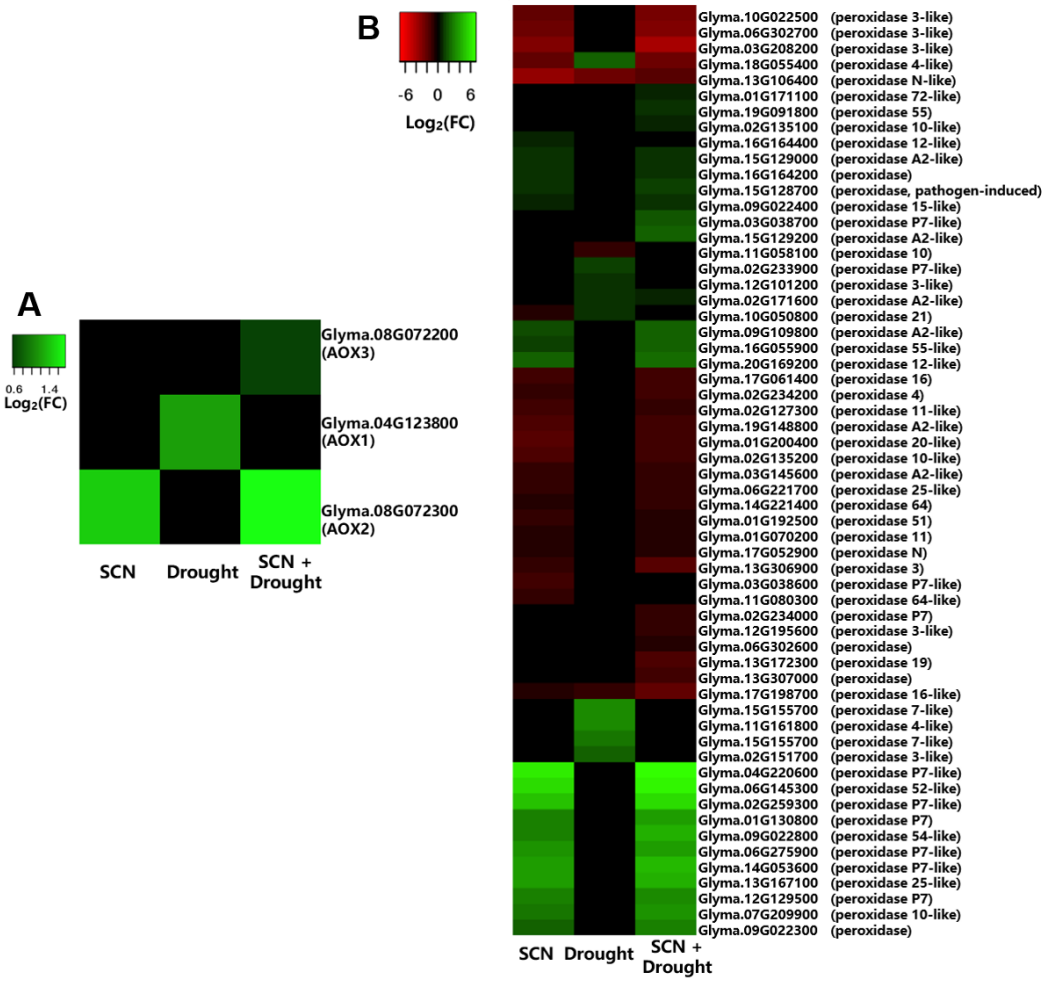



**Figure 24.** Significantly expressed Prx and AOX genes in $Log_2$Foldchage (FC).

### 2.7 Plant resistance and immune response

PR (pathogenesis related) genes and genes associated with autophagy functions are analyzed further due to their role in modulating the plant immune response to pathogens and stress (269-271), and their relative importance in the GO enrichment analysis. PR-1 genes are similarly regulated in SCN stress with and without the addition of drought, however, PTI (pathogenesis-related genes transcriptional activator) genes are differently

regulated after the addition of drought (Figure 25). Drought alone did not induce a significant expression change in any PR1, PTI genes.

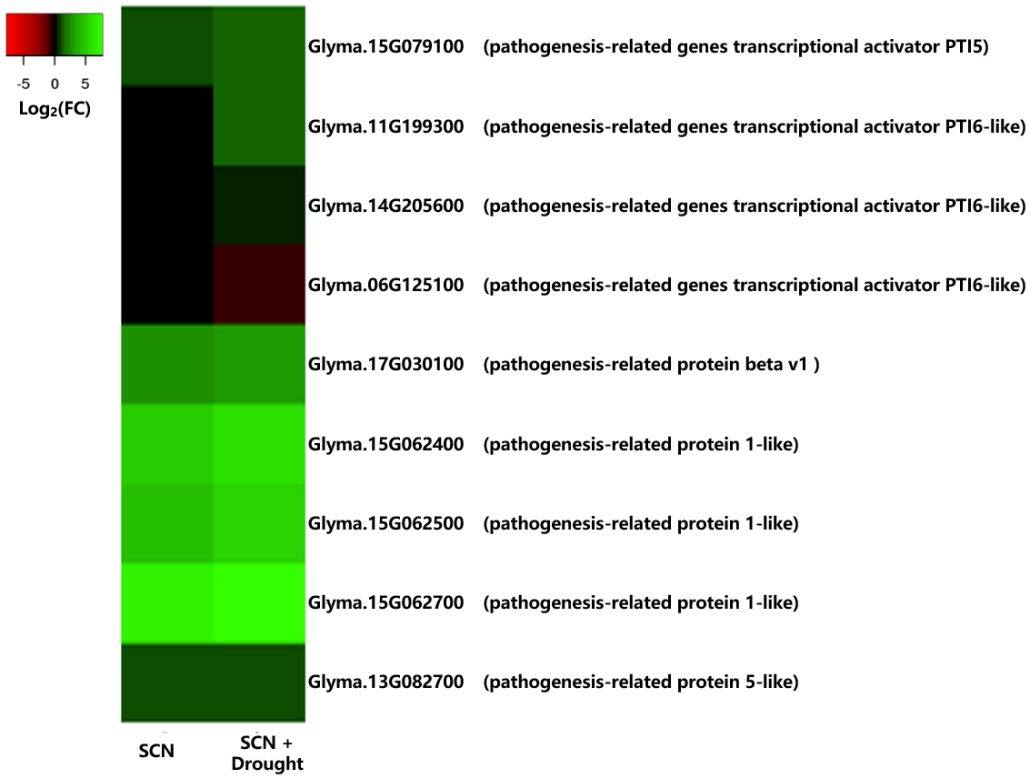



**Figure 25.** Significantly expressed PR and PTI genes in $Log_2$Foldchage (FC)

Autophagy genes which belong to overrepresented GO function, GO:0010508, were analyzed further. Overall, a larger number of autophagy related genes are differentially expressed in response to the combination stress than each stress individually, and unique upregulation at multiple autophagy related genes is found in the combination stress (Figure 26), indicating an importance of these genes in the immune response to multiple stresses. In addition, drought exclusively induced upregulation at AMP ligase gene, *Glyma.06G031600*, and shared upregulation with SCN-only response at histone binding protein *Glyma.07G13930*, but was not significantly expressed in any other autophagy related genes.

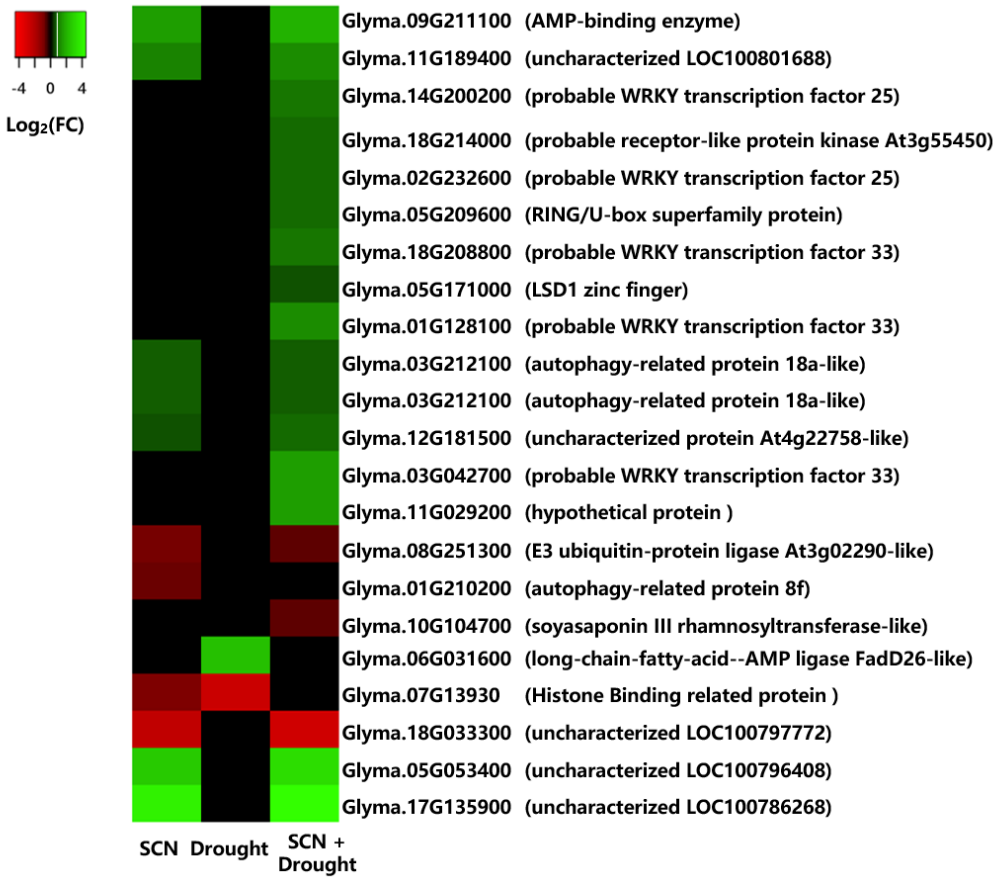



**Figure 26.** Significantly expressed genes in autophagy go function (GO:0010508), in $Log_2$Foldchage (FC)

Significantly regulated resistance genes (R genes) previously identified as responsive genes associated with SCN resistance in S54 (47) are analyzed to determine the impact of an additional stress on this potentially important regulatory step in SCN resistance. Disease resistance associated genes or R genes, Leucine-rich repeats (LRR), N-terminal coil-coiled domain with a nucleotide binding site (NBS) LRR (CC-NBS-LRR), and toll and interleukin receptor (TIR)-NBS-LR genes do not show significant differences in upregulation in response to SCN stress with and without the addition of drought stress except for LRR, *Glyma16G172300,* which was not significantly upregulated after the addition to drought

stress. *Glyma16G172300* was not a key candidate gene associated with SCN resistance found in S54 before. Dirigent, resistance-responsive genes are analyzed but not found to have significantly different profiles with the addition of drought stress.

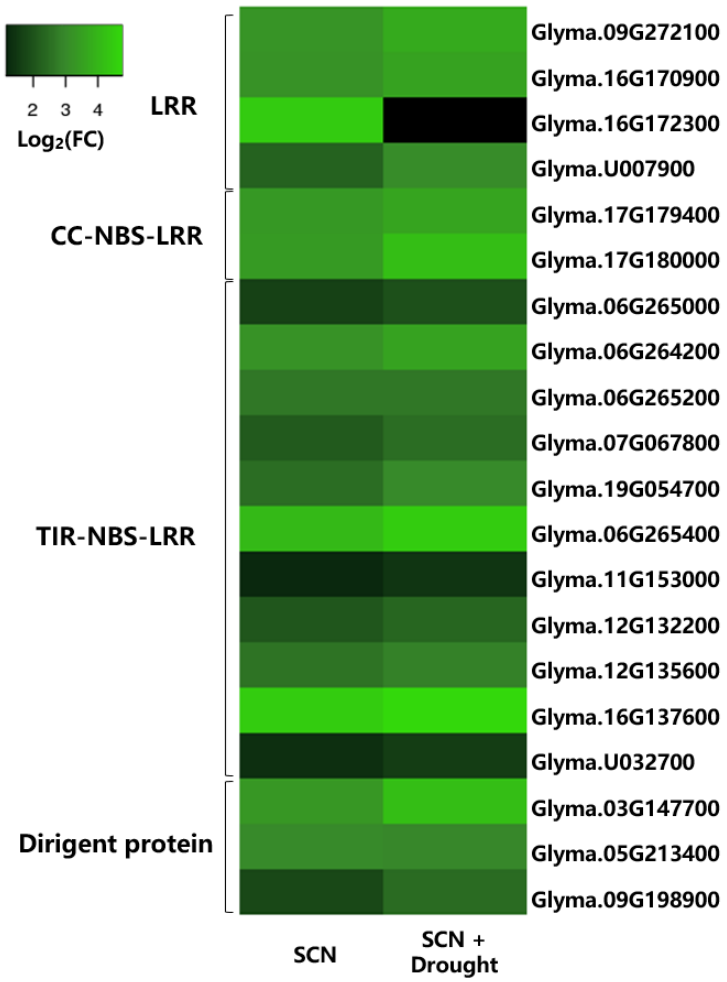



**Figure 27.** Significantly expressed R genes, in $Log_2$Foldchage (FC)

## 2.8 Validation by RT-qPCR

RNA-sequencing expression of 20 genes (S12 Table), was validated using RT-qPCR for all genotypes and condition replicates. RT-qPCR results of the 20 selected genes correlated

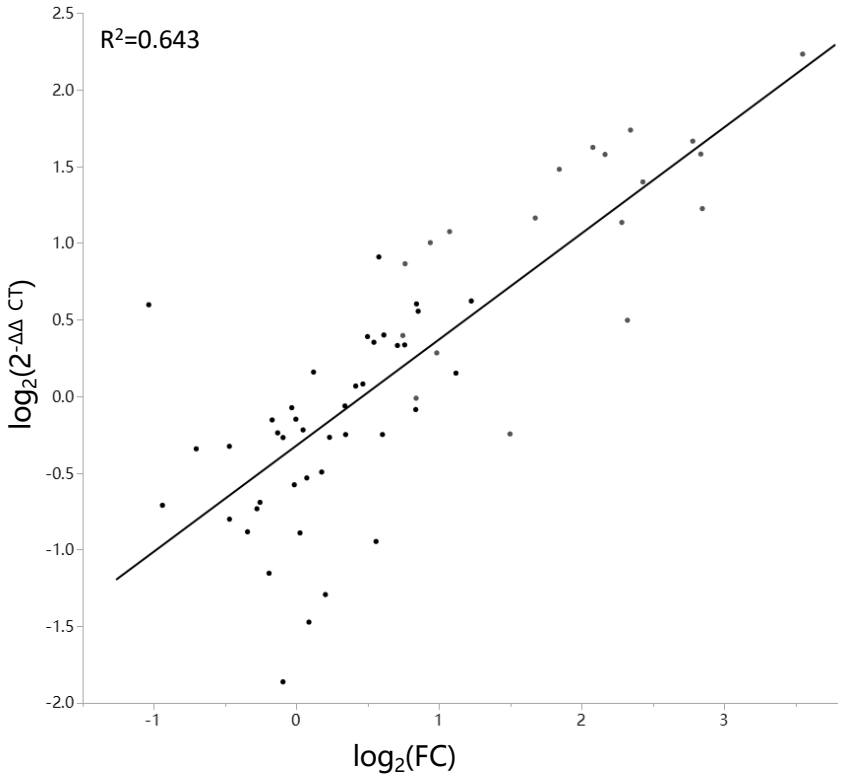



**Figure 28.** Positive correlation ($R^2$=0.643) between qPCR fold change expression values ($\log_2(2^{-\Delta\Delta CT})$) and correlated RNA-sequencing fold change expression values ($\log_2$(FC)) where FC= Fold change, for 20 genes (supplementary table 12). Individual points represent relative fold change of grouped replicates calculated by $2^{-\Delta\Delta CT}$ method and calculated fold change of group replicates from RNA-sequencing analysis.

with RNA-sequencing results. An expected indirect correlation (R=0.581) of RT-qPCR derived ΔCT to RNA-sequencing derived FPKM of all biological replicates and averaged technical replicates was found (S13 Figure). In addition, calculated fold change for from RT-qPCR correlates directly (R=0.643) with RNA-sequencing derived fold change (Figure 27). Results confirm that assumptions made from RNA-sequencing results are valid.

## 3. Discussion

Drought and SCN are the two most yield devastating abiotic and biotic stresses for soybean (223, 272, 273). The cultivated soybean, *G. max*, has a limited availability of natural resources for resistance and tolerance in the population. Considering all races in the highly adapted SCN population, SCN can overcome every type of resistance used in soybean crops (121-123). *G. soja*, the wild ancestor of *G. max*, is genetically diverse and offers an excellent natural toolkit for fighting abiotic and biotic stress in crops. Genetic widening of the cultivated crop can be done by integration of well-studied resistance alleles by genetic modification (31, 79). Our study makes use of SCN resistance found in a previously studied *G. soja* ecotype, S54, to investigate the impacts of a secondary stress, drought, on the expression of important resistance-related pathways. The results indicate that a combination stress should be considered a novel stress, and predicting the impact of combination stress on regulatory networks is not possible without this type of study. Due to the complex interactions between these two stresses, this type of investigation is needed in order to ensure that transformed crops for enhanced drought tolerance or SCN resistance will be able to perform in the field, and to further understand how current crops will manage under climatic changes.

### 3.1 Combination stress response: as distinct as drought or SCN stress response independently

*Combination stress*

We determined that that SCN and drought combination stress response is unique and separate from each stress individually. In addition, the combination of drought and SCN has a much more significant effect on soybean, resulting in the highest number of DEG's and a more enhanced level of expression on many shared DEG's. Drought alone had a much smaller effect on gene regulation than the combination of drought with SCN, in which weakening by either stress allowed for a more substantial impact by the other. Previous studies have shown that exposure to a pathogen can exaggerate the effect of the abiotic stress, and similarly, the susceptibly to a biotic stress can be heightened when enduring environmental stress (247, 274-278).

The effect of drought on SCN is unclear; conflicting studies have shown no effect of soil moisture on SCN count and increased SCN cyst number in wet soil (279, 280). Although, we safely rule out an overwhelming effect of drought on SCN given the highly inflated response in combination stress, suggesting that these treatments did not cancel each other, but instead combined forces to overwhelm the plant.

*Individual stress response*

Abiotic and biotic stresses are known to stimulate separate pathways, not surprisingly our results show that drought and SCN response consist of separate expression profiles in wild

soybean. Differential expression between the individual stress responses was most notable in the fact that SCN induced a much larger response in up and downregulated genes than drought. In addition, genes associated with hormone signaling pathways: ABA and brassinosteroid, and ROS related proteins: AOX and peroxidase genes, were regulated differently in drought stress compared SCN. Surprisingly, these genes were even regulated differently in drought stress compared to SCN and drought stress combination, in which regulation of different genes within the same functional groups were induced.

ABA signaling is important in abiotic stress response by controlling stomatal closure and various developmental traits(281). Differential expression of ABA signaling, specifically at the final protein of ABA production, abscisic aldehyde oxidase (AAO), is upregulated in drought stress but down regulated in SCN and combination SCN and drought stress. This indicates that ABA signaling is highly important in drought stress response in wild soybean, but not in SCN resistance. Only one gene, a cyclin D3-1 (CYCD3-1) gene, within the brassinosteroid signaling pathway was differently expressed, downregulated, in the drought treatment only, which is likely due to negative effects of drought stress, in turn downregulating plant growth (282). Only one, separate, CYCD3 gene was upregulated in SCN treatment alone, indicating that plant growth was not hindered by SCN stress in the SCN resistant S54.

High expression of AOX and peroxidase genes indicates high levels of ROS. Peroxidases can both produce and scavenge ROS (283). AOX is often induced in response to high ROS levels to reduce the impact of oxidative stress. In stress signaling, AOX is hypothesized to

control the signal of ROS generation and the strength of ROS signal dependent pathways in response to stress (284). Pervious findings propose that independent sets of defense genes can be activated in response to abiotic and biotic stress, which then activates similar or overlapping cascades downstream (285). This explains our findings that some similar functions are regulated in response to drought and SCN but by different proteins. The analyses of important molecular processes of each stress indicates that drought stress response was functionally very different, and SCN and combination stress had more similar patterns of expression. This indicates that SCN stress and the regulatory network that confers resistance to SCN overpowers the effect of drought stress in the combination stress reaction.

### 3.2 Complex cross-talk in combination stress response

A unique set of genes are expressed in response to combination drought and SCN stress in wild soybean, similar to previous findings in which combination stress is perceived as a novel type of stress by the plant (286, 287). Often in a combined stress on a plant, one stress response can inadvertently sabotage the effort to counteract the other. On the other hand, combined stress responses can work synergistically to overcome both stresses or maintain a stable level of resistance under an added stress(229, 242). We analyze important regulatory networks involved in pathogen resistance to determine the novel regulation of SCN resistance pathways under added drought stress.

*Transcription factor regulation*

Transcription factor regulation, particularly in the WRKY family, has a range of characterized responses to abiotic and biotic stress and resistance (158, 288-291), and has been recognized for its importance in SCN resistance specifically (47, 268, 292-294). Our results indicate that WRKY is not only important in SCN resistance for *G. soja* ecotype S54, but also highly upregulated with the addition of drought stress. This is likely a response to a severe novel stress – combination stress, in which an exaggerated response by WRKY family genes is produced in order to maintain the resistance to SCN even under added stress.

WRKY has been shown to regulate SA and JA pathways, and are key genes in the cross talk of abiotic and biotic stress in rice and tomato (265, 295). WRKY plays a role in PAMP triggered immunity and effector triggered immunity, triggered by R genes, in response to biotic stress. Both types of immunity can induce a local or systemic acquired resistance response which is modulated by accumulation of ROS and signaling hormones such as JA, SA, and ethylene (295). There is not a significant difference between the expression of R genes important to S54 before or after the addition to drought stress of SCN infected plants, which makes this a promising source of resistance in the natural environment. It has been shown that overexpression of WRKY genes can lead to activation of pathogen resistant genes, NPR1 (NON-EXPRESSOR OF PATHOGENESIS-RELATED GENES 1), within the SA pathway in order to induce resistance(296-298). Our results indicate that NPR1 is

upregulated in combination stress, which leads to the conclusion that combination stress induces an exaggerated response to SCN stress in order to maintain SCN resistance.

MYB transcription factor has also been shown to play an important role in combination abiotic and biotic stress by inducing immune responsive genes (265, 295). MYB transcription factor genes were identified as candidates for SCN race 5 resistance in wild soybean by GWAS, which is supported by our results (78). More promisingly, these genes are not significantly altered by the addition of drought stress.

MYC2 transcription factor, often associated with its important role of regulating JA signaling, is highly expressed in combination stress (299). MYC2 can fine-tune JA signaling by activating or repressing the pathway at multiple steps by regulation of other transcription factors and its interaction with JAZ (197, 198, 299). We cannot conclude if this pathway is induced or repressed by the upregulation of MYC2. Although, in SCN stress response alone, JAZ, a repressor of JA signaling (198), was upregulated. This suggests that that the pathway may be suppressed in response to SCN and then further suppressed by MYC2. Oppositely, MYC2 may play a role in the crosstalk between JA and SA, suppressing the SA pathway, as has been seen in plant-pathogen interactions before (300). JA can also interact with GA signaling, in which MYC2 regulates the interaction of the two pathways in order to prioritize the stress response or growth in changing conditions (192, 301). Additionally, MYC2 can interact a variety of other transcription factors and signaling pathways (302), and might be involved in a novel reaction to SCN and drought together.

*ROS accumulation and activation of pathogen related genes*

ROS are produced by plants to be used as signaling molecules in biotic and abiotic stress response. Production of ROS can cause oxidative stress, which is mitigated by ROS detoxifying proteins that sequester or modulate the production of ROS (303). In combination stress, expression of genes encoding ROS sequestering enzymes and PR proteins have been demonstrated to contribute to disease resistance in plants by abiotic stress stimulated accumulation of ABA and ROS that in turn mitigates the effects of the pathogen (304).

AOX modulates the stress signaling of accumulated ROS and attempts to protect the plant from oxidative stress (284). Higher observed AOX expression under combination stress indicates that more ROS has been accumulated. Interestingly AOX3 (alternative ID: AOX2b), which is only upregulated in the combination stress in our study, contains a unique motif in the promoter that is able to interact with pathogen related gene, NPR1(284, 305, 306). Our study shows that NPR1 is highly upregulated in the combination stress only, which can be explained by the combinations stress induced upregulation of AOX3. This suggests a synergistic relationship between drought and SCN stress response to induce NPR1 proteins.

Autophagy is the process of recycling molecules, which becomes very important in maintaining survivable conditions when under biotic or abiotic stress and consequently,

oxidative stress (307, 308). Functional analyses revealed that autophagy is important in the combination stress, which was supported by the different expression profile of genes related to autophagy. ROS accumulation has been shown to induce autophagy, and in response, autophagy will induce ROS scavenging (307). In this case, molecules and organelles damaged by oxidative stress can be recycled by autophagy (307, 309). Accumulation of SA can lead to increased ROS production and induce autophagy by interactions through NPR1, a receptor for SA (310, 311); our results suggest this interaction is also found in S54 under combination stress.

*Compounding suppression of ABA signaling in combination stress in order to promote SA signaling*

Production of ABA at the final protein in the process, AAO, is downregulated in both SCN and SCN and drought combination stress, but upregulated in drought stress. This indicates that ABA is not highly accumulated in either combination or SCN stress, and is not important in SCN resistance, but is important in drought stress alone. In our study, ABA signaling is suppressed at key steps in SCN and combination stress at abscisic acid receptors, PYR/PYL, and further in the combination stress only by upregulation of ABA suppressor, PP2C (312). This supports previous studies in which ABA signaling is seen to suppress SA signaling (281, 313), an important process in pathogen resistance and S54 SCN resistance specifically (47). Suppression of ABA accumulation in combination and SCN stress allows SA signaling to continue, which is not impacted by the addition of drought.

Here, SA signaling is highly upregulated at NPR1 genes and a PR1 precursor in the combination stress. Pathogen defense mediated by SA signaling has already been reported in many plants (155, 161, 165, 182), including SCN resistance in this ecotype specifically (47). It is not surprising that this pathway remains important even after the addition of drought stress. We suggest that ABA signaling is suppressed in response to SCN first by ABA receptors PYR/PYL and additionally suppressed after drought stress by SA signaling interruption at PP2C (312). This allows S54 to carry out SA mediated SCN resistance even in drought conditions.

*Alteration of growth-related pathways*

In the brassinosteroid pathway, CYCD3 was differentially regulated at one gene each in SCN stress and drought treatment individually, however 5 separate CYCD3 genes were upregulated in combination stress, suggesting upregulation of the brassinosteroid pathway and promotion of plant growth and development (282). Upregulation of BAK1 in both SCN and stress indicates this pathway remains important in combination stress. However, the expression of the brassinosteroid pathway changes with the addition of drought stress at BSK2, BRZ1, and XTH23 which might translate to differential growth of the cell wall in the combination stress (314, 315).

GA is important for plant growth and development; however, this process is not crucial in severe stress. GA signaling is negatively regulated by upregulation of DELLA GAI1 in SCN stress and further in the combination stress, suggesting an additive negative effect of drought and SCN on plant growth (316). Another DELLA gene, DWARF8 is exclusively

upregulated in combination stress. DWARF8 has not been widely studied in stress response but is known to be associated with growth and maturity time of a plant (317). The differential regulation of this gene in the combination stress only could indicate altered growth due to severe stress.

*Translation stalling in combination stress*

In addition to enhanced and specific regulation of genes at important pathways, overall translation was stalled in the combination stress. It is known that under severe stress, global translation rate is turned down in plants, indicative of translation stalling (318-320). In this situation, important stress-related mRNA may remain bound to ribosomes in polysomes to be translated while translation is halted for production of less critical proteins (321). Another possibility is that all translation is stalled due to toxicity in the plant as a protective quality control measure (320, 322, 323). Our results show that translation was highly downregulated in the combination stress specifically, indicating a similar mechanism of translation stalling, in which only critical proteins were manufactured and less urgent protein production was halted while in the severe-stress state.

*PAO increases spermine production in differential expression of polyamines*

Our results indicate that the polyamine pathway is disrupted in SCN and drought combination stress by downregulation of PAO4 and PAO5, which convert spermine to spermidine (262), and therefore downregulation would induce higher levels of spermine. High levels of spermine have been shown to induce pathogen resistance (158, 164) and abiotic stress tolerance(324-326). Arabidopsis mutants with the inability to produce

spermine exhibit hypersensitivity to drought stress (326), and spermine production has been associated with drought resistance(327-329). Although S54 is not drought resistant, we suggest that downregulation of specific PAO genes in combination stress leads to spermine accumulation which provides a protective effect against drought stress. In addition, the polyamine pathway and spermine production can trigger $H_20_2$ regulated hypersensitive response, cooperative effects with SA signaling to induce PR proteins or other modes of pathogen resistance, serve as a precursor to the production of secondary metabolites (157, 158, 164, 165, 199-203). The different regulation of the polyamine pathway by PAO downregulation in combination stress is an interesting discovery, which may impact each of the aforementioned pathways to promote a novel response.

## 4. Conclusion

Studying combination stresses allows us to model stress response in the natural environment. Drought and SCN stresses are not mutually exclusive in the natural environment, and therefore it is necessary to study them in combination. Combinations stress interactions are unpredictable, shown by the various avenues of cross-talk and outcomes; understudied, due to conventional method of focusing on one stress at a time; real and unavoidable in the natural environment; and likely the causative factor in transgenic field study failure. We suggest SCN and drought induce a complex response, including interactions between ROS signaling, SA signaling, JA signaling, and polyamine production. These findings can enhance crop improvement with genetic breeding techniques (genetic modification, gene editing) by determining precise candidate gene

capabilities in varying environments, and provides a model of applying crop wild relatives in crop improvement.

## 5. Materials and Methods

### 5.1 SCN and drought stress experiment

SCN race 5 were reared on Williams 82 soybean cultivar in a greenhouse at 27°C and 16h light/8h dark for over 30 generations. The nested sieve SCN extraction method was used to harvest SCN. Roots were washed over nested sieves (850 and 250 µm), and SCN were released from cysts by pressure into a 25 µm sieve (47). Extracted SCN eggs were purified by sucrose flotation (209) and hatched over several days on incubation trays over water. Second stage juveniles (J2) were collected from the water and diluted to a concentration of 1500 eggs/ml.

A controlled SCN, drought and nematode stress experiment was performed on SCN-resistant wild soybean accession (S54). Germinated soybeans were planted in cone planters in sterilized sand with nutrients (Stuewe & Sons, Tangent, Oregon, USA), in replicates of 12 per condition: SCN, drought, SCN and drought (combination), and control. After 2 days the SCN and combination treatment plants were inoculated with hatched J2 SCN race 5 (1500/plant) suspended in 0.09% agarose and a blank (1 ml 0.09% agarose) was added to control and drought plants. Growing conditions were kept constant at 27°C and 16h light/8h dark at 50% relative humidity in an environmental chamber. Drought and combination treatment plants were not watered after initial transplantation of germinated

plants into the sand cones. Roots were collected 6 days after SCN inoculation and tissue fragmented in liquid nitrogen in preparation for RNA sequencing. Each replicate contained roots from 4 individual plants, providing 3 replicates of each group in total.

### 5.2 RNA sequencing, alignment and assembly of transcriptomes

RNA of pool replicates was extracted using the RNeasy mini total RNA isolation kit (Qiagen, Valencia, CA, USA). RNA integrity, purity, and concentrations were assessed using an Agilent 2100 Bioanalyzer with an RNA 6000 Nano Chip (Agilent Technologies, Palo Alto, CA, USA). Messenger RNA (mRNA) was purified from the total RNA with oligo-dT beads provided in the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England BioLabs, Beverly, MA, USA). Complementary DNA (cDNA) libraries for Illumina sequencing were constructed using the NEBNext Ultra Directional RNA Library Prep Kit (NEB, Beverly, MA, USA) and NEBNext Multiplex Oligos for Illumina (NEB, Beverly, MA, USA). The mRNA was chemically fragmented and primed with random oligos for first-strand cDNA synthesis. The double-stranded cDNA was then purified, end repaired and "a-tailed" for adapter ligation. Following ligation, the samples were selected and sample-specific indexed. The final quantified libraries were pooled in equimolar amounts for sequencing on an Illumina HiSeq. 2500 utilizing a 125-bp read length with v4 sequencing chemistry (Illumina, San Diego, CA, USA).

Total RNA was obtained using the llumina 1.9 Hi-seq platform (NEB, Beverly, MA, USA). Each of the three replicates from four treatment groups (SCN, drought, combination, and control) was run on two separate lanes and merged, generating 12 libraries total. Raw fastq reads were checked for quality and adapter content with FastQC (version 0.11.6), and

adjusted using FASTx toolkit (version 0.0.13). Reads were mapped against reference genome *Glycine max* Wm82.a2.v1(211) with Tophat (version 2.1.1) and Bowtie2 (version 2.2.9). Cufflinks (Version 2.2.1) was used to assemble each transcript to the reference (212-215).

### 5.3 Comparative transcriptomics and analysis of DEGs

Assembled transcripts of treatment groups (drought, SCN, combination) were compared to the control condition in order to determine expression profiles of each condition separately. Differently expressed genes (DEGs) of the final assembly for each condition was completed using the Cufflinks (Version 2.2.1) pipeline. Genes with FDR significance ($q \leq 0.01$) were considered DEGs. Visualization of total expression profiles was used to isolate unique profiles using Venny 2.1.0 (153). Expression profiles of each treatment group and shared and unique profiles were characterized using GO annotations and enrichment data provided by SoyBase Gene Model Data Mining and Analysis Tool (http://www.soybase.org) (94, 216) and visualized using Heatmapper (330). Putative metabolic pathways for combination and individual stress DEGs are analyzed using the KEGG pathway database and visualized in Pathview 1.22.3 (218, 219).

### 5.4 RT-qPCR

Findings by RNA-sequencing are validated by RT-qPCR by comparing relative expression. Extracted RNA was quantified with NanoDrop 2009 (Thermo Fisher Scientific, Wilmington, DE, USA). Intron-spanning primers for 20 selected genes (Table S12), and reference gene UB1, were designed using Primer3 (331). RT-qPCR was

performed on ABI 7500 Fast real-time PCR system (Applied Biosystems, Foster City, CA, USA) using PerfeCTaTM SYBR® Green FastMixTM (Quanta Biosciences, Gaithersberg, MD, USA) with biological and technical triplicates. The $2^{-\Delta\Delta CT}$ method (221) was used to quantify relative expression.

REFERENCES

1. Ray DK, Mueller ND, West PC, Foley JA. Yield trends are insufficient to double global crop production by 2050. PLOS ONE. 2013;8(6):e66428.
2. USDA. Soybeans: Acreage planted, harvested, yield, production, value, and loan rate, U.S., 1960-2015. In: Agriculture ERSUSDo, editor. Oil Crops Yearbook 2016: United States Department of Agriculture; 2016.
3. Carter TE, Hymowitz T, Nelson RL. Biogeography, local adaptation, vavilov, and genetic diversity in soybean. Biological Resources and Migration. 2004:47-59.
4. Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, et al. Impacts of genetic bottlenecks on soybean genome diversity. Proceedings of the National Academy of Sciences. 2006;103(45):16666-71.
5. Qi X, Li MW, Xie M, Liu X, Ni M, Shao G, et al. Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nat Commun. 2014;5:4340.
6. Stupar RM. Into the wild: The soybean genome meets its undomesticated relative. Proceedings of the National Academy of Sciences. 2010;107(51):21947-8.
7. Zhang HY, Song BH. RNA-seq data comparisons of wild soybean genotypes in response to soybean cyst nematode (*Heterodera glycines*). Genom Data. 2017;14:36-9.
8. Qiu LJ, Xing LL, Guo Y, Wang J, Jackson SA, Chang RZ. A platform for soybean molecular breeding: the utilization of core collections for food security. Plant Mol Biol. 2013;83(1-2):41-50.
9. Kim M, Hyten DL, Niblack TL, Diers BW. Stacking resistance alleles from wild and domestic soybean sources improves soybean cyst nematode resistance (vol 51, pg 934, 2011). Crop Sci. 2011;51(5):2301-.
10. Zhang J, Wang J, Jiang W, Liu J, Yang S, Gai J, et al. Identification and analysis of NaHCO3 stress responsive genes in wild soybean (*Glycine soja)* roots by RNA-seq. Front Plant Sci. 2016;7:1842.
11. Zhang H, Li C, Davis EL, Wang J, Griffin JD, Kofsky J, et al. Genome-wide sssociation study of resistance to soybean cyst nematode (*Heterodera glycines*) HG Type 2.5.7 in wild soybean (*Glycine soja*). Front Plant Sci. 2016;7:1214.
12. Han Y, Zhao X, Liu D, Li Y, Lightfoot DA, Yang Z, et al. Domestication footprints anchor genomic regions of agronomic importance in soybeans. New Phytol. 2016;209(2):871-84.
13. Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS, et al. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor *(Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. New Phytol. 2010;188(1):242-53.
14. Qiu J, Wang Y, Wu S, Wang YY, Ye CY, Bai X, et al. Genome re-sequencing of semi-wild soybean reveals a complex *Soja* population structure and deep introgression. PLOS ONE. 2014;9(9):e108479.

15. Wang KJ, Li XH. Genetic diversity and gene flow dynamics revealed in the rare mixed populations of wild soybean (*Glycine soja*) and semi-wild type (*Glycine gracilis*) in China. Genet Resour Crop Ev. 2013;60(8):2303-18.
16. Chung WH, Jeong N, Kim J, Lee WK, Lee YG, Lee SH, et al. Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. DNA Res. 2014;21(2):153-67.
17. Sedivy EJ, Wu FQ, Hanzawa Y. Soybean domestication: the origin, genetic architecture and molecular bases. New Phytol. 2017;214(2):539-53.
18. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol. 2015;33(4):408-14.
19. Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, et al. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proc Natl Acad Sci U S A. 2010;107(51):22032-7.
20. Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, et al. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol. 2014;32(10):1045-52.
21. Zhang JP, Singh A, Mueller DS, Singh AK. Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. Plant Journal. 2015;84(6):1124-36.
22. Dong Y, Yang X, Liu J, Wang BH, Liu BL, Wang YZ. Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. Nat Commun. 2014;5.
23. Tian ZX, Wang XB, Lee R, Li YH, Specht JE, Nelson RL, et al. Artificial selection for determinate growth habit in soybean. P Natl Acad Sci USA. 2010;107(19):8563-8.
24. Zhou L, Wang SB, Jian J, Geng QC, Wen J, Song Q, et al. Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. Sci Rep. 2015;5:9350.
25. Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, et al. Fingerprinting soybean germplasm and its utility in genomic research. G3 (Bethesda). 2015;5(10):1999-2006.
26. Qiu Y-X, Fu C-X, Comes HP. Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. Molecular phylogenetics and evolution. 2011;59(1):225-44.
27. Guo J, Liu Y, Wang Y, Chen J, Li Y, Huang H, et al. Population structure of the wild soybean (*Glycine soja)* in China: implications from microsatellite analyses. Ann Bot. 2012;110(4):777-85.
28. Leamy LJ, Lee CR, Song Q, Mujacic I, Luo Y, Chen CY, et al. Environmental versus geographical effects on genomic variation in wild soybean (*Glycine soja*) across its native range in northeast Asia. Ecol Evol. 2016;6(17):6332-44.
29. Sakai M, Kanazawa A, Fujii A, Thseng FS, Abe J, Shimamoto Y. Phylogenetic relationships of the chloroplast genomes in the genus *Glycine* inferred from four intergenic spacer sequences. Plant Syst Evol. 2003;239(1-2):29-54.

30. Wang Y, Shahid MQ, Baloch FS. Phylogeographical studies of *Glycine soja*: implicating the refugium during the Quaternary glacial period and large-scale expansion after the Last Glacial Maximum. Turkish Journal of Agriculture and Forestry. 2016;40:825-38.
31. Wang Y, Shahid MQ, Ghouri F, Baloch FS, Wang Y, Huang H. Evaluation of the geographical pattern of genetic diversity of *Glycine soja* and *Glycine max* based on four single copy nuclear gene loci: For conservation of soybean germplasm. Biochemical Systematics and Ecology. 2015;62:229-35.
32. Lee TG, Kumar I, Diers BW, Hudson ME. Evolution and selection of Rhg1, a copy-number variant nematode-resistance locus. Mol Ecol. 2015;24(8):1774-91.
33. He SL, Wang YS, Li DZ, Yi TS. Environmental and historical determinants of patterns of genetic differentiation in wild soybean (*Glycine soja* Sieb. et Zucc). Sci Rep. 2016;6:22795.
34. Cook DE, Lee TG, Guo XL, Melito S, Wang K, Bayless AM, et al. Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science. 2012;338(6111):1206-9.
35. Tylka GL, Marett CC. Distribution of the soybean cyst nematode, *Heterodera glycines*, in the United States and Canada: 1954 to 2014. Plant Health Progress. 2014;15(2):13-5.
36. Maxted N, Kell SP. Establishment of a global network for the in situ conservation of crop wild relatives: Status and needs. Food and Agriculture Organization of the United Nations Commission on Genetic Resources for Food and Agriculture. 2009.
37. Zhang H, Mittal N, Leamy LJ, Barazani O, Song BH. Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. Evolutionary Applications. 2017;10(1):5-24.
38. Lam HM, Xu X, Liu X, Chen WB, Yang GH, Wong FL, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet. 2010;42(12):1053-U41.
39. Ha J, Abernathy B, Nelson W, Grant D, Wu X, Nguyen HT, et al. Integration of the draft sequence and physical map as a framework for genomic research in soybean (*Glycine max* (L.) Merr.) and wild soybean (*Glycine soja* Sieb. and Zucc.). G3 (Bethesda). 2012;2(3):321-9.
40. Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, et al. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLOS ONE. 2013;8(1):e54985.
41. Maughan PJ, Maroof MAS, Buss GR. Molecular-marker analysis of seed weight: Genomic locations, gene action, and evidence for orthologous evolution among three legume species. Theor Appl Genet. 1996;93(4):574-9.
42. Hesler LS. Resistance to soybean aphid among wild soybean lines under controlled conditions. Crop Protection. 2013;53:139-46.
43. Zhang S, Zhang Z, Bales C, Gu C, DiFonzo C, Li M, et al. Mapping novel aphid resistance QTL from wild soybean, *Glycine soja* 85-32. Theor Appl Genet. 2017;130(9):1941-52.
44. Zhang S, Zhang Z, Wen Z, Gu C, An YC, Bales C, et al. Fine mapping of the soybean aphid-resistance genes Rag6 and Rag3c from *Glycine soja* 85-32. Theor Appl Genet. 2017;130(12):2601-15.

45. Rossi M, Goggin FL, Milligan SB, Kaloshian I, Ullman DE, Williamson VM. The nematode resistance gene Mi of tomato confers resistance against the potato aphid. P Natl Acad Sci USA. 1998;95(17):9750-4.
46. Lee JS, Yoo MH, Jung JK, Bilyeu KD, Lee JD, Kang S. Detection of novel QTLs for foxglove aphid resistance in soybean. Theor Appl Genet. 2015;128(8):1481-8.
47. Zhang H, Kjemtrup-Lovelace S, Li C, Luo Y, Chen LP, Song BH. Comparative RNA-seq analysis uncovers a complex regulatory network for soybean cyst nematode resistance in wild soybean (*Glycine soja*). Sci Rep. 2017;7(1):9699.
48. Hu ZA, Wang HX. Salt tolerance of wild soybean (*Glycine soja)* in populations evaluated by a new method. Soybean Genetics Newsletter. 1997;24:79 - 80.
49. Zhang Q, Wang H, Hu Z. RAPD markers associated with salt tolerance in wild soybean populations. Soybean genetics newsletter. 1999.
50. Lee J-D, Shannon JG, Vuong TD, Nguyen HT. Inheritance of salt tolerance in wild soybean (*Glycine soja* Sieb. and Zucc.) accession PI483463. Journal of Heredity. 2009;100(6):798-801.
51. Yang D-s, Zhang J, Li M-x, Shi L-x. Metabolomics analysis reveals the salt-tolerant mechanism in *Glycine soja*. Journal of Plant Growth Regulation. 2017;36(2):460-71.
52. Ning W, Zhai H, Yu J, Liang S, Yang X, Xing X, et al. Overexpression of *Glycine soja* WRKY20 enhances drought tolerance and improves plant yields under drought stress in transgenic soybean. Molecular Breeding. 2017;37(2).
53. Anderson JE, Kono TJ, Stupar RM, Kantar MB, Morrell PL. Environmental association analyses identify candidates for abiotic stress tolerance in *Glycine soja*, the wild progenitor of cultivated soybeans. G3 (Bethesda). 2016;6(4):835-43.
54. Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Rosner BA, et al. Dietary fat intake and the risk of coronary heart disease in women. New Engl J Med. 1997;337(21):1491-9.
55. Leamy LJ, Zhang H, Li C, Chen CY, Song BH. A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*). Bmc Genomics. 2017;18(1):18.
56. Chen YW, Nelson RL. Genetic variation and relationships among cultivated, wild, and semiwild soybean. Crop Sci. 2004;44(1):316-25.
57. Diers BW, Keim P, Fehr WR, Shoemaker RC. Rflp analysis of soybean seed protein and oil content. Theor Appl Genet. 1992;83(5):608-12.
58. Chen Y, L. Nelson R. Variation in early plant height in wild soybean2006.
59. Chen GH, Wiatrak P. Seeding rate effects on soybean height, yield, and economic return. Agron J. 2011;103(5):1301-7.
60. Filho JM. Seed vigor testing: an overview of the past, present and future perspective. Sci Agr. 2015;72(4):363-74.
61. Concibido VC, La Vallee B, McLaird P, Pineda N, Meyer J, Hummel L, et al. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor Appl Genet. 2003;106(4):575-82.
62. Seo JK, Ohshima K, Lee HG, Son M, Choi HS, Lee SH, et al. Molecular variability and genetic structure of the population of Soybean mosaic virus based on the analysis of complete genome sequences. Virology. 2009;393(1):91-103.

63. Fang C, Ma Y, Yuan L, Wang Z, Yang R, Zhou Z, et al. Chloroplast DNA underwent independent selection from nuclear genes during soybean domestication and improvement. J Genet Genomics. 2016;43(4):217-21.
64. Asaf S, Khan AL, Aaqil Khan M, Muhammad Imran Q, Kang SM, Al-Hosni K, et al. Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other *Glycine* species. PLoS One. 2017;12(8):e0182281.
65. Goto H, McPherson MA, Comstock BA, Stojsin D, Ohsawa R. Likelihood assessment for gene flow of transgenes from imported genetically modified soybean (*Glycine max* (L.) Merr.) to wild soybean (*Glycine soja* Seib. et Zucc.) in Japan as a component of environmental risk assessment. Breed Sci. 2017;67(4):348-56.
66. OECD, Food, Nations AOotU. OECD-FAO Agricultural Outlook 2016-20252016.
67. Tester M, Langridge P. Breeding Technologies to Increase Crop Production in a Changing World. Science. 2010;327(5967):818-22.
68. Li Y-h, Zhao S-c, Ma J-x, Li D, Yan L, Li J, et al. Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. Bmc Genomics. 2013;14(1):579.
69. Wen ZX, Boyse JF, Song QJ, Cregan PB, Wang DC. Genomic consequences of selection and genome-wide association mapping in soybean. Bmc Genomics. 2015;16.
70. Kim MY, Van K, Kang YJ, Kim KH, Lee SH. Tracing soybean domestication history: From nucleotide to genome. Breed Sci. 2012;61(5):445-52.
71. Akram S, Hussain BN, Al Bari MA, Burritt DJ, Hossain MA. Genetic Variability and Association Analysis of Soybean (Glycine max (L.) Merrill) for Yield and Yield Attributing Traits. Plant Gene and Trait. 2016;7(13).
72. Isely D, editor Vigor tests. Proc Assoc Off Seed Anal; 1957.
73. Borevitz JO, Nordborg M. The impact of genomics on the study of natural variation in Arabidopsis. Plant physiology. 2003;132(2):718-25.
74. Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, et al. Complement Factor H Polymorphism in Age-Related Macular Degeneration. Science. 2005;308(5720):385-9.
75. Aranzana MJ, Kim S, Zhao KY, Bakker E, Horton M, Jakob K, et al. Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. Plos Genet. 2005;1(5):531-9.
76. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods. 2013;9(1):29.
77. Brachi B, Morris GP, Borevitz JO. Genome-wide association studies in plants: the missing heritability is in the field. Genome biology. 2011;12(10):232.
78. Zhang H, Li C, Davis EL, Wang J, Griffin JD, Kofsky J, et al. Genome-wide association study of resistance to soybean cyst nematode (*Heterodera glycines*) HG Type 2.5.7 in wild soybean (*Glycine soja*). Front Plant Sci. 2016;7:1214.
79. Kofsky J, Zhang HY, Song BH. The Untapped Genetic Reservoir: The Past, Current, and Future Applications of the Wild Soybean (Glycine soja). Frontiers in Plant Science. 2018;9.
80. Prince SJ, Song L, Qiu D, Maldonado Dos Santos JV, Chai C, Joshi T, et al. Genetic variants in root architecture-related genes in a *Glycine soja* accession, a potential resource to improve cultivated soybean. Bmc Genomics. 2015;16:132.

81. Asekova S, Kulkarni KP, Patil G, Kim M, Song JT, Nguyen HT, et al. Genetic analysis of shoot fresh weight in a cross of wild (G. soja) and cultivated (G. max) soybean. Molecular Breeding. 2016;36(7).
82. Wee C-D, Hashiguchi M, Ishigaki G, Muguerza M, Oba C, Abe J, et al. Evaluation of seed components of wild soybean (Glycine soja) collected in Japan using near-infrared reflectance spectroscopy. Plant Genetic Resources: Characterization and Utilization. 2018;16(2):94-102.
83. Yuan C, Zhang L, Zhao H, Wang Y, Liu X, Dong Y, et al. RNA-seq analysis for soybean cyst nematode resistance of Glycine soja (wild soybean). Oil Crop Science. 2019;4(1):33-46.
84. Yu N, Diers BW. Fine mapping of the SCN resistance QTL cqSCN-006 and cqSCN-007 from Glycine soja PI 468916. Euphytica. 2017;213(2):54.
85. Winter SMJ, Shelp BJ, Anderson TR, Welacky TW, Rajcan I. QTL associated with horizontal resistance to soybean cyst nematode in Glycine soja PI464925B. Theor Appl Genet. 2007;114(3):461-72.
86. Wang D, Diers BW, Arelli PR, Shoemaker RC. Loci underlying resistance to Race 3 of soybean cyst nematode in Glycine soja plant introduction 468916. Theoretical and Applied Genetics. 2001;103(4):561-6.
87. Kim M, Diers BW. Fine Mapping of the SCN Resistance QTL cqSCN-006 and cqSCN-007 from Glycine soja PI 468916. Crop Sci. 2013;53(3):775-85.
88. Kabelka EA, Carlson SR, Diers BW. Localization of Two Loci that Confer Resistance to Soybean Cyst Nematode from Glycine soja PI 468916. Crop Sci. 2005;45(6):2473-81.
89. Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources. 2012;4(2):359-61.
90. Wang D, Graef G, Procopiuk A, Diers B. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor Appl Genet. 2004;108(3):458-67.
91. Mansur L, Orf J, Lark K. Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (Glycine max L. Merr.). Theor Appl Genet. 1993;86(8):914-8.
92. Guzman P, Diers BW, Neece D, St Martin S, LeRoy A, Grau C, et al. QTL associated with yield in three backcross-derived populations of soybean. Crop Sci. 2007;47(1):111-22.
93. Sun D, Li W, Zhang Z, Chen Q, Ning H, Qiu L, et al. Quantitative trait loci analysis for the developmental behavior of soybean (Glycinemax L. Merr.). Theor Appl Genet. 2006;112(4):665-73.
94. Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic acids research. 2009;38(suppl_1):D843-D6.
95. Griesen D, Su D, Bérczi A, Asard H. Localization of an Ascorbate-Reducible Cytochrome b561 in the Plant Tonoplast. Plant Physiology. 2004;134(2):726-34.
96. Larkin RM, Stefano G, Ruckle ME, Stavoe AK, Sinkler CA, Brandizzi F, et al. REDUCED CHLOROPLAST COVERAGE genes from Arabidopsis thaliana help to

establish the size of the chloroplast compartment. Proceedings of the National Academy of Sciences. 2016;113(8):E1116-E25.
97. Verelst W, Asard H. Analysis of an Arabidopsis thaliana protein family, structurally related to cytochromes b561 and potentially involved in catecholamine biochemistry in plants. J Plant Physiol. 2004;161(2):175-81.
98. Weng J, Xie C, Hao Z, Wang J, Liu C, Li M, et al. Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (Zea mays L.) inbred lines. PLoS One. 2011;6(12):e29229.
99. Lin Z, Ho C-W, Grierson D. AtTRP1 encodes a novel TPR protein that interacts with the ethylene receptor ERS1 and modulates development in Arabidopsis. Journal of Experimental Botany. 2009;60(13):3697-714.
100. Schapire AL, Valpuesta V, Botella MA. TPR Proteins in Plant Hormone Signaling. Plant Signaling & Behavior. 2006;1(5):229-30.
101. Greenboim-Wainberg Y, Maymon I, Borochov R, Alvarez J, Olszewski N, Ori N, et al. Cross talk between gibberellin and cytokinin: the Arabidopsis GA response inhibitor SPINDLY plays a positive role in cytokinin signaling. The Plant Cell. 2005;17(1):92-102.
102. Zhang J, Song Q, Cregan PB, Nelson RL, Wang X, Wu J, et al. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (Glycine max) germplasm. Bmc Genomics. 2015;16(1):217.
103. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. The American Journal of Human Genetics. 2009;84(2):210-23.
104. Browning BL, Browning SR. Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society. 2007;31(5):365-75.
105. Goh L, Yap VB. Effects of normalization on quantitative traits in association test. BMC Bioinformatics. 2009;10:415-.
106. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012;28(18):2397-9.
107. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633-5.
108. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870-4.
109. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945-59.
110. Leamy LJ, Pomp D, Eisen E, Cheverud JM. Pleiotropy of quantitative trait loci for organ weights and limb bone lengths in mice. Physiol Genomics. 2002;10(1):21-9.
111. Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, et al. Accounting for multiple comparisons in a genome-wide association study (GWAS). Bmc Genomics. 2010;11:724-.
112. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic acids research. 2011;40(D1):D1178-D86.

113. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. genesis. 2015;53(8):474-85.
114. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674-6.
115. Shin J-H, Blay S, McNeney B, Graham J. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. Journal of Statistical Software. 2006;16(3):1-10.
116. Wilson RF. Seed Composition. In: Boerma HR, Specht JE, editors. Soybeans: Improvement, Production, and Uses. Agronomy Monograph. Madison, WI: American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America; 2004. p. 621-77.
117. Wrather JA, Koenning SR. Estimates of disease effects on soybean yields in the United States 2003 to 2005. Journal of nematology. 2006;38(2):173-80.
118. Tylka GL, Marett CC. Distribution of the soybean cyst nematode, Heterodera glycines, in the United States and Canada: 1954 to 2014. Plant Health Progress. 2014;15(2):85-7.
119. Mitchum MG. Soybean Resistance to the Soybean Cyst Nematode Heterodera glycines: An Update. Phytopathology. 2016;106(12):1444-50.
120. Wu X, Blake S, Sleper DA, Shannon JG, Cregan P, Nguyen HT. QTL, additive and epistatic effects for SCN resistance in PI 437654. Theor Appl Genet. 2009;118(6):1093-105.
121. Gardner M, Heinz R, Wang J, Mitchum MG. Genetics and Adaptation of Soybean Cyst Nematode to Broad Spectrum Soybean Resistance. G3: Genes|Genomes|Genetics. 2017;7(3):835.
122. Niblack TL. Soybean cyst nematode management reconsidered. Plant Dis. 2005;89(10):1020-6.
123. Niblack T, Colgrove A, Colgrove K, Bond J. Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. Plant Health Progress. 2008;9(1):29.
124. Caldwell BE, Brim CA, Ross JP. Inheritance of Resistance of Soybeans to the Cyst Nematode, Heterodera Glycines1. Agron J. 1960;52(11):635-6.
125. Hauge BM, Wang ML, Parsons JD, Parnell LD. Nucleic acid molecules and other molecules associated with soybean cyst nematode resistance. Google Patents; 2006.
126. Lightfoot D, Meksem K. Isolated polynucleotides and polypeptides relating to loci underlying resistance to soybean cyst nematode and soybean sudden death syndrome and methods employing same. Google Patents; 2002.
127. Liu X, Liu S, Jamai A, Bendahmane A, Lightfoot DA, Mitchum MG, et al. Soybean cyst nematode resistance in soybean is independent of the Rhg4 locus LRR-RLK gene. Functional & integrative genomics. 2011;11(4):539-49.
128. Melito S, Heuberger AL, Cook D, Diers BW, MacGuidwin AE, Bent AF. A nematode demographics assay in transgenic roots reveals no significant impacts of the Rhg1locus LRR-Kinase on soybean cyst nematode resistance. BMC Plant Biology. 2010;10(1):104.

129. Cook DE, Bayless AM, Wang K, Guo X, Song Q, Jiang J, et al. Distinct Copy Number, Coding Sequence, and Locus Methylation Patterns Underlie Rhg1-Mediated Soybean Resistance to Soybean Cyst Nematode. Plant Physiology. 2014;165(2):630-47.
130. Liu S, Kandoth PK, Warren SD, Yeckel G, Heinz R, Alden J, et al. A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. Nature. 2012;492:256.
131. Meksem K, Liu S, Pramod K, Lakhssassi N, Colantonio V, Kang JW, et al. The GmSNAP18 is the Peking-type rhg1-a gene for resistance to soybean cyst nematode. Phytopathology. 2017;107(1):6-.
132. Yu N, Lee TG, Rosa DP, Hudson M, Diers BW. Impact of Rhg1 copy number, type, and interaction with Rhg4 on resistance to Heterodera glycines in soybean. Theor Appl Genet. 2016;129(12):2403-12.
133. Patil GB, Lakhssassi N, Wan J, Song L, Zhou Z, Klepadlo M, et al. Whole-genome re-sequencing reveals the impact of the interaction of copy number variants of the rhg1 and Rhg4 genes on broad-based resistance to soybean cyst nematode. Plant biotechnology journal. 2019.
134. Wu X-Y, Zhou G-C, Chen Y-X, Wu P, Liu L-W, Ma F-F, et al. Soybean cyst nematode resistance emerged via artificial selection of duplicated serine hydroxymethyltransferase genes. Frontiers in plant science. 2016;7:998.
135. Peng D, Peng H, Wu D, Huang W, Ye W, Cui J. First report of soybean cyst nematode (Heterodera glycines) on soybean from Gansu and Ningxia China. Plant Dis. 2016;100(1):229-.
136. Yan GP, Baidoo R. Current Research Status of Heterodera glycines Resistance and Its Implication on Soybean Breeding. Engineering-Prc. 2018;4(4):534-41.
137. He SL, Wang YS, Li DZ, Yi TS. Environmental and Historical Determinants of Patterns of Genetic Differentiation in Wild Soybean (Glycine soja Sieb. et Zucc). Scientific Reports. 2016;6.
138. Ning W, Zhai H, Yu J, Liang S, Yang X, Xing X, et al. Overexpression of Glycine soja WRKY20 enhances drought tolerance and improves plant yields under drought stress in transgenic soybean. Molecular breeding. 2017;37(2):19.
139. Munns R, James RA, Xu B, Athman A, Conn SJ, Jordans C, et al. Wheat grain yield on saline soils is improved by an ancestral Na+ transporter gene. Nature biotechnology. 2012;30(4):360.
140. Arbelaez JD, Moreno LT, Singh N, Tung C-W, Maron LG, Ospina Y, et al. Development and GBS-genotyping of introgression lines (ILs) using two wild species of rice, O. meridionalis and O. rufipogon, in a common recurrent parent, O. sativa cv. Curinga. Molecular Breeding. 2015;35(2):81.
141. Von Korff M, Wang H, Léon J, Pillen K. Development of candidate introgression lines using an exotic barley accession (Hordeum vulgare ssp. spontaneum) as donor. Theor Appl Genet. 2004;109(8):1736-45.
142. Naz AA, Arifuzzaman M, Muzammil S, Pillen K, Léon J. Wild barley introgression lines revealed novel QTL alleles for root and related shoot traits in the cultivated barley (Hordeum vulgare L.). BMC genetics. 2014;15(1):107.
143. Hajjar R, Hodgkin T. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica. 2007;156(1-2):1-13.

144. Zhou Y-L, Xu J-L, Zhou S-C, Yu J, Xie X-W, Xu M-R, et al. Pyramiding Xa23 and Rxo1 for resistance to two bacterial diseases into an elite indica rice variety using molecular approaches. Molecular breeding. 2009;23(2):279-87.
145. Ballini E, Berruyer R, Morel JB, Lebrun MH, Nottéghem JL, Tharreau D. Modern elite rice varieties of the 'Green Revolution'have retained a large introgression from wild rice around the Pi33 rice blast resistance locus. New Phytol. 2007;175(2):340-50.
146. Menda N, Strickler SR, Edwards JD, Bombarely A, Dunham DM, Martin GB, et al. Analysis of wild-species introgressions in tomato inbreds uncovers ancestral origins. Bmc Plant Biology. 2014;14:16.
147. Kumar A, Tiwari K, Datta D, Singh M. Marker assisted gene pyramiding for enhanced Tomato leaf curl virus disease resistance in tomato cultivars. Biologia plantarum. 2014;58(4):792-7.
148. Concibido VC, Diers BW, Arelli PR. A decade of QTL mapping for cyst nematode resistance in soybean. Crop Sci. 2004;44(4):1121-31.
149. Meksem K, Ruben E, Hyten DL, Schmidt ME, Lightfoot DA. High-throughput genotyping for a polymorphism linked to soybean cyst nematode resistance gene Rhg4 by using TaqmanTM probes. Molecular Breeding. 2001;7(1):63-71.
150. Lakhssassi N, Liu S, Bekal S, Zhou Z, Colantonio V, Lambert K, et al. Characterization of the Soluble NSF Attachment Protein gene family identifies two members involved in additive resistance to a plant pathogen. Scientific Reports. 2017;7:45226.
151. Yu N, Diers BW. Fine mapping of the SCN resistance QTL cqSCN-006 and cqSCN-007 from Glycine soja PI 468916. Euphytica. 2017;213(2).
152. Riechmann JL, Meyerowitz EM. The AP2/EREBP family of plant transcription factors. Biological chemistry. 1998;379:633-46.
153. Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. http://bioinfogp cnb csic es/tools/venny/index html. 2007.
154. Dixon RA. Natural products and plant disease resistance. Nature. 2001;411(6839):843-7.
155. Durner J, Shah J, Klessig DF. Salicylic acid and disease resistance in plants. Trends in Plant Science. 1997;2(7):266-74.
156. Lang J, Genot B, Hirt H, Colcombet J. Constitutive activity of the Arabidopsis MAP Kinase 3 confers resistance to Pseudomonas syringae and drives robust immune responses. Plant Signaling & Behavior. 2017;12(8):e1356533.
157. Wittstock U, Gershenzon J. Constitutive plant toxins and their role in defense against herbivores and pathogens. Current Opinion in Plant Biology. 2002;5(4):300-7.
158. Zhang J, Peng Y, Guo Z. Constitutive expression of pathogen-inducible OsWRKY31 enhances disease resistance and affects root growth and auxin response in transgenic rice plants. Cell Research. 2007;18:508.
159. Angelini R, Bragaloni M, Federico R, Infantino A, Porta-Pugua A. Involvement of polyamines, diamine oxidase and peroxidase in resistance of chickpea to Ascochyta rabiei. Journal of Plant Physiology. 1993;142(6):704-9.
160. Yoda H, Fujimura K, Takahashi H, Munemura I, Uchimiya H, Sano H. Polyamines as a common source of hydrogen peroxide in host- and nonhost hypersensitive response during pathogen infection. Plant Molecular Biology. 2009;70(1-2):103-12.

161. Ali MA, Anjam MS, Nawaz MA, Lam HM, Chung G. Signal Transduction in Plant-Nematode Interactions. International Journal of Molecular Sciences. 2018;19(6).
162. McConn M, Creelman RA, Bell E, Mullet JE, Browse J. Jasmonate is essential for insect defense Arabidopsis. P Natl Acad Sci USA. 1997;94(10):5473-7.
163. Nahar K, Kyndt T, De Vleesschauwer D, Hofte M, Gheysen G. The Jasmonate Pathway Is a Key Player in Systemically Induced Defense against Root Knot Nematodes in Rice. Plant Physiology. 2011;157(1):305-16.
164. Zhang L, Zhang F, Melotto M, Yao J, He SY. Jasmonate signaling and manipulation by pathogens and insects. Journal of experimental botany. 2017;68(6):1371-85.
165. Reymond P, Farmer EE. Jasmonate and salicylate as global signals for defense gene expression. Current Opinion in Plant Biology. 1998;1(5):404-11.
166. Takahashi Y. The role of polyamines in plant disease resistance. Environmental Control in Biology. 2016;54(1):17-21.
167. Walters D. Resistance to plant pathogens: possible roles for free polyamines and polyamine catabolism. New Phytol. 2003;159(1):109-15.
168. Howland A, Monnig N, Mathesius J, Nathan M, Mitchum MG. Survey of Heterodera glycines Population Densities and Virulence Phenotypes During 2015–2016 in Missouri. Plant Dis. 2018;102(12):2407-10.
169. Bayless AM, Zapotocny RW, Grunwald DJ, Amundson KK, Diers BW, Bent AF. An atypical N-ethylmaleimide sensitive factor enables the viability of nematode-resistant Rhg1 soybeans. Proceedings of the National Academy of Sciences. 2018;115(19):E4512-E21.
170. Bayless AM, Smith JM, Song J, McMinn PH, Teillet A, August BK, et al. Disease resistance through impairment of α-SNAP–NSF interaction and vesicular trafficking by soybean Rhg1. Proceedings of the National Academy of Sciences. 2016;113(47):E7375-E82.
171. Bekal S, Domier LL, Gonfa B, Lakhssassi N, Meksem K, Lambert KN. A SNARE-like protein and biotin are implicated in soybean cyst nematode virulence. PLoS One. 2015;10(12):e0145601.
172. Zander M, Thurow C, Gatz C. TGA Transcription Factors Activate the Salicylic Acid-Suppressible Branch of the Ethylene-Induced Defense Program by Regulating ORA59 Expression. Plant physiology. 2014;165(4):1671-83.
173. Chang KN, Zhong S, Weirauch MT, Hon G, Pelizzola M, Li H, et al. Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. Elife. 2013;2:e00675.
174. McGrath KC, Dombrecht B, Manners JM, Schenk PM, Edgar CI, Maclean DJ, et al. Repressor-and activator-type ethylene response factors functioning in jasmonate signaling and disease resistance identified via a genome-wide screen of Arabidopsis transcription factor gene expression. Plant physiology. 2005;139(2):949-59.
175. Nakano T, Suzuki K, Ohtsuki N, Tsujimoto Y, Fujimura T, Shinshi H. Identification of genes of the plant-specific transcription-factor families cooperatively regulated by ethylene and jasmonate in Arabidopsis thaliana. Journal of plant research. 2006;119(4):407-13.

176. Hu Y, You J, Li C, Williamson VM, Wang C. Ethylene response pathway modulates attractiveness of plant roots to soybean cyst nematode Heterodera glycines. Scientific reports. 2017;7:41282-.
177. Klink VP, Hosseini P, Matsye PD, Alkharouf NW, Matthews BF. Syncytium gene expression in Glycine max [PI 88788] roots undergoing a resistant reaction to the parasitic nematode Heterodera glycines. Plant Physiology and Biochemistry. 2010;48(2-3):176-93.
178. Li R, Rashotte AM, Singh NK, Weaver DB, Lawrence KS, Locy RD. Integrated signaling networks in plant responses to sedentary endoparasitic nematodes: a perspective. Plant Cell Reports. 2015;34(1):5-22.
179. Mazarei M, Puthoff DP, Hart JK, Rodermel SR, Baum TJ. Identification and characterization of a soybean ethylene-responsive element-binding protein gene whose mRNA expression changes during soybean cyst nematode infection. Molecular Plant-Microbe Interactions. 2002;15(6):577-86.
180. Nahar K, Kyndt T, De Vleesschauwer D, Höfte M, Gheysen G. The Jasmonate Pathway Is a Key Player in Systemically Induced Defense against Root Knot Nematodes in Rice. Plant Physiology. 2011;157(1):305-16.
181. Tucker ML, Xue P, Yang R. 1-Aminocyclopropane-1-carboxylic acid (ACC) concentration and ACC synthase expression in soybean roots, root tips, and soybean cyst nematode (Heterodera glycines)-infected roots. Journal of experimental botany. 2009;61(2):463-72.
182. Feys BJ, Parker JE. Interplay of signaling pathways in plant disease resistance. Trends Genet. 2000;16(10):449-55.
183. Farmer EE, Ryan CA. Octadecanoid precursors of jasmonic acid activate the synthesis of wound-inducible proteinase inhibitors. The Plant Cell. 1992;4(2):129-34.
184. Ruan J, Zhou Y, Zhou M, Yan J, Khurshid M, Weng W, et al. Jasmonic Acid Signaling Pathway in Plants. International Journal of Molecular Sciences. 2019;20(10).
185. Okada K, Abe H, Arimura G-i. Jasmonates Induce Both Defense Responses and Communication in Monocotyledonous and Dicotyledonous Plants. Plant and Cell Physiology. 2014;56(1):16-27.
186. Avila CA, Arévalo-Soliz LM, Jia L, Navarre DA, Chen Z, Howe GA, et al. Loss of function of FATTY ACID DESATURASE7 in tomato enhances basal aphid resistance in a salicylate-dependent manner. Plant physiology. 2012;158(4):2028-41.
187. Beneventi MA, da Silva OB, de Sá MEL, Firmino AAP, de Amorim RMS, Albuquerque ÉVS, et al. Transcription profile of soybean-root-knot nematode interaction reveals a key role of phythormones in the resistance reaction. Bmc Genomics. 2013;14(1):322.
188. Zhou G, Qi J, Ren N, Cheng J, Erb M, Mao B, et al. Silencing OsHI-LOX makes rice more susceptible to chewing herbivores, but enhances resistance to a phloem feeder. The Plant Journal. 2009;60(4):638-48.
189. Felton GW, Korth KL, Bi JL, Wesley SV, Huhman DV, Mathews MC, et al. Inverse relationship between systemic resistance of plants to microorganisms and to insect herbivory. Current Biology. 1999;9(6):317-20.
190. Gupta V, Willits MG, Glazebrook J. Arabidopsis thaliana EDS4 contributes to salicylic acid (SA)-dependent expression of defense responses: evidence for inhibition of

jasmonic acid signaling by SA. Molecular Plant-Microbe Interactions. 2000;13(5):503-11.
191. Turner JG, Ellis C, Devoto A. The Jasmonate Signal Pathway. The Plant Cell. 2002;14(suppl 1):S153-S64.
192. Yang D-L, Yao J, Mei C-S, Tong X-H, Zeng L-J, Li Q, et al. Plant hormone jasmonate prioritizes defense over growth by interfering with gibberellin signaling cascade. Proceedings of the National Academy of Sciences. 2012;109(19):E1192-E200.
193. Musser RO, Hum-Musser SM, Eichenseer H, Peiffer M, Ervin G, Murphy JB, et al. Herbivory: caterpillar saliva beats plant defences. Nature. 2002;416(6881):599.
194. Kus JV, Zaton K, Sarkar R, Cameron RK. Age-Related Resistance in Arabidopsis Is a Developmentally Regulated Defense Response to <em>Pseudomonas syringae</em&gt. The Plant Cell. 2002;14(2):479.
195. Zhai Q, Zhang X, Wu F, Feng H, Deng L, Xu L, et al. Transcriptional mechanism of jasmonate receptor COI1-mediated delay of flowering time in Arabidopsis. The Plant Cell. 2015;27(10):2814-28.
196. Xie D-X, Feys BF, James S, Nieto-Rostro M, Turner JG. COI1: an Arabidopsis gene required for jasmonate-regulated defense and fertility. Science. 1998;280(5366):1091-4.
197. Thines B, Katsir L, Melotto M, Niu Y, Mandaokar A, Liu G, et al. JAZ repressor proteins are targets of the SCFCOI1 complex during jasmonate signalling. Nature. 2007;448:661.
198. Chini A, Fonseca S, Fernández G, Adie B, Chico JM, Lorenzo O, et al. The JAZ family of repressors is the missing link in jasmonate signalling. Nature. 2007;448:666.
199. Walters DR. Polyamines and plant disease. Phytochemistry. 2003;64(1):97-107.
200. Walters DR. Polyamines in plant–microbe interactions. Physiological and Molecular Plant Pathology. 2000;57(4):137-46.
201. Fu X-Z, Chen C-W, Wang Y, Liu J-H, Moriguchi T. Ectopic expression of MdSPDS1 in sweet orange (Citrus sinensis Osbeck) reduces canker susceptibility: involvement of H2O2 production and transcriptional alteration. BMC Plant Biology. 2011;11(1):55.
202. Seifi HS, Shelp BJ. Spermine Differentially Refines Plant Defense Responses Against Biotic and Abiotic Stresses. Frontiers in Plant Science. 2019;10(117).
203. Pal M, Janda T. Role of polyamine metabolism in plant pathogen interactions. J Plant Sci Phytopathol. 2017;1:095-0100.
204. Berta G, Altamura MM, Fusconi A, Cerruti F, Capitani F, Bagni N. The plant cell wall is altered by inhibition of polyamine biosynthesis. New Phytol. 1997;137(4):569-77.
205. Martin-Tanguy J. Conjugated polyamines and reproductive development: biochemical, molecular and physiological approaches. Physiologia Plantarum. 1997;100(3):675-88.
206. Hewezi T, Howe PJ, Maier TR, Hussey RS, Mitchum MG, Davis EL, et al. Arabidopsis Spermidine Synthase Is Targeted by an Effector Protein of the Cyst Nematode Heterodera schachtii. Plant Physiology. 2010;152(2):968-84.
207. Capell T, Escobar C, Liu H, Burtin D, Lepri O, Christou P. Over-expression of the oat arginine decarboxylase cDNA in transgenic rice (Oryza sativa L.) affects normal development patterns in vitro and results in putrescine accumulation in transgenic plants. Theor Appl Genet. 1998;97(1):246-54.

208. Bhatnagar P, Glasheen BM, Bains SK, Long SL, Minocha R, Walter C, et al. Transgenic Manipulation of the Metabolism of Polyamines in Poplar Cells. Plant Physiology. 2001;125(4):2139.
209. Matthews B, MacDonald M, Thai V, Tucker M. Molecular characterization of arginine kinases in the soybean cyst nematode (Heterodera glycines). Journal of nematology. 2003;35(3):252.
210. Niblack T, Tylka GL, Arelli P, Bond J, Diers B, Donald P, et al. A standard greenhouse method for assessing soybean cyst nematode resistance in soybean: SCE08 (standardized cyst evaluation 2008). Plant Health Progress. 2009;10(1):33.
211. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2010;463(7278):178.
212. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics. 2011;27(17):2325-9.
213. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biology. 2011;12(3):R22.
214. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology. 2012;31:46.
215. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology. 2010;28:511.
216. Morales AMAP, O'Rourke JA, van de Mortel M, Scheider KT, Bancroft TJ, Borém A, et al. Transcriptome analyses and virus induced gene silencing identify genes in the Rpp4-mediated Asian soybean rust resistance pathway. Functional Plant Biology. 2013;40(10):1029-47.
217. Jin J, He K, Tang X, Li Z, Lv L, Zhao Y, et al. An Arabidopsis Transcriptional Regulatory Map Reveals Distinct Functional and Evolutionary Features of Novel Transcription Factors. Mol Biol Evol. 2015;32(7):1767-73.
218. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics. 2013;29(14):1830-1.
219. Luo W, Pant G, Bhavnasi YK, Blanchard SG, Jr, Brouwer C. Pathview Web: user friendly pathway visualization and data integration. Nucleic Acids Research. 2017;45(W1):W501-W8.
220. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature biotechnology. 2011;29(1):24.
221. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. methods. 2001;25(4):402-8.
222. Food and Agriculture Organization of the United Nations (FAO) [Internet]. 2018. Available from: http://www.fao.org/faostat/en/#data/QC.
223. Allen TW, Bradley CA, Sisson AJ, Byamukama E, Chilvers MI, Coker CM, et al. Soybean Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada, from 2010 to 2014. Plant Health Progress. 2017;18(1):19-27.
224. Chaves MM, Flexas J, Pinheiro C. Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. Annals of botany. 2009;103(4):551-60.

225. Prasch CM, Sonnewald U. Simultaneous Application of Heat, Drought, and Virus to Arabidopsis Plants Reveals Significant Shifts in Signaling Networks. Plant Physiology. 2013;162(4):1849-66.
226. Zandalinas SI, Balfagón D, Arbona V, Gómez-Cadenas A. Modulation of Antioxidant Defense System Is Associated with Combined Drought and Heat Stress Tolerance in Citrus. Frontiers in Plant Science. 2017;8:953.
227. Suzuki N, Rivero RM, Shulaev V, Blumwald E, Mittler R. Abiotic and biotic stress combinations. New Phytol. 2014;203(1):32-43.
228. Anderson JP, Badruzsaufari E, Schenk PM, Manners JM, Desmond OJ, Ehlert C, et al. Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in Arabidopsis. The Plant Cell. 2004;16(12):3460-79.
229. Ramegowda V, Senthil-Kumar M. The interactive effects of simultaneous biotic and abiotic stresses on plants: mechanistic understanding from drought and pathogen combination. Journal of plant physiology. 2015;176:47-54.
230. Lindner M, Maroschek M, Netherer S, Kremer A, Barbati A, Garcia-Gonzalo J, et al. Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. Forest ecology and management. 2010;259(4):698-709.
231. Shaheen T, Riaz MS, Zafar Y. Soybean production and drought stress. Abiotic and Biotic Stresses in Soybean Production: Elsevier; 2016. p. 177-96.
232. Kunert KJ, Vorster BJ, Fenta BA, Kibido T, Dionisio G, Foyer CH. Drought Stress Responses in Soybean Roots and Nodules. Frontiers in Plant Science. 2016;7(1015).
233. El Sabagh A, Hossain A, Islam MS, barutçular C, Fahad S, Ratnasekera D, et al. ROLE OF OSMOPROTECTANTS AND SOIL AMENDMENTS FOR SUSTAINABLE SOYBEAN (Glycine max L.) PRODUCTION UNDER DROUGHT CONDITION: A REVIEW. Journal of Experimental Biology and Agricultural Sciences. 2018;6:32-41.
234. Frederick JR, Camp CR, Bauer PJ. Drought-stress effects on branch and mainstem seed yield and yield components of determinate soybean. Crop Sci. 2001;41(3):759-63.
235. He J, Du Y-L, Wang T, Turner NC, Yang R-P, Jin Y, et al. Conserved water use improves the yield performance of soybean (Glycine max (L.) Merr.) under drought. Agricultural Water Management. 2017;179:236-45.
236. Sadeghipour O, Abbasi S. Soybean response to drought and seed inoculation. World Appl Sci J. 2012;17(1):55-60.
237. Zipper SC, Qiu J, Kucharik CJ. Drought effects on US maize and soybean production: spatiotemporal patterns and historical changes. Environmental Research Letters. 2016;11(9):094021.
238. Mertz-Henning ML, Ferreira CL, Henning AF, Mandarino MGJ, Santos DE, Oliveira CNDM, et al. Effect of Water Deficit-Induced at Vegetative and Reproductive Stages on Protein and Oil Content in Soybean Grains. Agronomy. 2018;8(1).
239. Riggs R. Worldwide distribution of soybean-cyst nematode and its economic importance. Journal of Nematology. 1977;9(1):34.
240. Lian Y, Guo J, Li H, Wu Y, Wei H, Wang J, et al. A New Race (X12) of Soybean Cyst Nematode in China. Journal of nematology. 2017;49(3):321-6.

241. Yan G, Baidoo R. Current Research Status of Heterodera glycines Resistance and Its Implication on Soybean Breeding. Engineering-Prc. 2018;4(4):534-41.
242. Mittler R. Abiotic stress, the field environment and stress combination. Trends in plant science. 2006;11(1):15-9.
243. Rizhsky L, Liang H, Shuman J, Shulaev V, Davletova S, Mittler R. When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. Plant physiology. 2004;134(4):1683-96.
244. Ozturk ZN, Talamé V, Deyholos M, Michalowski CB, Galbraith DW, Gozukirmizi N, et al. Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Molecular Biology. 2002;48(5):551-73.
245. Kreps JA, Wu Y, Chang H-S, Zhu T, Wang X, Harper JF. Transcriptome Changes for Arabidopsis in Response to Salt, Osmotic, and Cold Stress. Plant Physiology. 2002;130(4):2129.
246. Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K, et al. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Current Opinion in Plant Biology. 2006;9(4):436-42.
247. Atkinson NJ, Urwin PE. The interaction of plant biotic and abiotic stresses: from genes to the field. Journal of experimental botany. 2012;63(10):3523-43.
248. Yalpani N, Enyedi AJ, León J, Raskin I. Ultraviolet light and ozone stimulate accumulation of salicylic acid, pathogenesis-related proteins and virus resistance in tobacco. Planta. 1994;193(3):372-6.
249. Sharma YK, León J, Raskin I, Davis KR. Ozone-induced responses in Arabidopsis thaliana: the role of salicylic acid in the accumulation of defense-related transcripts and induced resistance. Proceedings of the National Academy of Sciences. 1996;93(10):5099-104.
250. Bowler C, Fluhr R. The role of calcium and activated oxygens as signals for controlling cross-tolerance. Trends in Plant Science. 2000;5(6):241-6.
251. Sandermann H. Molecular ecotoxicology: from man-made pollutants to multiple environmental stresses. Molecular ecotoxicology of plants: Springer; 2004. p. 1-16.
252. Grodzki W, McManus M, Knížek M, Meshkova V, Mihalciuc V, Novotny J, et al. Occurrence of spruce bark beetles in forest stands at different levels of air pollution stress. Environmental Pollution. 2004;130(1):73-83.
253. Atkinson NJ, Lilley CJ, Urwin PE. Identification of genes involved in the response of Arabidopsis to simultaneous biotic and abiotic stresses. Plant physiology. 2013;162(4):2028-41.
254. Kissoudis C, van de Wiel C, Visser RG, van der Linden G. Enhancing crop resilience to combined abiotic and biotic stress through the dissection of physiological and molecular crosstalk. Frontiers in Plant Science. 2014;5:207.
255. Dangi AK, Sharma B, Khangwal I, Shukla P. Combinatorial interactions of biotic and abiotic stresses in plants and their molecular mechanisms: systems biology approach. Molecular biotechnology. 2018;60(8):636-50.
256. Mittler R, Blumwald E. Genetic engineering for modern agriculture: challenges and perspectives. Annual review of plant biology. 2010;61:443-62.

257. Pandey P, Irulappan V, Bagavathiannan MV, Senthil-Kumar M. Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. Frontiers in plant science. 2017;8:537.
258. Nakashima K, Yamaguchi-Shinozaki K. ABA signaling in stress-response and seed development. Plant Cell Reports. 2013;32(7):959-70.
259. Planas-Riverola A, Gupta A, Betegón-Putze I, Bosch N, Ibañes M, Caño-Delgado AI. Brassinosteroid signaling in plant development and adaptation to stress. Development. 2019;146(5):dev151894.
260. Eckardt NA. GA Signaling: Direct Targets of DELLA Proteins. The Plant Cell. 2007;19(10):2970-.
261. De Vleesschauwer D, Seifi HS, Filipe O, Haeck A, Huu SN, Demeestere K, et al. The DELLA protein SLR1 integrates and amplifies salicylic acid-and jasmonic acid-dependent innate immunity in rice. Plant physiology. 2016;170(3):1831-47.
262. Claeys H, Skirycz A, Maleux K, Inzé D. DELLA signaling mediates stress-induced cell differentiation in Arabidopsis leaves through modulation of anaphase-promoting complex/cyclosome activity. Plant physiology. 2012;159(2):739-47.
263. Wang KL-C, Li H, Ecker JR. Ethylene Biosynthesis and Signaling Networks. The Plant Cell. 2002;14(suppl 1):S131-S51.
264. Chen D, Shao Q, Yin L, Younis A, Zheng B. Polyamine Function in Plants: Metabolism, Regulation on Development, and Roles in Abiotic Stress Responses. Frontiers in Plant Science. 2019;9(1945).
265. Vemanna RS, Bakade R, Bharti P, Kumar MKP, Sreeman SM, Senthil-Kumar M, et al. Cross-Talk Signaling in Rice During Combined Drought and Bacterial Blight Stress. Frontiers in Plant Science. 2019;10(193).
266. Roy S. Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. Plant signaling & behavior. 2016;11(1):e1117723-e.
267. Shi W-Y, Du Y-T, Ma J, Min D-H, Jin L-G, Chen J, et al. The WRKY Transcription Factor GmWRKY12 Confers Drought and Salt Tolerance in Soybean. International journal of molecular sciences. 2018;19(12):4087.
268. Yang Y, Zhou Y, Chi Y, Fan B, Chen Z. Characterization of Soybean WRKY Gene Family and Identification of Soybean WRKY Genes that Promote Resistance to Soybean Cyst Nematode. Scientific Reports. 2017;7(1):17804.
269. Conrath U, Pieterse CMJ, Mauch-Mani B. Priming in plant–pathogen interactions. Trends in Plant Science. 2002;7(5):210-6.
270. Hayward AP, Dinesh-Kumar SP. What Can Plant Autophagy Do for an Innate Immune Response? Annual Review of Phytopathology. 2011;49(1):557-76.
271. Tang J, Bassham DC. Autophagy in crop plants: what's new beyond Arabidopsis? Open Biol. 2018;8(12):180162.
272. Specht J, Hume D, Kumudini S. Soybean yield potential—a genetic and physiological perspective. Crop Sci. 1999;39(6):1560-70.
273. Manavalan LP, Guttikonda SK, Phan Tran L-S, Nguyen HT. Physiological and Molecular Approaches to Improve Drought Resistance in Soybean. Plant and Cell Physiology. 2009;50(7):1260-76.

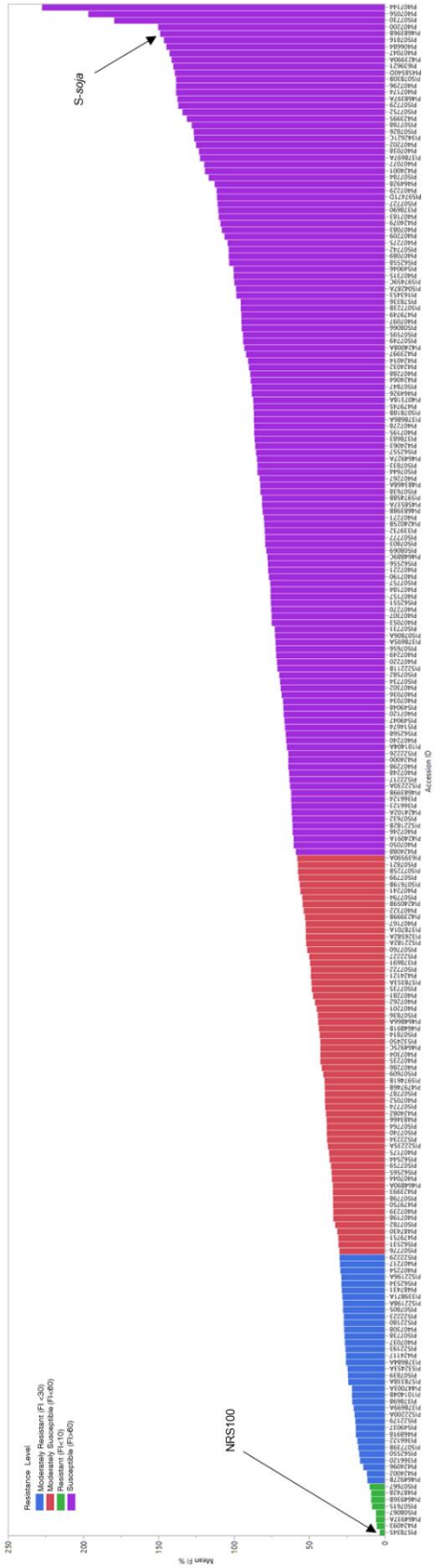
274. Smit AL, Vamerali T. The influence of potato cyst nematodes (Globodera pallida) and drought on rooting dynamics of potato (Solanum tuberosum L.). European Journal of Agronomy. 1998;9(2):137-46.
275. Audebert A, Coyne DL, Dingkuhn M, Plowright RA. The influence of cyst nematodes (Heterodera sacchari) and drought on water relations and growth of upland rice in Côte d'Ivoire. Plant and Soil. 2000;220(1):235-42.
276. Goel AK, Lundberg D, Torres MA, Matthews R, Akimoto-Tomiyama C, Farmer L, et al. The Pseudomonas syringae type III effector HopAM1 enhances virulence on water-stressed plants. Molecular Plant-Microbe Interactions. 2008;21(3):361-70.
277. Amtmann A, Troufflard S, Armengaud P. The effect of potassium nutrition on pest and disease resistance in plants. Physiologia plantarum. 2008;133(4):682-91.
278. English-Loeb GM. Plant drought stress and outbreaks of spider mites: a field test. Ecology. 1990;71(4):1401-11.
279. Young L, Heatherly L. Soybean cyst nematode effect on soybean grown at controlled soil water potentials. Crop Sci. 1988;28(3):543-5.
280. Heatherly LG, Young L, Epps J, Hartwig E. Effect of Upper-Profile Soil Water Potential on Numbers of Cysts of Heterodera glycines on Soybeans 1. Crop Sci. 1982;22(4):833-5.
281. Vishwakarma K, Upadhyay N, Kumar N, Yadav G, Singh J, Mishra RK, et al. Abscisic Acid Signaling and Abiotic Stress Tolerance in Plants: A Review on Current Knowledge and Future Prospects. Frontiers in plant science. 2017;8:161-.
282. Hu Y, Bao F, Li J. Promotive effect of brassinosteroids on cell division involves a distinct CycD3-induction pathway in Arabidopsis. The Plant Journal. 2000;24(5):693-701.
283. Mittler R. Oxidative stress, antioxidants and stress tolerance. Trends in Plant Science. 2002;7(9):405-10.
284. Polidoros AN, Mylona PV, Arnholdt-Schmitt B. Aox gene structure, transcript variation and expression in plants. Physiologia Plantarum. 2009;137(4):342-53.
285. Fraire-Velázquez S, Rodríguez-Guerra R, Sánchez-Calderón L. Abiotic and biotic stress response crosstalk in plants. Abiotic Stress Response in Plants—Physiological, Biochemical and Genetic Perspectives. 2011:346.
286. Choi H-K, Iandolino A, da Silva FG, Cook DR. Water Deficit Modulates the Response of Vitis vinifera to the Pierce's Disease Pathogen Xylella fastidiosa. Molecular Plant-Microbe Interactions. 2013;26(6):643-57.
287. Gupta A, Sarkar AK, Senthil-Kumar M. Global transcriptional analysis reveals unique and shared responses in Arabidopsis thaliana exposed to combined drought and pathogen stress. Frontiers in plant science. 2016;7:686.
288. Fan C, Yao H, Qiu Z, Ma H, Zeng B. Genome-wide analysis of Eucalyptus grandis WRKY genes family and their expression profiling in response to hormone and abiotic stress treatment. Gene. 2018;678:38-48.
289. Rushton PJ, Somssich IE, Ringler P, Shen QJ. WRKY transcription factors. Trends in plant science. 2010;15(5):247-58.
290. Pandey SP, Somssich IE. The Role of WRKY Transcription Factors in Plant Immunity. Plant Physiology. 2009;150(4):1648.
291. Jiang J, Ma S, Ye N, Jiang M, Cao J, Zhang J. WRKY transcription factors in plant responses to stresses. Journal of integrative plant biology. 2017;59(2):86-101.

292. Jain S, Chittem K, Brueggeman R, Osorno JM, Richards J, Nelson BD, Jr. Comparative Transcriptome Analysis of Resistant and Susceptible Common Bean Genotypes in Response to Soybean Cyst Nematode Infection. PLOS ONE. 2016;11(7):e0159338.
293. Han Y, Zhao X, Cao G, Wang Y, Li Y, Liu D, et al. Genetic characteristics of soybean resistance to HG type 0 and HG type 1.2.3.5.7 of the cyst nematode analyzed by genome-wide association mapping. Bmc Genomics. 2015;16(1):598.
294. Neupane S, Purintun J, Mathew F, Varenhorst A, Nepal M. Molecular Basis of Soybean Resistance to Soybean Aphids and Soybean Cyst Nematodes. Plants. 2019;8.
295. Bai Y, Sunarti S, Kissoudis C, Visser RGF, van der Linden CG. The Role of Tomato WRKY Genes in Plant Responses to Combined Abiotic and Biotic Stresses. Frontiers in plant science. 2018;9:801-.
296. Jiang Y, Duan Y, Yin J, Ye S, Zhu J, Zhang F, et al. Genome-wide identification and characterization of the Populus WRKY transcription factor family and analysis of their expression in response to biotic and abiotic stresses. Journal of experimental botany. 2014;65(22):6629-44.
297. Qiu Y, Yu D. Over-expression of the stress-induced OsWRKY45 enhances disease resistance and drought tolerance in Arabidopsis. Environmental and experimental botany. 2009;65(1):35-47.
298. Xing D-H, Lai Z-B, Zheng Z-Y, Vinod K, Fan B-F, Chen Z-X. Stress-and pathogen-induced Arabidopsis WRKY48 is a transcriptional activator that represses plant basal defense. Molecular plant. 2008;1(3):459-70.
299. Dombrecht B, Xue GP, Sprague SJ, Kirkegaard JA, Ross JJ, Reid JB, et al. MYC2 differentially modulates diverse jasmonate-dependent functions in Arabidopsis. The Plant cell. 2007;19(7):2225-45.
300. Laurie-Berry N, Joardar V, Street IH, Kunkel BN. The Arabidopsis thaliana JASMONATE INSENSITIVE 1 gene is required for suppression of salicylic acid-dependent defenses during infection by Pseudomonas syringae. Molecular Plant-Microbe Interactions. 2006;19(7):789-800.
301. Wild M, Davière J-M, Cheminant S, Regnault T, Baumberger N, Heintz D, et al. The Arabidopsis DELLA RGA-LIKE3 is a direct target of MYC2 and modulates jasmonate signaling responses. The Plant Cell. 2012;24(8):3307-19.
302. Kazan K, Manners JM. MYC2: The Master in Action. Molecular Plant. 2013;6(3):686-703.
303. Choudhury FK, Rivero RM, Blumwald E, Mittler R. Reactive oxygen species, abiotic stress and stress combination. The Plant Journal. 2017;90(5):856-67.
304. Ramegowda V, Senthil-Kumar M, Ishiga Y, Kaundal A, Udayakumar M, Mysore K. Drought stress acclimation imparts tolerance to Sclerotinia sclerotiorum and Pseudomonas syringae in Nicotiana benthamiana. International journal of molecular sciences. 2013;14(5):9497-513.
305. Després C, Chubak C, Rochon A, Clark R, Bethune T, Desveaux D, et al. The Arabidopsis NPR1 disease resistance protein is a novel cofactor that confers redox regulation of DNA binding activity to the basic domain/leucine zipper transcription factor TGA1. The Plant Cell. 2003;15(9):2181-91.

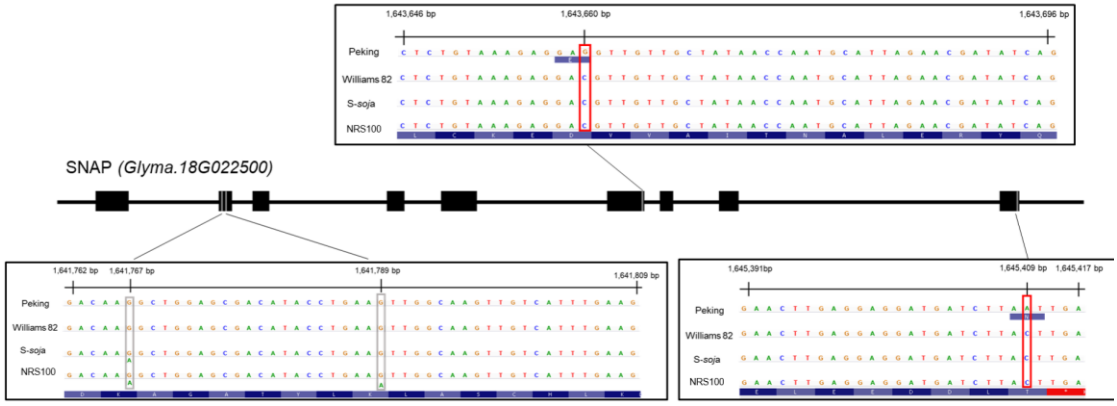
306. Polidoros AN, Mylona PV, Pasentsis K, Scandalios JG, Tsaftaris AS. The maize alternative oxidase 1a (Aox1a) gene is regulated by signals related to oxidative stress. Redox Report. 2005;10(2):71-8.
307. Signorelli S, Tarkowski ŁP, Van den Ende W, Bassham DC. Linking autophagy to abiotic and biotic stress responses. Trends in plant science. 2019.
308. Han S, Yu B, Wang Y, Liu Y. Role of plant autophagy in stress response. Protein & Cell. 2011;2(10):784-91.
309. Avin-Wittenberg T. Autophagy and its role in plant abiotic stress management. Plant, Cell & Environment. 2019;42(3):1045-53.
310. Yoshimoto K, Jikumaru Y, Kamiya Y, Kusano M, Consonni C, Panstruga R, et al. Autophagy Negatively Regulates Cell Death by Controlling NPR1-Dependent Salicylic Acid Signaling during Senescence and the Innate Immune Response in Arabidopsis. The Plant Cell. 2009;21(9):2914-27.
311. Wu Y, Zhang D, Chu Jee Y, Boyle P, Wang Y, Brindle Ian D, et al. The Arabidopsis NPR1 Protein Is a Receptor for the Plant Defense Hormone Salicylic Acid. Cell Reports. 2012;1(6):639-47.
312. Manohar M, Wang D, Manosalva PM, Choi HW, Kombrink E, Klessig DF. Members of the abscisic acid co-receptor PP2C protein family mediate salicylic acid–abscisic acid crosstalk. Plant Direct. 2017;1(5):e00020.
313. Berens ML, Wolinska KW, Spaepen S, Ziegler J, Nobori T, Nair A, et al. Balancing trade-offs between biotic and abiotic stress responses through leaf age-dependent variation in stress hormone cross-talk. Proceedings of the National Academy of Sciences. 2019;116(6):2364-73.
314. İşkil R, Surgun-Acar Y. Expression analysis of cell wall assembly and remodelling-related genes in Arabidopsis roots subjected to boron stress and brassinosteroid at different developmental stages. Acta Botanica Brasilica. 2018;32(4):546-54.
315. Rao X, Dixon RA. Brassinosteroid mediated cell wall remodeling in grasses under abiotic stress. Frontiers in Plant Science. 2017;8:806.
316. Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, et al. The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses. Genes Dev. 1997;11(23):3194-205.
317. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES. Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet. 2001;28(3):286-9.
318. Matsuura H, Ishibashi Y, Shinmyo A, Kanaya S, Kato K. Genome-Wide Analyses of Early Translational Responses to Elevated Temperature and High Salinity in Arabidopsis thaliana. Plant and Cell Physiology. 2010;51(3):448-62.
319. Merret R, Nagarajan VK, Carpentier M-C, Park S, Favory J-J, Descombin J, et al. Heat-induced ribosome pausing triggers mRNA co-translational decay in Arabidopsis thaliana. Nucleic acids research. 2015;43(8):4121-32.
320. Merchante C, Stepanova AN, Alonso JM. Translation regulation in plants: an interesting past, an exciting present and a promising future. The Plant Journal. 2017;90(4):628-53.
321. Matsuura H, Takenami S, Kubo Y, Ueda K, Ueda A, Yamaguchi M, et al. A computational and experimental approach reveals that the 5′-proximal region of the 5′-

UTR has a Cis-regulatory signature responsible for heat stress-regulated mRNA translation in Arabidopsis. Plant and cell physiology. 2013;54(4):474-83.
322. Wang L, Li H, Zhao C, Li S, Kong L, Wu W, et al. The inhibition of protein translation mediated by AtGCN1 is essential for cold tolerance in Arabidopsis thaliana. Plant, cell & environment. 2017;40(1):56-68.
323. Lykke-Andersen J, Bennett EJ. Protecting the proteome: eukaryotic cotranslational quality control pathways. J Cell Biol. 2014;204(4):467-76.
324. Hussain SS, Ali M, Ahmad M, Siddique KHM. Polyamines: Natural and engineered abiotic and biotic stress tolerance in plants. Biotechnology Advances. 2011;29(3):300-11.
325. Yamaguchi K, Takahashi Y, Berberich T, Imai A, Miyazaki A, Takahashi T, et al. The polyamine spermine protects against high salt stress in Arabidopsis thaliana. FEBS letters. 2006;580(30):6783-8.
326. Yamaguchi K, Takahashi Y, Berberich T, Imai A, Takahashi T, Michael AJ, et al. A protective role for the polyamine spermine against drought stress in Arabidopsis. Biochemical and biophysical research communications. 2007;352(2):486-90.
327. Xu C, Wu X, Zhang H. Impact of D-Arg on drought resistance and endogenous polyamines in mycorrhizal Pinus massoniana. Journal of Nanjing Forestry University (Natural Sciences Edition). 2009;33(4):19-23.
328. Montesinos-Pereira D, Barrameda-Medina Y, Romero L, Ruiz J, Sánchez-Rodríguez E. Genotype differences in the metabolism of proline and polyamines under moderate drought in tomato plants. Plant Biology. 2014;16(6):1050-7.
329. LIU Y-c, ZUO Z-w, HU J-j. Effects of Exogenous Polyamines on Growth and Drought Resistance of Apple Seedlings [J]. Journal of Northwest Forestry University. 2010;1.
330. Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, et al. Heatmapper: web-enabled heat mapping for all. Nucleic acids research. 2016;44(W1):W147-W53.
331. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--new capabilities and interfaces. Nucleic acids research. 2012;40(15):e115-e.

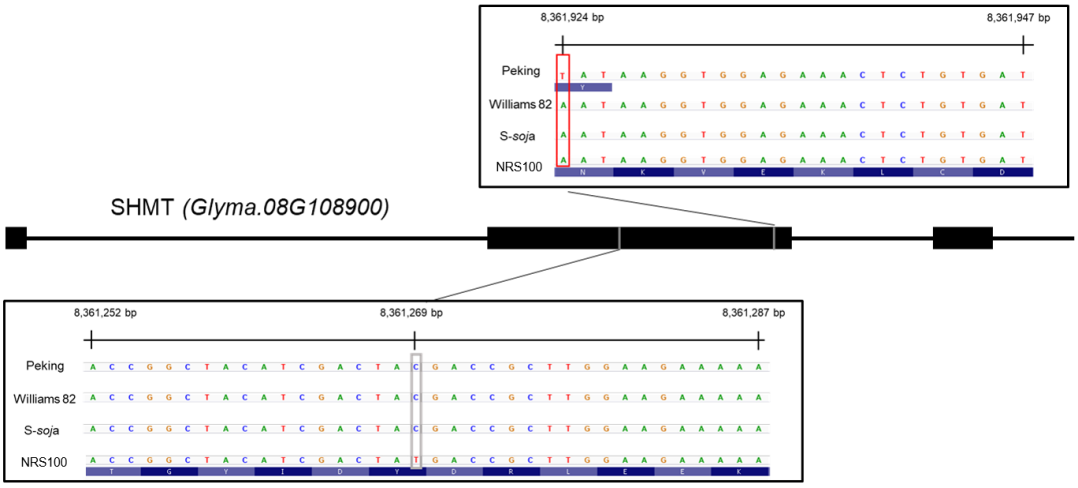
# APPENDIX A



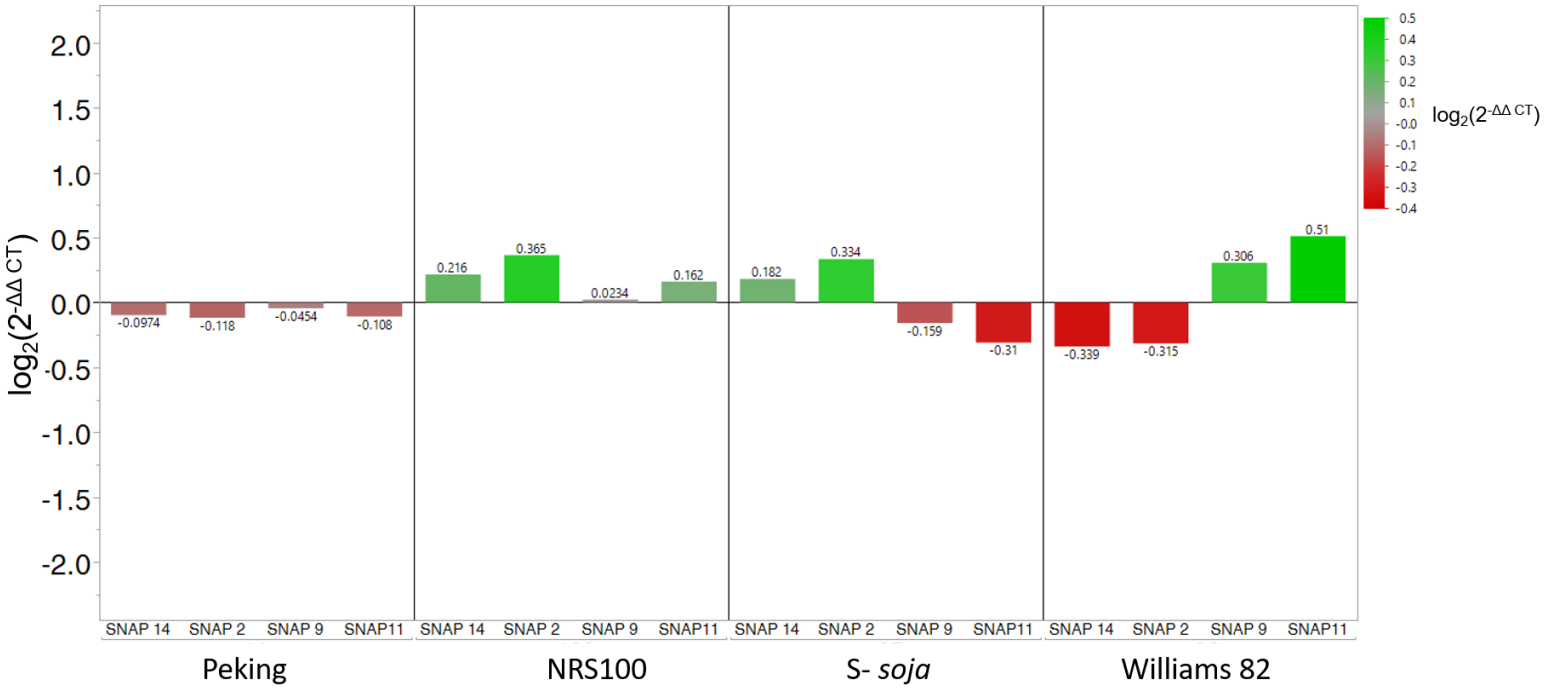

**Supplementary Figure 1.** Graph of all accessions screened for SCN resistance with resulting FI%. NRS100 is the most resistant of all screened accessions.



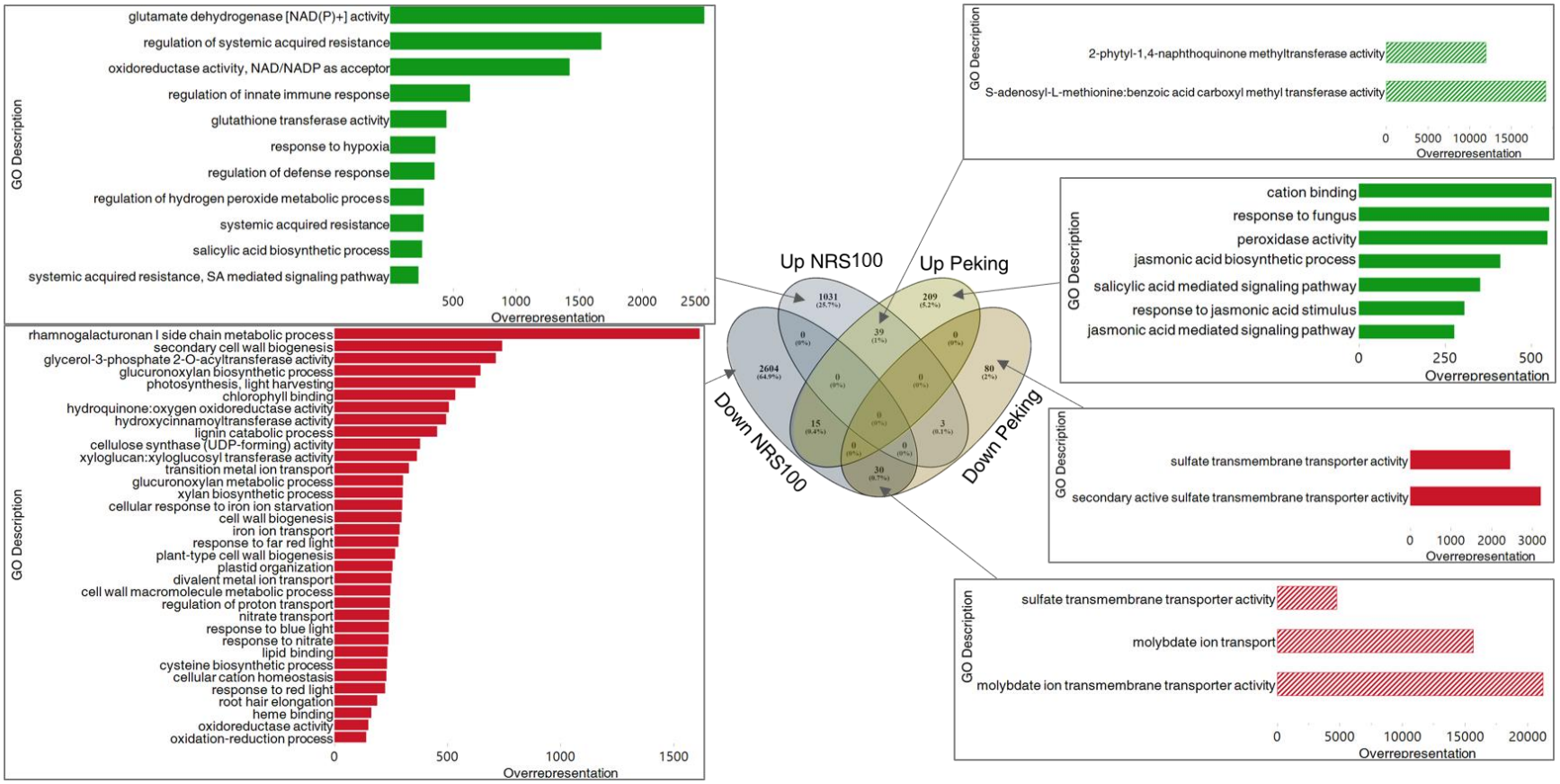

**Supplementary Figure 2.** *Rhg1* gene SNAP *(Glyma.18G022500)* comparison of nucleotide sequence and amino acid sequence between Peking, Williams 82, S-*soja*, and NRS100. Variations are highlighted in gray. Synonymous variations are outlined in gray. Non-synonymous variations are outlined in red.



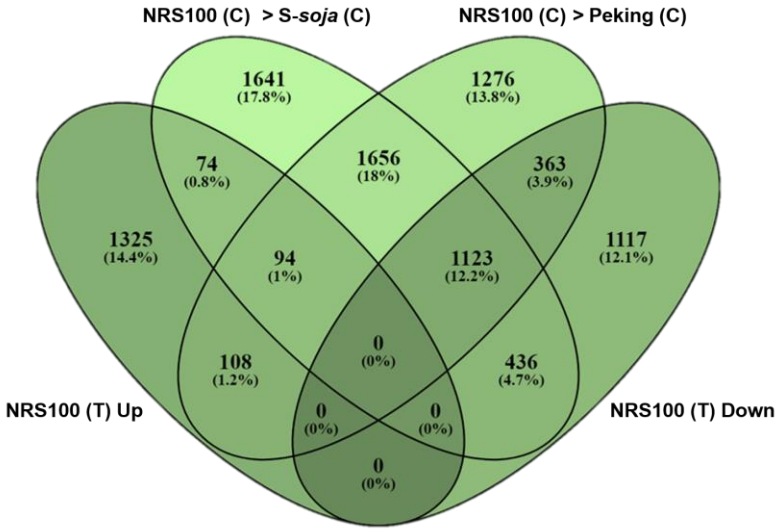

**Supplementary Figure 3.** *Rhg4* gene SHMT *(Glyma.08G108900)* comparison of nucleotide sequence and amino acid sequence between Peking, Williams 82, S-soja, and NRS100. Variations are highlighted in gray. Synonymous variations are outlined in gray. Non-synonymous variations are outlined in red.



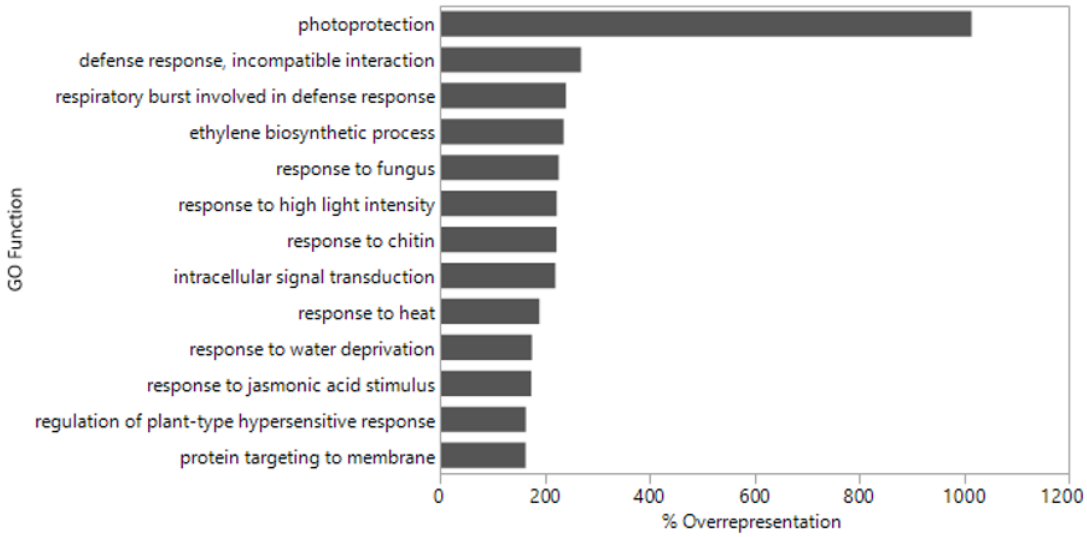

**Supplementary Figure 4.** Relative fold change of grouped replicates calculated by $2^{-\Delta\Delta CT}$ method of SNAP genes. No significant fold change found.



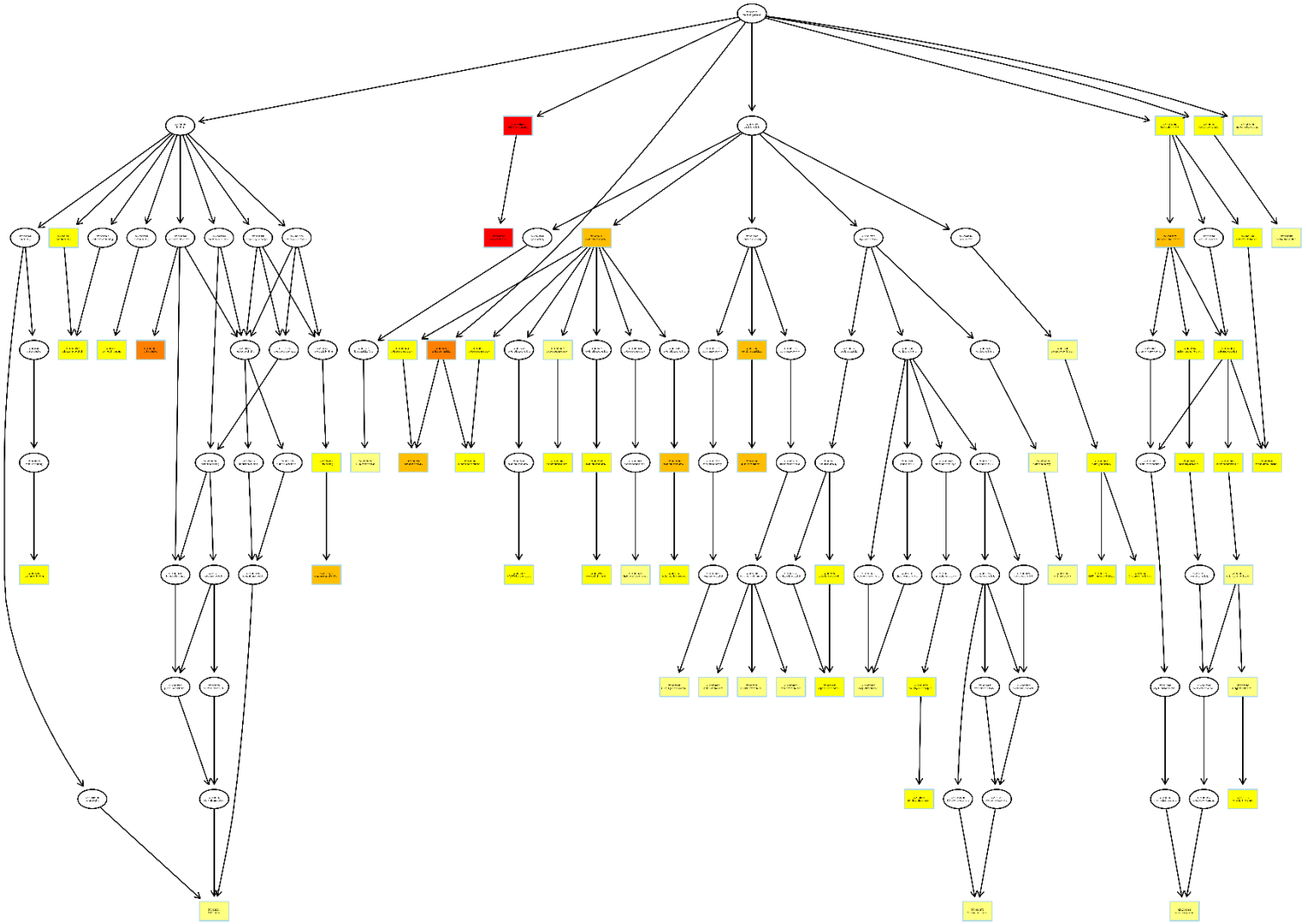

**Supplementary Figure 5.** Overrepresented GO functions for NRS100 and Peking resistance specific genes. Significance cut-off at Bonferroni adjusted p-value <0.01



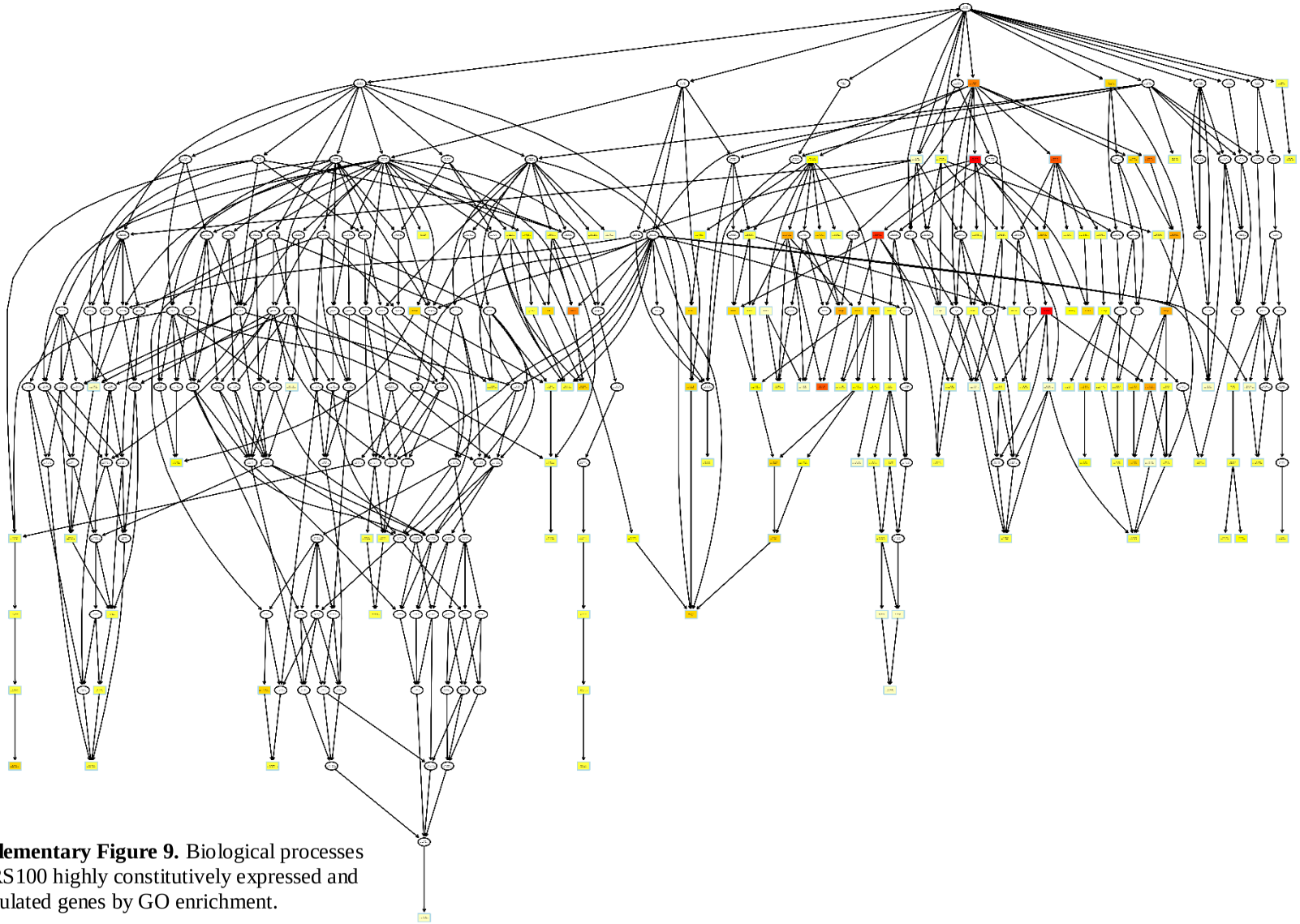

**Supplementary Figure 6.** Significantly highly expressed genes in NRS100 control expression profile (NRS100 (C)), compared to S-*soja* control expression profile (S-soja (C)) and to Peking control expression profile (Peking (C)), q-value<0.01. NRS100 SCN induced expression profiles, both up (NRS100 (T) Up) and down (NRS100 (T) Down), compared to the NRS-specific control expressed genes.



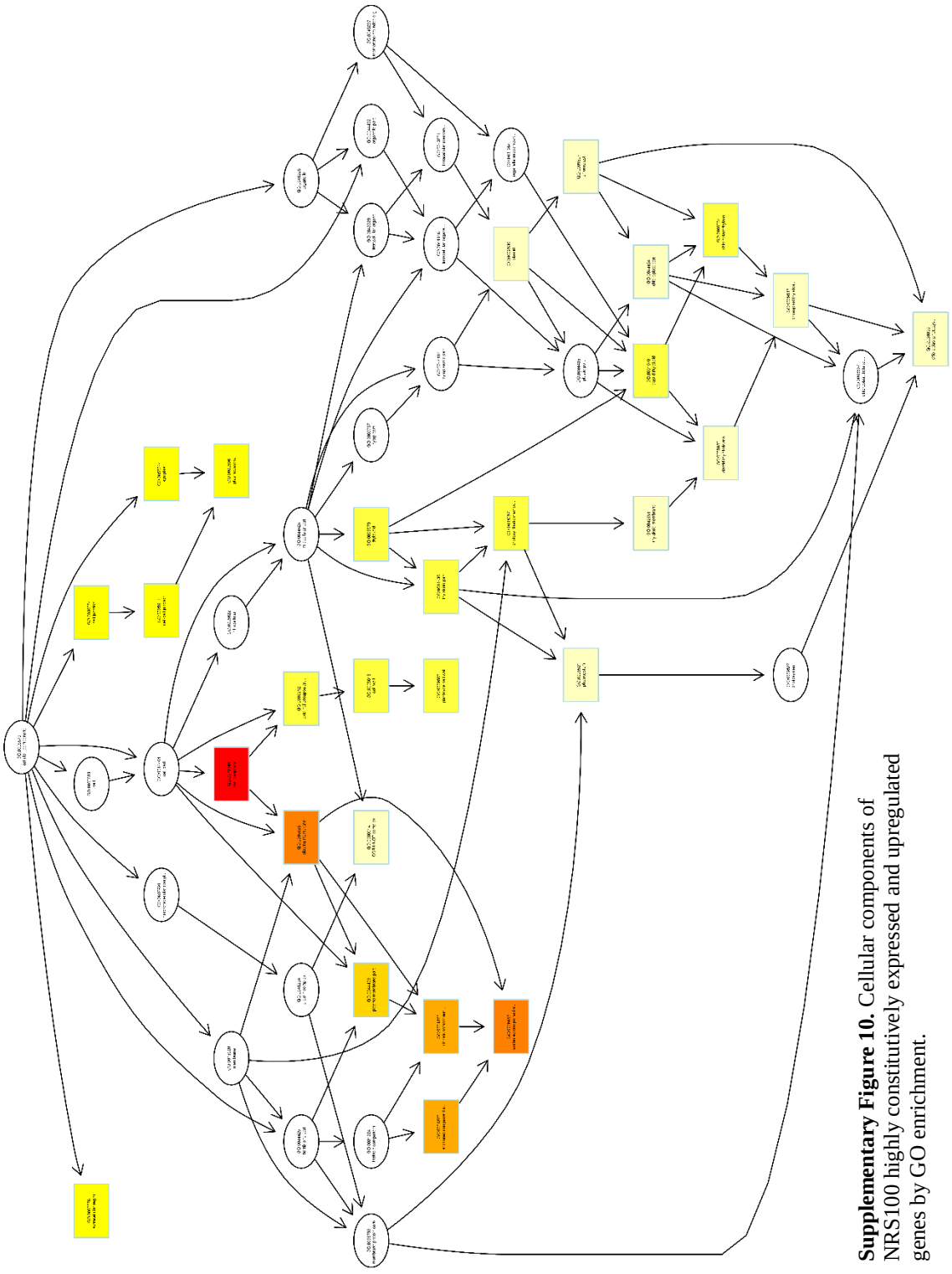

**Supplementary Figure 7.** Significantly enriched GO functions (p<0.01) of the 1656 NRS-specific significantly constitutively expressed genes.



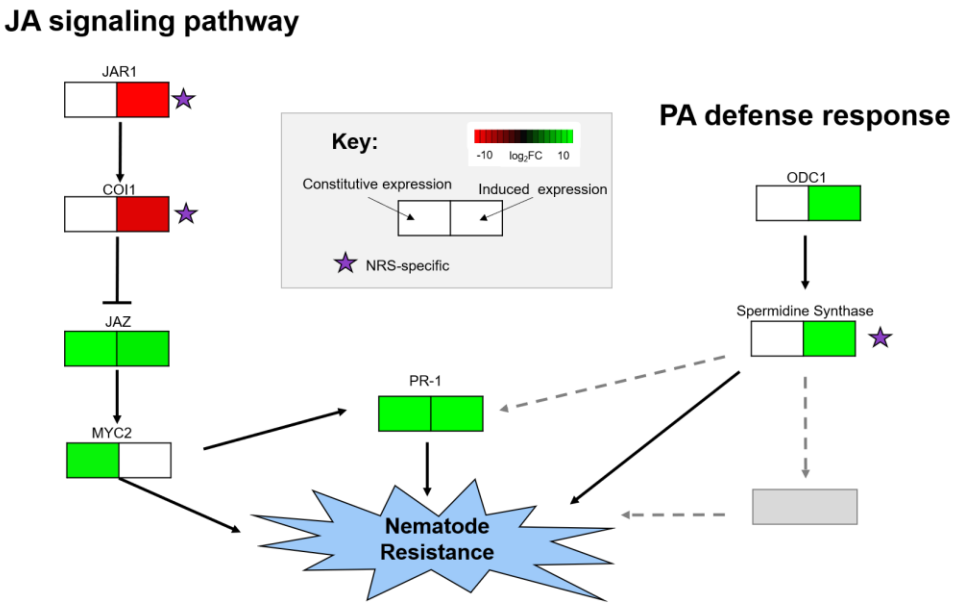
**Supplementary Figure 8.** Molecular functions of NRS100 highly constitutively expressed and upregulated genes by GO enrichment.



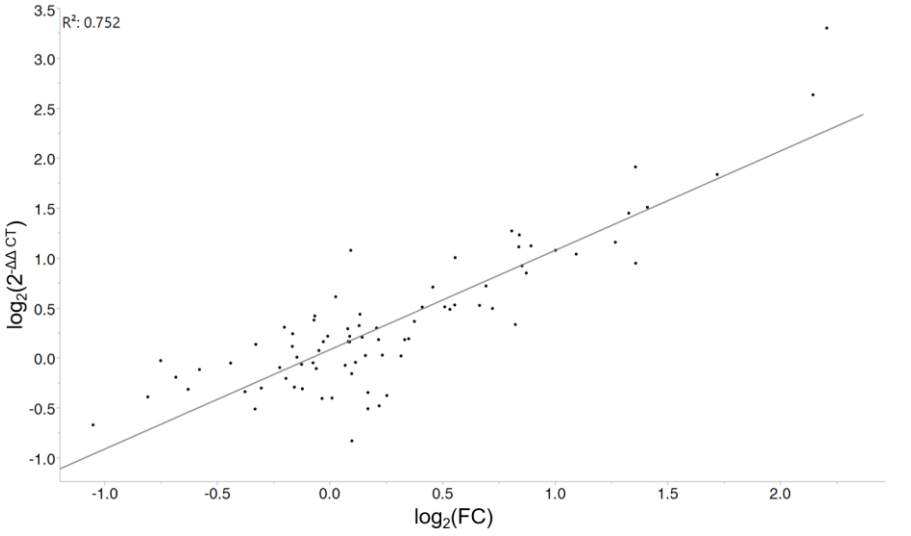
**Supplementary Figure 9.** Biological processes of NRS100 highly constitutively expressed and upregulated genes by GO enrichment.



**Supplementary Figure 10.** Cellular components of NRS100 highly constitutively expressed and upregulated genes by GO enrichment.




**Supplementary Figure 11.** Heatmap/Pathway of DEGs by $log_2$FC (fold change) of constitute expression relative to other genotypes on left of box and induced response to SCN treatment on right of box. NRS100 specific regulation is shown with purple stars. Adapted from Pathveiw output.




**Supplementary Figure 12.** Positive correlation ($R^2$=0.752) between qPCR fold change expression values ($log_2(2^{-\Delta\Delta CT})$) and correlated RNA-sequencing fold change expression values ($log_2$(FC)) where FC= Fold change, for 20 genes (supplementary table 6). Individual points represent relative fold change of grouped replicates calculated by $2^{-\Delta\Delta CT}$ method and calculated fold change of group replicates from RNA-sequencing analysis.

**Supplementary Table 1.** RNA sequencing results before and after quality filter for NRS, Peking, S-*soja*, and Williams 82 in control (C) and treatment (T) conditions.

| Sample ID | Total raw reads | Total cleaned reads | Quality checked read rate | Reads mapped | Multiple-alignment rate | Overall Mapping Rate |
|---|---|---|---|---|---|---|
| NRS100_C1 | 21,063,908 | 19,761,419 | 93.82% | 17,649,075 | 19.70% | 89.30% |
| NRS100_C2 | 17,934,885 | 16,889,827 | 94.17% | 15,429,378 | 10.30% | 91.40% |
| NRS100_C3 | 26,401,150 | 24,786,549 | 93.88% | 22,527,878 | 16.70% | 90.90% |
| NRS100_T1 | 21,400,799 | 20,131,429 | 94.07% | 17,325,667 | 13.20% | 86.10% |
| NRS100_T2 | 21,168,967 | 19,920,214 | 94.10% | 17,478,920 | 11.50% | 87.70% |
| NRS100_T3 | 24,977,825 | 23,508,971 | 94.12% | 20,126,490 | 11.30% | 85.60% |
| S-*soja* _C1 | 21,713,908 | 20,418,540 | 94.03% | 18,734,942 | 13.20% | 91.80% |
| S-*soja* _C2 | 22,480,424 | 21,140,634 | 94.04% | 19,371,545 | 13.20% | 91.60% |
| S-*soja* _C3 | 18,217,521 | 17,128,116 | 94.02% | 15,687,553 | 13.50% | 91.60% |
| S-*soja* _T1 | 22,025,661 | 20,699,732 | 93.98% | 17,575,720 | 12.90% | 84.90% |
| S-*soja* _T2 | 18,306,577 | 17,213,674 | 94.03% | 14,773,707 | 10.60% | 85.80% |
| S-*soja* _T3 | 20,541,015 | 19,341,615 | 94.16% | 16,819,076 | 9.80% | 87.00% |
| Peking_C1 | 23,521,521 | 22,138,476 | 94.12% | 20,372,740 | 10.10% | 92.00% |
| Peking_C2 | 19,348,795 | 18,197,184 | 94.05% | 16,720,114 | 11.40% | 91.90% |
| Peking_C3 | 22,428,803 | 21,091,236 | 94.04% | 19,166,712 | 11.90% | 90.90% |
| Peking_T1 | 22,536,397 | 21,195,218 | 94.05% | 19,156,398 | 10.00% | 90.40% |
| Peking_T2 | 25,234,276 | 23,745,769 | 94.10% | 21,672,021 | 9.90% | 91.30% |
| Peking_T3 | 29,687,822 | 27,941,769 | 94.12% | 25,563,489 | 10.20% | 91.50% |
| Williams_82_C1 | 22,149,804 | 20,833,495 | 94.06% | 19,333,485 | 10.10% | 92.80% |
| Williams_82_C2 | 19,200,624 | 18,034,095 | 93.92% | 16,641,849 | 13.30% | 92.30% |
| Williams_82_C3 | 24,633,007 | 23,152,320 | 93.99% | 21,419,108 | 12.80% | 92.50% |
| Williams_82_T1 | 25,363,759 | 24,056,915 | 94.85% | 21,800,389 | 9.70% | 90.60% |
| Williams_82_T2 | 27,632,370 | 26,243,565 | 94.97% | 24,041,494 | 8.70% | 91.60% |
| Williams_82_T3 | 29,277,958 | 27,806,301 | 94.97% | 25,569,291 | 9.5%) | 92.00% |

**Supplementary Table 2.** Significant upregulated genes in NRS100 overrepresented enriched GO terms.

| GO ID | GO count | Expressed | Expected_ | Corrected_P | GO_desc |
|---|---|---|---|---|---|
| GO:0009627 | 751 | 54 | 20.44458 | 6.03E-08 | systemic acquired resistance |
| GO:0031347 | 345 | 33 | 9.391982 | 2.52E-07 | regulation of defense response |
| GO:0009697 | 653 | 45 | 17.77671 | 7.64E-06 | salicylic acid biosynthetic process |
| GO:0010310 | 535 | 39 | 14.56438 | 1.60E-05 | regulation of hydrogen peroxide metabolic process |
| GO:0004353 | 8 | 5 | 0.201269 | 0.000175436 | glutamate dehydrogenase [NAD(P)+] activity |
| GO:0045088 | 64 | 11 | 1.742281 | 0.00065232 | regulation of innate immune response |
| GO:0009862 | 688 | 42 | 18.72952 | 0.000716186 | systemic acquired resistance, salicylic acid mediated signaling pathway |
| GO:0001666 | 195 | 19 | 5.308512 | 0.000962761 | response to hypoxia |
| GO:0004364 | 125 | 14 | 3.144833 | 0.001002302 | glutathione transferase activity |
| GO:0010112 | 11 | 5 | 0.299455 | 0.003770086 | regulation of systemic acquired resistance |
| GO:0016639 | 14 | 5 | 0.352221 | 0.005544378 | oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor |
| GO:0080142 | 7 | 4 | 0.190562 | 0.011562246 | regulation of salicylic acid biosynthetic process |
| GO:0009867 | 800 | 43 | 21.77851 | 0.014133788 | jasmonic acid mediated signaling pathway |
| GO:0043295 | 28 | 6 | 0.704443 | 0.019326592 | glutathione binding |
| GO:0043531 | 543 | 30 | 13.66116 | 0.023266026 | ADP binding |

**Supplementary Table 3.** Significant (Bonferroni p>.05) GO identifiers associated with results from venn diagram (Figure 3A, s10 Figure) of resistance specific genes associated with Peking and NRS100 resistance. Up and down regulated GO ID are specific do genes exclusive to up and down regulation of that genotype. Shared up and down regulated GO IDs are from shared genes in NRS100 and Peking resistance specific responses.

| GO ID | Genome total count | Expressed | Expected | Overrepresentation | p-value | Description |
|---|---|---|---|---|---|---|
| **Up Regulated in NRS** | | | | | | |
| GO:0009627 | 751 | 54 | 20.354305 | 265.3001436 | 5.16975E-08 | systemic acquired resistance |
| GO:0031347 | 345 | 33 | 9.3505129 | 352.9218175 | 2.26941E-07 | regulation of defense response |
| GO:0009697 | 653 | 45 | 17.698217 | 254.2628993 | 6.74541E-06 | salicylic acid biosynthetic process |
| GO:0010310 | 535 | 39 | 14.500071 | 268.9642059 | 1.42707E-05 | regulation of hydrogen peroxide metabolic process |
| GO:0004353 | 8 | 5 | 0.2004984 | 2493.785851 | 0.000172221 | glutamate dehydrogenase [NAD(P)+] activity |
| GO:0045088 | 64 | 11 | 1.7345879 | 634.1563908 | 0.000626476 | regulation of innate immune response |
| GO:0009862 | 688 | 42 | 18.64682 | 225.2394792 | 0.000666556 | systemic acquired resistance, salicylic acid mediated signaling pathway |
| GO:0001666 | 195 | 19 | 5.2850725 | 359.5031101 | 0.000905478 | response to hypoxia |
| GO:0004364 | 125 | 14 | 3.132787 | 446.8864245 | 0.000960171 | glutathione transferase activity |
| GO:0010112 | 11 | 5 | 0.2981323 | 1677.10781 | 0.003697018 | regulation of systemic acquired resistance |
| **Down Regulated in NRS** | | | | | | |
| GO:0009834 | 135 | 62 | 8.3544883 | 742.1160684 | 6.857E-37 | secondary cell wall biogenesis |
| GO:0010413 | 495 | 93 | 30.633124 | 303.5929371 | 2.03785E-20 | glucuronoxylan metabolic process |
| GO:0045492 | 497 | 93 | 30.756894 | 302.3712351 | 2.74313E-20 | xylan biosynthetic process |
| GO:0015706 | 486 | 73 | 30.076158 | 242.7171729 | 1.89893E-10 | nitrate transport |
| GO:0010167 | 479 | 71 | 29.642962 | 239.5172239 | 1.09083E-09 | response to nitrate |
| GO:0000041 | 201 | 41 | 12.438905 | 329.6110126 | 2.17046E-09 | transition metal ion transport |
| GO:0010106 | 237 | 44 | 14.666768 | 299.9979207 | 1.00029E-08 | cellular response to iron ion starvation |
| GO:0009832 | 313 | 52 | 19.370036 | 268.4558796 | 1.17032E-08 | plant-type cell wall biogenesis |
| GO:0019344 | 459 | 66 | 28.40526 | 232.3513307 | 1.97867E-08 | cysteine biosynthetic process |
| GO:0006826 | 241 | 43 | 14.914309 | 288.3137316 | 6.49597E-08 | iron ion transport |
| GO:0010218 | 251 | 44 | 15.53316 | 283.2649689 | 7.14464E-08 | response to far red light |
| GO:0009765 | 44 | 17 | 2.7229443 | 624.3241851 | 1.60475E-07 | photosynthesis, light harvesting |
| GO:0016168 | 58 | 19 | 3.5536248 | 534.6653331 | 1.95928E-07 | chlorophyll binding |
| GO:0052716 | 58 | 18 | 3.5536248 | 506.5250524 | 1.53896E-06 | hydroquinone:oxygen oxidoreductase activity |
| GO:0042546 | 185 | 34 | 11.448743 | 296.9758286 | 2.77389E-06 | cell wall biogenesis |
| GO:0010417 | 35 | 14 | 2.1659784 | 646.3591563 | 3.91328E-06 | glucuronoxylan biosynthetic process |
| GO:0046274 | 64 | 18 | 3.9606463 | 454.4712818 | 1.67176E-05 | lignin catabolic process |
| GO:0055114 | 2341 | 204 | 144.87302 | 140.812973 | 2.04603E-05 | oxidation-reduction process |
| GO:0010400 | 6 | 6 | 0.3713106 | 1615.897891 | 4.1089E-05 | rhamnogalacturonan I side chain metabolic process |
| GO:0070838 | 237 | 37 | 14.666768 | 252.2709787 | 5.32909E-05 | divalent metal ion transport |
| GO:0009637 | 269 | 40 | 16.647091 | 240.2822142 | 6.20507E-05 | response to blue light |
| GO:0016491 | 1468 | 135 | 89.94347 | 150.0942764 | 8.90188E-05 | oxidoreductase activity |
| GO:0009657 | 201 | 32 | 12.438905 | 257.2573756 | 0.00026004 | plastid organization |
| GO:0044036 | 222 | 34 | 13.738492 | 247.4798571 | 0.000287572 | cell wall macromolecule metabolic process |
| GO:0030003 | 274 | 39 | 16.956517 | 230.0000647 | 0.000362172 | cellular cation homeostasis |
| GO:0048767 | 504 | 59 | 31.19009 | 189.1626499 | 0.000504086 | root hair elongation |
| GO:0016762 | 76 | 17 | 4.6564739 | 365.0831152 | 0.000799047 | xyloglucan:xyloglucosyl transferase activity |
| GO:0010114 | 252 | 35 | 15.595045 | 224.4302626 | 0.002297266 | response to red light |
| GO:0010155 | 191 | 29 | 11.820054 | 245.3457531 | 0.002478003 | regulation of proton transport |
| GO:0020037 | 750 | 75 | 45.952045 | 163.213628 | 0.002891104 | heme binding |
| GO:0050734 | 33 | 10 | 2.02189 | 494.5867515 | 0.007120777 | hydroxycinnamoyltransferase activity |
| GO:0008289 | 187 | 27 | 11.457377 | 235.6560404 | 0.007172937 | lipid binding |
| GO:0016760 | 56 | 13 | 3.431086 | 378.8887793 | 0.009108529 | cellulose synthase (UDP-forming) activity |
| GO:0090447 | 16 | 7 | 0.9803103 | 714.0596225 | 0.00988302 | glycerol-3-phosphate 2-O-acyltransferase activity |
| GO:0015250 | 91 | 17 | 5.5755148 | 304.9045798 | 0.010506593 | water channel activity |
| GO:0016759 | 74 | 15 | 4.5339351 | 330.8384351 | 0.011831642 | cellulose synthase activity |
| GO:0009413 | 7 | 5 | 0.4331957 | 1154.212779 | 0.013409727 | response to flooding |
| GO:0010345 | 21 | 8 | 1.2995871 | 615.5801489 | 0.014256745 | suberin biosynthetic process |
| GO:0016747 | 260 | 33 | 15.930042 | 207.1557586 | 0.015237587 | transferase activity, transferring acyl groups other than amino-acyl groups |
| GO:0009932 | 248 | 33 | 15.347504 | 215.018671 | 0.015574335 | cell tip growth |
| GO:0008324 | 68 | 14 | 4.1663188 | 336.0280576 | 0.018531918 | cation transmembrane transporter activity |
| GO:0016706 | 298 | 36 | 18.258279 | 197.1708258 | 0.019220978 | oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors |
| GO:0019684 | 320 | 39 | 19.803231 | 196.9375554 | 0.020332362 | photosynthesis, light reaction |
| GO:0010817 | 212 | 29 | 13.119641 | 221.042636 | 0.022570572 | regulation of hormone levels |
| GO:0009416 | 440 | 49 | 27.229443 | 179.9522651 | 0.023559122 | response to light stimulus |
| GO:0009773 | 121 | 20 | 7.4880969 | 267.0905605 | 0.024652611 | photosynthetic electron transport in photosystem I |
| GO:0046688 | 29 | 9 | 1.7946679 | 501.4855523 | 0.027876403 | response to copper ion |
| GO:0006073 | 67 | 14 | 4.1463016 | 337.6503055 | 0.02816383 | cellular glucan metabolic process |
| GO:0016798 | 129 | 20 | 7.9037518 | 253.0443845 | 0.033923242 | hydrolase activity, acting on glycosyl bonds |
| GO:0046168 | 8 | 5 | 0.4950808 | 1009.936182 | 0.034028433 | glycerol-3-phosphate catabolic process |
| GO:0008361 | 144 | 22 | 8.9114542 | 246.8732889 | 0.035063373 | regulation of cell size |
| GO:0009055 | 970 | 87 | 59.431312 | 146.3874808 | 0.036500059 | electron carrier activity |
| GO:0015018 | 14 | 6 | 0.8577715 | 699.4869771 | 0.047774951 | galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase activity |
| **Up Regulated in Peking** | | | | | | |
| GO:0009620 | 310 | 17 | 3.0831421 | 551.3855439 | 5.88362E-06 | response to fungus |
| GO:0043169 | 335 | 13 | 2.3271475 | 558.6238197 | 0.000110916 | cation binding |
| GO:0009695 | 417 | 17 | 4.1473234 | 409.9029223 | 0.000384972 | jasmonic acid biosynthetic process |
| GO:0009753 | 756 | 23 | 7.5188885 | 305.8962752 | 0.000882651 | response to jasmonic acid stimulus |
| GO:0009863 | 458 | 16 | 4.5550939 | 351.2551115 | 0.005356641 | salicylic acid mediated signaling pathway |
| GO:0004601 | 237 | 9 | 1.64637 | 546.6571945 | 0.006520455 | peroxidase activity |
| GO:0009867 | 800 | 22 | 7.9564958 | 276.5036331 | 0.006772526 | jasmonic acid mediated signaling pathway |
| GO:0004473 | 13 | 3 | 0.0903072 | 3321.993721 | 0.012589926 | malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) activity |
| GO:0016619 | 13 | 3 | 0.0903072 | 3321.993721 | 0.012589926 | malate dehydrogenase (oxaloacetate-decarboxylating) activity |
| GO:0004567 | 3 | 2 | 0.0208401 | 9596.870748 | 0.020142076 | beta-mannosidase activity |
| GO:0047668 | 3 | 2 | 0.0208401 | 9596.870748 | 0.020142076 | amygdalin beta-glucosidase activity |
| GO:0080081 | 3 | 2 | 0.0208401 | 9596.870748 | 0.020142076 | 4-methylumbelliferyl-beta-D-glucopyranoside beta-glucosidase activity |
| GO:0080082 | 3 | 2 | 0.0208401 | 9596.870748 | 0.020142076 | esculin beta-glucosidase activity |
| GO:0080083 | 3 | 2 | 0.0208401 | 9596.870748 | 0.020142076 | beta-gentiobiose beta-glucosidase activity |
| GO:0004470 | 19 | 3 | 0.1319875 | 2272.943072 | 0.041362028 | malic enzyme activity |
| GO:0016652 | 19 | 3 | 0.1319875 | 2272.943072 | 0.041362028 | oxidoreductase activity, acting on NAD(P)H, NAD(P) as acceptor |
| **Down Regulated in Peking** | | | | | | |
| GO:0008271 | 36 | 4 | 0.1246776 | 3208.275463 | 0.000668772 | secondary active sulfate transmembrane transporter activity |
| GO:0015116 | 47 | 4 | 0.1627735 | 2457.402482 | 0.001965842 | sulfate transmembrane transporter activity |
| GO:0008272 | 49 | 4 | 0.1993236 | 2006.786901 | 0.009230133 | sulfate transport |
| GO:0005504 | 9 | 2 | 0.0311694 | 6416.550926 | 0.037788551 | fatty acid binding |
| GO:0018131 | 6 | 2 | 0.024407 | 8194.379845 | 0.046146255 | oxazole or thiazole biosynthetic process |
| **Shared Up Regulation** | | | | | | |
| GO:0080150 | 5 | 2 | 0.0104511 | 19136.81818 | 0.00192795 | S-adenosyl-L-methionine:benzoic acid carboxyl methyl transferase activity |
| GO:0052624 | 8 | 2 | 0.0167217 | 11960.51136 | 0.005376293 | 2-phytyl-1,4-naphthoquinone methyltransferase activity |
| **Shared Down regulation** | | | | | | |
| GO:0015116 | 47 | 3 | 0.0632554 | 4742.680851 | 0.001314923 | sulfate transmembrane transporter activity |
| GO:0015098 | 7 | 2 | 0.009421 | 21229.14286 | 0.001369833 | molybdate ion transmembrane transporter activity |
| GO:0015689 | 7 | 2 | 0.0127761 | 15654.1744 | 0.004582665 | molybdate ion transport |
| GO:0015114 | 21 | 2 | 0.028263 | 7076.380952 | 0.013534433 | phosphate ion transmembrane transporter activity |

**Supplementary Table 4.** Significant downregulated genes in NRS100 overrepresented enriched GO terms.

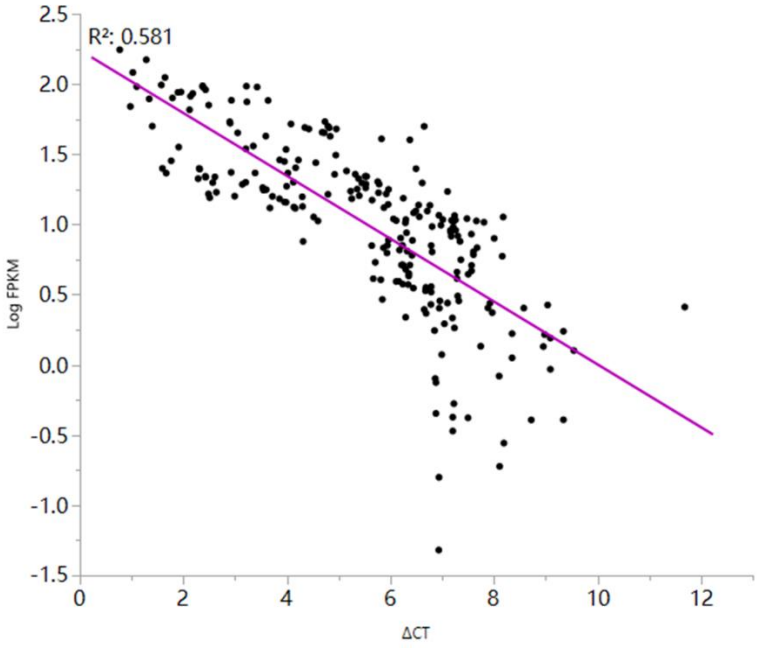
| GO ID | GO count | Expressed | Expected_ex | Corrected_P | GO_desc |
|---|---|---|---|---|---|
| GO:0009834 | 135 | 62 | 8.460091119 | 1.36E-36 | secondary cell wall biogenesis |
| GO:0010413 | 495 | 93 | 31.0203341 | 4.60E-20 | glucuronoxylan metabolic process |
| GO:0045492 | 497 | 93 | 31.14566879 | 6.18E-20 | xylan biosynthetic process |
| GO:0009765 | 44 | 20 | 2.757363031 | 1.18E-10 | photosynthesis, light harvesting |
| GO:0015706 | 486 | 73 | 30.45632803 | 4.83E-10 | nitrate transport |
| GO:0019344 | 459 | 70 | 28.76430981 | 4.83E-10 | cysteine biosynthetic process |
| GO:0010167 | 479 | 71 | 30.01765664 | 1.56E-09 | response to nitrate |
| GO:0000041 | 201 | 41 | 12.59613567 | 3.15E-09 | transition metal ion transport |
| GO:0010218 | 251 | 46 | 15.72950275 | 8.20E-09 | response to far red light |
| GO:0010106 | 237 | 44 | 14.85215996 | 1.47E-08 | cellular response to iron ion starvation |
| GO:0009832 | 313 | 52 | 19.61487793 | 1.80E-08 | plant-type cell wall biogenesis |
| GO:0006826 | 241 | 43 | 15.10282933 | 9.40E-08 | iron ion transport |
| GO:0042546 | 185 | 34 | 11.5934582 | 3.73E-06 | cell wall biogenesis |
| GO:0010417 | 35 | 14 | 2.193356957 | 4.56E-06 | glucuronoxylan biosynthetic process |
| GO:0009637 | 269 | 42 | 16.8575149 | 9.78E-06 | response to blue light |
| GO:0055114 | 2341 | 206 | 146.7042467 | 1.92E-05 | oxidation-reduction process |
| GO:0046274 | 64 | 18 | 4.010709864 | 2.01E-05 | lignin catabolic process |
| GO:0070838 | 237 | 38 | 14.85215996 | 2.33E-05 | divalent metal ion transport |
| GO:0009657 | 201 | 34 | 12.59613567 | 3.24E-05 | plastid organization |
| GO:0010400 | 6 | 6 | 0.37600405 | 4.42E-05 | rhamnogalacturonan I side chain metabolic process |
| GO:0030003 | 274 | 40 | 17.1708516 | 0.000200732 | cellular cation homeostasis |
| GO:0010114 | 252 | 37 | 15.79217009 | 0.000349449 | response to red light |
| GO:0044036 | 222 | 34 | 13.91214984 | 0.00037714 | cell wall macromolecule metabolic process |
| GO:0048767 | 504 | 59 | 31.58434018 | 0.000959254 | root hair elongation |
| GO:0010155 | 191 | 30 | 11.96946225 | 0.001035818 | regulation of proton transport |
| GO:0019684 | 320 | 42 | 20.05354932 | 0.002097215 | photosynthesis, light reaction |
| GO:0015979 | 452 | 52 | 28.32563841 | 0.005329286 | photosynthesis |
| GO:0009416 | 440 | 50 | 27.57363031 | 0.010927635 | response to light stimulus |
| GO:0009413 | 7 | 5 | 0.438671391 | 0.014222921 | response to flooding |
| GO:0010345 | 21 | 8 | 1.316014174 | 0.015555878 | suberin biosynthetic process |
| GO:0009932 | 248 | 33 | 15.54150072 | 0.017338653 | cell tip growth |
| GO:0010817 | 212 | 29 | 13.28547642 | 0.026844375 | regulation of hormone levels |
| GO:0009773 | 121 | 20 | 7.582748336 | 0.029171977 | photosynthetic electron transport in photosystem I |
| GO:0046688 | 29 | 9 | 1.817352907 | 0.030636232 | response to copper ion |
| GO:0006073 | 67 | 14 | 4.198711889 | 0.032103644 | cellular glucan metabolic process |
| GO:0046168 | 8 | 5 | 0.501338733 | 0.036070209 | glycerol-3-phosphate catabolic process |
| GO:0008361 | 144 | 22 | 9.024097194 | 0.041865148 | regulation of cell size |

**Supplementary Table 5**. Selected genes and primers for RT-qPCR.

| Lab Name | Gene ID | Left Primer | Right Primer |
|---|---|---|---|
| Gm182580 | Glyma.18G022400 | CGTGTAGAGTCCTTGAAGTACAGC | ACCAGAGCTGTGATAGCCAACC |
| Gm182590 (SNAP) | Glyma.18G022500 | TCGCCAAATCATGGGACAAGGC | CAATGTGCAGCATCGACATGGG |
| rhg 1 2160 | Gm182160 | AGGTCACGTGTTGCCGTTG | AAACCACACCAATAACAACAAAGCTCT |
| SHMT | Glyma.08G108900 | TGAAAAAGACTTTGAGCAGATTGG | TTGCCATGCTCCTTCTGGAT |
| rhg1 a block | Glyma.18G022300 | TGAGCTAACCATCGAGGAAGA | GCTACCCCTGCAGTGACAA |
| GmSNAP11 | Glyma.11G234500 | CGAAGAGATAGCTCGCCAAT | GGCAGATGCCAGCATTAAG |
| GmSNAP02 | Glyma.02G260400 | TGAAGATATTGCTCGCCAGTC | GCAAATGCCAGCATTAAGAA |
| GmSNAP14 | Glyma.14G054900 | TGAAGATATTGCTCGCCAGTC | GCAAATGCCAGCATTAAGAA |
| GmSNAP09 | Glyma.09G279400 | CCAAGAGAGCGGAGAACAAG | TTGGCGAGTTTGAAGGAAGT |
| JAR1-B | Glyma.03G256200 | TTCGTTGATGCAGGCTACAC | TGGAATGTTCCTCTCCGAAC |
| COI1 | Glyma.18G030200 | TGTCACAGGGCTGTTCAGAG | CATGGTCAAGCAACACAAGG |
| JAZ-A | Glyma.11G038600 | GGCAAGGGAATGTCTCAAAA | TGGCACTGGTTGGAATGATA |
| JAZ-B | Glyma.17G043700 | CCCAAAAGCCCTTCTTCTTC | AGCCTGCGAATGGAACTCT |
| JAZ-C | Glyma.17G047700 | CCTGCTGAAAAATTGGAGGA | AGTGTGAGCACATGCAGAGG |
| JAZ-F | Glyma.09G071600 | TGACACCAAAGGACCTGACA | GGTTGGAATCTCCTCTGAAGG |
| JAZ-G | Glyma.13G112000 | GCCAAAGCACCCTATCAAAT | GCACCTAATCCAAGCCATGA |
| MYC2-C | Glyma.01G096600 | GGGTGAGGAGGACAAAGTCA | AGGGGCCAGAGATTAAGGAA |
| ODC1 | Glyma.04G020200 | TGAACCTGAAGCCCATTTTC | TGGTGTGTCGATGTCTGGTT |
| SS B | Glyma.05G036300 | TCCCCCAGTGAACTTCAAAC | CTACCCCATTTTGCTCTGGA |
| SSA_5 | Glyma.17G091100 | TGATTGGATTCATGCTCTGC | CACCCCATTTTTCTCTGGAT |
| Reference UB1 | GmUB1 | GTGTAATGTTGGATGTGTTCCC | ACACAATTGAGTTCAACACAAACCG |

APPENDX B

**Supplementary Figure 13.** Correlation ($R^2$=0.581) between qPCR expression values (ΔCT) and correlated RNA-sequencing expression values (FPKM) for 20 genes (supplementary table 12). Individual points represent relative expression of individual replicates calculated by RT-qPCR analysis $2^{-\Delta\Delta CT}$ method and calculated FPKM expression of individual replicates from RNA-sequencing analysis.




**Supplementary Table 6.** Significant upregulated enriched GO terms in drought treatment exclusively

| GO ID | Genome Count | Expressed_GO | Expected_expression | Corrected_P | GO_desc |
|---|---|---|---|---|---|
| | | | | Up Drought | |
| GO:0019375 | 234 | 23 | 2.241532672 | 4.66E-14 | galactolipid biosynthetic process |
| GO:0016036 | 330 | 26 | 3.161135819 | 1.12E-13 | cellular response to phosphate starvation |
| GO:0010054 | 51 | 12 | 0.488539172 | 1.88E-11 | trichoblast differentiation |
| GO:0009247 | 13 | 6 | 0.124529593 | 3.87E-07 | glycolipid biosynthetic process |
| GO:0046506 | 6 | 4 | 0.057475197 | 3.95E-05 | sulfolipid biosynthetic process |
| GO:0080040 | 4 | 3 | 0.038316798 | 0.00112321 | positive regulation of cellular response to phosphate starvation |
| GO:0010223 | 33 | 5 | 0.316113582 | 0.00481507 | secondary shoot formation |
| GO:0010623 | 7 | 3 | 0.067054396 | 0.00962054 | developmental programmed cell death |
| GO:0046274 | 64 | 6 | 0.613068765 | 0.011208341 | lignin catabolic process |
| GO:0016121 | 10 | 3 | 0.095791995 | 0.03228913 | carotene catabolic process |
| GO:0016124 | 10 | 3 | 0.095791995 | 0.03228913 | xanthophyll catabolic process |
| GO:0010413 | 495 | 15 | 4.741703729 | 0.033405737 | glucuronoxylan metabolic process |
| GO:0045492 | 497 | 15 | 4.760862128 | 0.034909973 | xylan biosynthetic process |
| GO:1901684 | 11 | 3 | 0.105371194 | 0.044083548 | arsenate ion transmembrane transport |
| GO:0006817 | 52 | 5 | 0.498118372 | 0.045435198 | phosphate ion transport |

**Supplementary Table 7.** Significant upregulated enriched GO terms in combination treatment exclusively

| GO ID | Genome Count | Expressed_GO | Expected_expression | Corrected_P | GO_desc |
|---|---|---|---|---|---|
| | | | | Up Combination | |
| GO:0010200 | 1135 | 135 | 49.31754528 | 2.98E-24 | response to chitin |
| GO:0002679 | 420 | 68 | 18.24966433 | 1.34E-18 | respiratory burst involved in defense response |
| GO:0035556 | 479 | 68 | 20.81330765 | 1.86E-15 | intracellular signal transduction |
| GO:0010363 | 1019 | 106 | 44.2771618 | 2.18E-14 | regulation of plant-type hypersensitive response |
| GO:0006612 | 1020 | 106 | 44.32061338 | 2.27E-14 | protein targeting to membrane |
| GO:0009863 | 458 | 63 | 19.90082444 | 1.51E-13 | salicylic acid mediated signaling pathway |
| GO:0043069 | 525 | 66 | 22.81208041 | 2.63E-12 | negative regulation of programmed cell death |
| GO:0030968 | 501 | 58 | 21.76924245 | 3.79E-09 | endoplasmic reticulum unfolded protein response |
| GO:0006412 | 763 | 2 | 33.15355687 | 8.36E-09 | translation |
| GO:0009963 | 300 | 41 | 13.03547452 | 3.09E-08 | positive regulation of flavonoid biosynthetic process |
| GO:0031348 | 766 | 74 | 33.28391161 | 3.78E-08 | negative regulation of defense response |
| GO:0009738 | 634 | 65 | 27.54830282 | 4.46E-08 | abscisic acid mediated signaling pathway |
| GO:0009697 | 653 | 66 | 28.37388288 | 7.87E-08 | salicylic acid biosynthetic process |
| GO:0050832 | 941 | 83 | 40.88793842 | 2.42E-07 | defense response to fungus |
| GO:0009611 | 1031 | 88 | 44.79858078 | 5.11E-07 | response to wounding |
| GO:0009627 | 751 | 69 | 32.63213789 | 1.89E-06 | systemic acquired resistance |
| GO:0009867 | 800 | 71 | 34.76126539 | 3.86E-06 | jasmonic acid mediated signaling pathway |
| GO:0045087 | 274 | 35 | 11.9057334 | 4.90E-06 | innate immune response |
| GO:0031347 | 345 | 39 | 14.9907957 | 2.14E-05 | regulation of defense response |
| GO:0009793 | 1180 | 17 | 51.27286646 | 0.000103159 | embryo development ending in seed dormancy |
| GO:0009693 | 270 | 32 | 11.73192707 | 0.000125662 | ethylene biosynthetic process |
| GO:0009862 | 688 | 60 | 29.89468824 | 0.000129916 | systemic acquired resistance, salicylic acid mediated signaling pathway |
| GO:0001510 | 418 | 1 | 18.16276117 | 0.00079114 | RNA methylation |
| GO:0000165 | 575 | 49 | 24.9846595 | 0.002794434 | MAPK cascade |
| GO:0009909 | 744 | 9 | 32.32797682 | 0.004697536 | regulation of flower development |
| GO:0010508 | 8 | 5 | 0.347612654 | 0.005866077 | positive regulation of autophagy |
| GO:0006468 | 2386 | 145 | 103.675474 | 0.008057313 | protein phosphorylation |
| GO:0009695 | 417 | 38 | 18.11930959 | 0.009956987 | jasmonic acid biosynthetic process |
| GO:0006364 | 532 | 5 | 23.11624149 | 0.016359402 | rRNA processing |
| GO:0009612 | 152 | 19 | 6.604640425 | 0.019222689 | response to mechanical stimulus |
| GO:0009873 | 311 | 30 | 13.51344192 | 0.021596002 | ethylene mediated signaling pathway |
| GO:0010228 | 754 | 11 | 32.76249263 | 0.03155419 | vegetative to reproductive phase transition of meristem |
| GO:0010207 | 372 | 2 | 16.16398841 | 0.036463262 | photosystem II assembly |
| GO:0009220 | 315 | 1 | 13.68724825 | 0.045841081 | pyrimidine ribonucleotide biosynthetic process |
| GO:0009753 | 756 | 56 | 32.8493958 | 0.046084423 | response to jasmonic acid stimulus |

**Supplementary Table 8.** Significant upregulated enriched GO terms in SCN and combination treatment exclusively

| GO ID | Genome Count | Expressed_GO | Expected_expression | Corrected_P | GO_desc |
|---|---|---|---|---|---|
| Up SCN and Combination | | | | | |
| GO:0010200 | 1135 | 166 | 68.10272928 | 1.24E-25 | response to chitin |
| GO:0031347 | 345 | 74 | 20.70082961 | 2.90E-20 | regulation of defense response |
| GO:0043069 | 525 | 93 | 31.50126244 | 1.75E-19 | negative regulation of programmed cell death |
| GO:0010363 | 1019 | 141 | 61.14245034 | 2.81E-19 | regulation of plant-type hypersensitive response |
| GO:0006612 | 1020 | 141 | 61.20245275 | 3.05E-19 | protein targeting to membrane |
| GO:0002679 | 420 | 80 | 25.20100995 | 1.42E-18 | respiratory burst involved in defense response |
| GO:0009867 | 800 | 118 | 48.00192372 | 3.62E-18 | jasmonic acid mediated signaling pathway |
| GO:0009697 | 653 | 103 | 39.18157024 | 8.07E-18 | salicylic acid biosynthetic process |
| GO:0009863 | 458 | 82 | 27.48110133 | 2.56E-17 | salicylic acid mediated signaling pathway |
| GO:0009862 | 688 | 103 | 41.2816544 | 3.75E-16 | systemic acquired resistance, salicylic acid mediated signaling pathway |
| GO:0031348 | 766 | 109 | 45.96184196 | 1.66E-15 | negative regulation of defense response |
| GO:0000165 | 575 | 89 | 34.50138268 | 1.09E-14 | MAPK cascade |
| GO:0010310 | 535 | 84 | 32.10128649 | 3.65E-14 | regulation of hydrogen peroxide metabolic process |
| GO:0006412 | 763 | 2 | 45.78183475 | 6.90E-14 | translation |
| GO:0045088 | 64 | 26 | 3.840153898 | 2.37E-13 | regulation of innate immune response |
| GO:0050832 | 941 | 118 | 56.46226278 | 1.48E-12 | defense response to fungus |
| GO:0043900 | 264 | 51 | 15.84063483 | 1.32E-11 | regulation of multi-organism process |
| GO:0009595 | 278 | 52 | 16.68066849 | 2.86E-11 | detection of biotic stimulus |
| GO:0009753 | 756 | 98 | 45.36181792 | 4.37E-11 | response to jasmonic acid stimulus |
| GO:0009627 | 751 | 97 | 45.06180589 | 6.90E-11 | systemic acquired resistance |
| GO:0006952 | 1116 | 128 | 66.96268359 | 7.44E-11 | defense response |
| GO:0051707 | 255 | 47 | 15.30061319 | 8.21E-10 | response to other organism |
| GO:0009620 | 310 | 52 | 18.60074544 | 2.37E-09 | response to fungus |
| GO:0009611 | 1031 | 113 | 61.8624792 | 4.23E-08 | response to wounding |
| GO:0009617 | 407 | 59 | 24.42097869 | 4.26E-08 | response to bacterium |
| GO:0001666 | 195 | 37 | 11.70046891 | 8.50E-08 | response to hypoxia |
| GO:0051567 | 443 | 1 | 26.58106526 | 3.15E-07 | histone H3-K9 methylation |
| GO:0045087 | 274 | 44 | 16.44065887 | 4.49E-07 | innate immune response |
| GO:0009751 | 380 | 54 | 22.80091377 | 7.79E-07 | response to salicylic acid stimulus |
| GO:0009793 | 1180 | 24 | 70.80283749 | 1.18E-06 | embryo development ending in seed dormancy |
| GO:0009693 | 270 | 42 | 16.20064926 | 2.91E-06 | ethylene biosynthetic process |
| GO:0030968 | 501 | 63 | 30.06120473 | 4.28E-06 | endoplasmic reticulum unfolded protein response |
| GO:0000226 | 379 | 1 | 22.74091136 | 9.28E-06 | microtubule cytoskeleton organization |
| GO:0009723 | 726 | 81 | 43.56174578 | 9.48E-06 | response to ethylene stimulus |
| GO:0009909 | 744 | 11 | 44.64178906 | 9.71E-06 | regulation of flower development |
| GO:0019288 | 581 | 6 | 34.8613971 | 1.18E-05 | isopentenyl diphosphate biosynthetic process, mevalonate-independent pathway |
| GO:0010112 | 11 | 8 | 0.660026451 | 1.34E-05 | regulation of systemic acquired resistance |
| GO:0010207 | 372 | 1 | 22.32089453 | 1.36E-05 | photosystem II assembly |
| GO:0001510 | 418 | 2 | 25.08100514 | 1.43E-05 | RNA methylation |
| GO:0006098 | 421 | 2 | 25.26101236 | 1.50E-05 | pentose-phosphate shunt |
| GO:0042538 | 461 | 57 | 27.66110854 | 3.98E-05 | hyperosmotic salinity response |
| GO:0010027 | 469 | 4 | 28.14112778 | 5.98E-05 | thylakoid membrane organization |
| GO:0035556 | 479 | 58 | 28.74115183 | 7.29E-05 | intracellular signal transduction |
| GO:0009640 | 502 | 5 | 30.12120714 | 9.04E-05 | photomorphogenesis |
| GO:0009612 | 152 | 27 | 9.120365507 | 0.000112621 | response to mechanical stimulus |
| GO:0009738 | 634 | 70 | 38.04152455 | 0.000140579 | abscisic acid mediated signaling pathway |
| GO:1900056 | 23 | 10 | 1.380055307 | 0.000175163 | negative regulation of leaf senescence |
| GO:0015979 | 452 | 4 | 27.1210869 | 0.00017682 | photosynthesis |
| GO:0006342 | 372 | 2 | 22.32089453 | 0.00019408 | chromatin silencing |
| GO:0009814 | 196 | 31 | 11.76047131 | 0.000211355 | defense response, incompatible interaction |
| GO:0006979 | 632 | 69 | 37.92151974 | 0.000228712 | response to oxidative stress |
| GO:0006995 | 149 | 26 | 8.940358293 | 0.000274041 | cellular response to nitrogen starvation |
| GO:0009695 | 417 | 51 | 25.02100274 | 0.000300689 | jasmonic acid biosynthetic process |
| GO:0000278 | 398 | 3 | 23.88095705 | 0.000321311 | mitotic cell cycle |
| GO:0008283 | 388 | 3 | 23.28093301 | 0.000677129 | cell proliferation |
| GO:0006306 | 421 | 4 | 25.26101236 | 0.000690432 | DNA methylation |
| GO:0042742 | 1063 | 100 | 63.78255615 | 0.000895347 | defense response to bacterium |
| GO:0009560 | 289 | 1 | 17.34069494 | 0.001267943 | embryo sac egg cell differentiation |
| GO:0006281 | 290 | 1 | 17.40069735 | 0.00126945 | DNA repair |
| GO:0009873 | 311 | 40 | 18.66074785 | 0.001428245 | ethylene mediated signaling pathway |
| GO:0006396 | 330 | 2 | 19.80079354 | 0.001723404 | RNA processing |
| GO:0010120 | 35 | 11 | 2.100084163 | 0.00200821 | camalexin biosynthetic process |
| GO:0010088 | 9 | 6 | 0.540021642 | 0.002144911 | phloem development |
| GO:0010228 | 754 | 16 | 45.24181311 | 0.00241237 | vegetative to reproductive phase transition of meristem |
| GO:0009630 | 358 | 3 | 21.48086087 | 0.002722509 | gravitropism |
| GO:0009625 | 83 | 17 | 4.980199586 | 0.003130685 | response to insect |
| GO:0009637 | 269 | 1 | 16.14064685 | 0.003924651 | response to blue light |
| GO:0016192 | 549 | 9 | 32.94132015 | 0.003935083 | vesicle-mediated transport |
| GO:0007165 | 1328 | 116 | 79.68319338 | 0.00482547 | signal transduction |
| GO:0006346 | 306 | 2 | 18.36073582 | 0.005006232 | methylation-dependent chromatin silencing |
| GO:0007131 | 263 | 1 | 15.78063242 | 0.00573157 | reciprocal meiotic recombination |
| GO:0042254 | 264 | 1 | 15.84063483 | 0.005774493 | ribosome biogenesis |
| GO:0007154 | 186 | 27 | 11.16044727 | 0.006434846 | cell communication |
| GO:0052542 | 116 | 20 | 6.96027894 | 0.007306987 | defense response by callose deposition |
| GO:0006886 | 410 | 5 | 24.60098591 | 0.007314858 | intracellular protein transport |
| GO:0031048 | 261 | 1 | 15.66062761 | 0.009029791 | chromatin silencing by small RNA |
| GO:0080142 | 7 | 5 | 0.420016833 | 0.009987281 | regulation of salicylic acid biosynthetic process |
| GO:0009624 | 221 | 30 | 13.26053143 | 0.010400341 | response to nematode |
| GO:0070588 | 34 | 10 | 2.040081758 | 0.011291901 | calcium ion transmembrane transport |
| GO:0010218 | 251 | 1 | 15.06060357 | 0.012334304 | response to far red light |
| GO:0015031 | 286 | 2 | 17.16068773 | 0.014846739 | protein transport |
| GO:0006606 | 243 | 1 | 14.58058433 | 0.017737057 | protein import into nucleus |
| GO:0043090 | 266 | 33 | 15.96063964 | 0.022929972 | amino acid import |
| GO:0006261 | 234 | 1 | 14.04056269 | 0.025645517 | DNA-dependent DNA replication |
| GO:0009759 | 30 | 9 | 1.80007214 | 0.025790474 | indole glucosinolate biosynthetic process |
| GO:0035196 | 342 | 4 | 20.52082239 | 0.029644558 | production of miRNAs involved in gene silencing by miRNA |
| GO:0000394 | 343 | 4 | 20.5808248 | 0.029687785 | RNA splicing, via endonucleolytic cleavage and ligation |
| GO:0051726 | 345 | 4 | 20.70082961 | 0.030052048 | regulation of cell cycle |
| GO:0006486 | 348 | 4 | 20.88083682 | 0.031319638 | protein glycosylation |
| GO:0050776 | 25 | 8 | 1.500060116 | 0.042893348 | regulation of immune response |
| GO:0009864 | 32 | 9 | 1.920076949 | 0.045606417 | induced systemic resistance, jasmonic acid mediated signaling pathway |

**Supplementary Table 9.** Significant Downregulated enriched GO terms in SCN, Drought, and all three treatments exclusively

| GO ID | Genome Count | Expressed_GO | Expected_expression | Corrected_P | GO_desc |
|---|---|---|---|---|---|
| **Down Drought** | | | | | |
| GO:0042802 | 318 | 5 | 0.596057971 | 0.018806807 | identical protein binding |
| GO:0071333 | 9 | 2 | 0.02357442 | 0.027014017 | cellular response to glucose stimulus |
| **Down SCN** | | | | | |
| GO:0009834 | 135 | 11 | 1.489875958 | 0.000161232 | secondary cell wall biogenesis |
| **Down Drought, SCN, and Combination** | | | | | |
| GO:0008289 | 187 | 5 | 0.386060108 | 0.002366365 | lipid binding |
| GO:0008515 | 34 | 3 | 0.070192747 | 0.002529211 | sucrose transmembrane transporter activity |
| GO:0005504 | 9 | 2 | 0.018580433 | 0.007802209 | fatty acid binding |
| GO:0051119 | 53 | 3 | 0.109418105 | 0.009623042 | sugar transmembrane transporter activity |
| GO:0015770 | 34 | 3 | 0.09170877 | 0.009645032 | sucrose transport |
| GO:0005319 | 11 | 2 | 0.022709418 | 0.011887933 | lipid transporter activity |
| GO:0071446 | 11 | 2 | 0.029670485 | 0.034995632 | cellular response to salicylic acid stimulus |
| GO:0071470 | 13 | 2 | 0.035065118 | 0.049455299 | cellular response to osmotic stress |

**Supplementary Table 10.** Significant Downregulated enriched GO terms in combination exclusively

| GO ID | Genome Count | Expressed_GO | Expected_expression | Corrected_P | GO_desc |
|---|---|---|---|---|---|
| **Down Combination** | | | | | |
| GO:0006412 | 763 | 183 | 32.61428151 | 1.12011E-81 | translation |
| GO:0001510 | 418 | 125 | 17.86732591 | 2.17778E-67 | RNA methylation |
| GO:0042254 | 264 | 71 | 11.28462689 | 2.67978E-34 | ribosome biogenesis |
| GO:0009220 | 315 | 70 | 13.46461163 | 3.9863E-28 | pyrimidine ribonucleotide biosynthetic process |
| GO:0006084 | 178 | 38 | 7.608574192 | 5.26126E-14 | acetyl-CoA metabolic process |
| GO:0042545 | 266 | 43 | 11.37011649 | 1.94355E-11 | cell wall modification |
| GO:0016132 | 271 | 43 | 11.58384048 | 3.79295E-11 | brassinosteroid biosynthetic process |
| GO:0009664 | 337 | 46 | 14.40499721 | 1.32066E-09 | plant-type cell wall organization |
| GO:0016126 | 363 | 48 | 15.51636198 | 1.41606E-09 | sterol biosynthetic process |
| GO:0006606 | 243 | 37 | 10.38698612 | 7.63606E-09 | protein import into nucleus |
| GO:0006633 | 237 | 32 | 10.13051732 | 4.06246E-06 | fatty acid biosynthetic process |
| GO:0009697 | 653 | 3 | 27.91235364 | 8.54982E-06 | salicylic acid biosynthetic process |
| GO:0042742 | 1063 | 12 | 45.43772116 | 1.55864E-05 | defense response to bacterium |
| GO:0042991 | 71 | 16 | 3.034880717 | 2.28291E-05 | transcription factor import into nucleus |
| GO:0000271 | 274 | 33 | 11.71207488 | 4.02833E-05 | polysaccharide biosynthetic process |
| GO:0000462 | 13 | 7 | 0.555682385 | 0.000261658 | maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) |
| GO:0009753 | 756 | 7 | 32.31506792 | 0.000284619 | response to jasmonic acid stimulus |
| GO:0009832 | 313 | 34 | 13.37912203 | 0.000306728 | plant-type cell wall biogenesis |
| GO:0048829 | 50 | 12 | 2.137239942 | 0.000617783 | root cap development |
| GO:0043481 | 310 | 33 | 13.25088764 | 0.000715726 | anthocyanin accumulation in tissues in response to UV light |
| GO:0010817 | 212 | 26 | 9.061897352 | 0.000732895 | regulation of hormone levels |
| GO:0006414 | 81 | 15 | 3.462328705 | 0.000936013 | translational elongation |
| GO:0009932 | 248 | 28 | 10.60071011 | 0.001505882 | cell tip growth |
| GO:0010200 | 1135 | 18 | 48.51534667 | 0.001840559 | response to chitin |
| GO:0044267 | 46 | 11 | 1.966260746 | 0.001878767 | cellular protein metabolic process |
| GO:0016192 | 549 | 4 | 23.46689456 | 0.002499513 | vesicle-mediated transport |
| GO:0006952 | 1116 | 18 | 47.7031955 | 0.003171897 | defense response |
| GO:0044070 | 18 | 7 | 0.769406379 | 0.004046803 | regulation of anion transport |
| GO:0006626 | 235 | 26 | 10.04502773 | 0.005051799 | protein targeting to mitochondrion |
| GO:0006334 | 128 | 18 | 5.471334251 | 0.005187748 | nucleosome assembly |
| GO:0010075 | 450 | 40 | 19.23515947 | 0.006428815 | regulation of meristem growth |
| GO:0048767 | 504 | 43 | 21.54337861 | 0.007821147 | root hair elongation |
| GO:0006486 | 348 | 1 | 14.87518999 | 0.015928674 | protein glycosylation |
| GO:0042026 | 15 | 6 | 0.641171982 | 0.016823815 | protein refolding |
| GO:0050832 | 941 | 15 | 40.2228557 | 0.02150667 | defense response to fungus |
| GO:0019745 | 69 | 12 | 2.949391119 | 0.021775494 | pentacyclic triterpenoid biosynthetic process |
| GO:0006085 | 23 | 7 | 0.983130373 | 0.026013215 | acetyl-CoA biosynthetic process |
| GO:0007169 | 348 | 32 | 14.87518999 | 0.029029429 | transmembrane receptor protein tyrosine kinase signaling pathway |
| GO:0009165 | 201 | 22 | 8.591704565 | 0.030456759 | nucleotide biosynthetic process |
| GO:0031348 | 766 | 11 | 32.74251591 | 0.033343623 | negative regulation of defense response |
| GO:0010400 | 6 | 4 | 0.256468793 | 0.038726983 | rhamnogalacturonan I side chain metabolic process |
| GO:0009834 | 135 | 17 | 5.770547842 | 0.040285085 | secondary cell wall biogenesis |
| GO:0009738 | 634 | 8 | 27.10020246 | 0.049685245 | abscisic acid mediated signaling pathway |

**Supplementary Table 11.** Significant Downregulated enriched GO terms in SCN and combination

| GO ID | Genome Count | Expressed_GO | Expected_expression | Corrected_P | GO_desc |
|---|---|---|---|---|---|
| Down SCN and Combination | | | | | |
| GO:0009834 | 135 | 32 | 4.676182848 | 1.73347E-15 | secondary cell wall biogenesis |
| GO:0010413 | 495 | 58 | 17.14600378 | 2.39015E-13 | glucuronoxylan metabolic process |
| GO:0045492 | 497 | 58 | 17.21528056 | 2.86838E-13 | xylan biosynthetic process |
| GO:0055114 | 2341 | 141 | 81.08847442 | 5.15586E-08 | oxidation-reduction process |
| GO:0000041 | 201 | 28 | 6.962316685 | 1.99858E-07 | transition metal ion transport |
| GO:0015706 | 486 | 45 | 16.83425825 | 1.19565E-06 | nitrate transport |
| GO:0010106 | 237 | 29 | 8.209298778 | 2.10574E-06 | cellular response to iron ion starvation |
| GO:0006826 | 241 | 29 | 8.347852343 | 3.09925E-06 | iron ion transport |
| GO:0010167 | 479 | 43 | 16.59178951 | 6.82341E-06 | response to nitrate |
| GO:0006457 | 779 | 3 | 26.98330695 | 1.71803E-05 | protein folding |
| GO:0071732 | 36 | 10 | 1.246982093 | 0.00017235 | cellular response to nitric oxide |
| GO:0071219 | 24 | 8 | 0.831321395 | 0.000613349 | cellular response to molecule of bacterial origin |
| GO:0030001 | 247 | 25 | 8.555682692 | 0.000978243 | metal ion transport |
| GO:0006412 | 763 | 5 | 26.42909269 | 0.001208703 | translation |
| GO:0010345 | 21 | 7 | 0.727406221 | 0.003100405 | suberin biosynthetic process |
| GO:0009793 | 1180 | 14 | 40.87330193 | 0.003722725 | embryo development ending in seed dormancy |
| GO:0042546 | 185 | 20 | 6.408102421 | 0.004095828 | cell wall biogenesis |
| GO:0016036 | 330 | 28 | 11.43066918 | 0.007298561 | cellular response to phosphate starvation |
| GO:0046854 | 79 | 12 | 2.736432926 | 0.010727871 | phosphatidylinositol phosphorylation |
| GO:0071369 | 27 | 7 | 0.93523657 | 0.019842686 | cellular response to ethylene stimulus |
| GO:0006886 | 410 | 1 | 14.2017405 | 0.020809893 | intracellular protein transport |
| GO:0006364 | 532 | 3 | 18.42762426 | 0.025892857 | rRNA processing |
| GO:0009909 | 744 | 7 | 25.77096325 | 0.03085878 | regulation of flower development |
| GO:0040009 | 7 | 4 | 0.24246874 | 0.034786656 | regulation of growth rate |
| GO:0010054 | 51 | 9 | 1.766557965 | 0.038387487 | trichoblast differentiation |
| GO:0046274 | 64 | 10 | 2.216857054 | 0.043713297 | lignin catabolic process |
| GO:0071281 | 78 | 11 | 2.701794534 | 0.049590134 | cellular response to iron ion |

**Supplementary Table 12.** Selected genes and primers for RT-qPCR.

| Gene ID | Left Primer | Right Primer |
|---|---|---|
| Glyma.01G096600 | GGGTGAGGAGGACAAAGTCA | AGGGGCCAGAGATTAAGGAA |
| Glyma.02G260400 | TGAAGATATTGCTCGCCAGTC | GCAAATGCCAGCATTAAGAA |
| Glyma.03G256200 | TTCGTTGATGCAGGCTACAC | TGGAATGTTCCTCTCCGAAC |
| Glyma.04G020200 | TGAACCTGAAGCCCATTTTC | TGGTGTGTCGATGTCTGGTT |
| Glyma.05G036300 | TCCCCCAGTGAACTTCAAAC | CTACCCCATTTTGCTCTGGA |
| Glyma.08G108900 | TGAAAAAGACTTTGAGCAGATTGG | TTGCCATGCTCCTTCTGGAT |
| Glyma.09G071600 | TGACACCAAAGGACCTGACA | GGTTGGAATCTCCTCTGAAGG |
| Glyma.09G279400 | CCAAGAGAGCGGAGAACAAG | TTGGCGAGTTTGAAGGAAGT |
| Glyma.11G038600 | GGCAAGGGAATGTCTCAAAA | TGGCACTGGTTGGAATGATA |
| Glyma.11G234500 | CGAAGAGATAGCTCGCCAAT | GGCAGATGCCAGCATTAAG |
| Glyma.13G112000 | GCCAAAGCACCCTATCAAAT | GCACCTAATCCAAGCCATGA |
| Glyma.14G054900 | TGAAGATATTGCTCGCCAGTC | GCAAATGCCAGCATTAAGAA |
| Glyma.17G043700 | CCCAAAAGCCCTTCTTCTTC | AGCCTGCGAATGGAACTCT |
| Glyma.17G047700 | CCTGCTGAAAAATTGGAGGA | AGTGTGAGCACATGCAGAGG |
| Glyma.17G091100 | TGATTGGATTCATGCTCTGC | CACCCCATTTTTCTCTGGAT |
| Glyma.18G022300 | TGAGCTAACCATCGAGGAAGA | GCTACCCCTGCAGTGACAA |
| Glyma.18G022400 | CGTGTAGAGTCCTTGAAGTACAGC | ACCAGAGCTGTGATAGCCAACC |
| Glyma.18G022500 | TCGCCAAATCATGGGACAAGGC | CAATGTGCAGCATCGACATGGG |
| Glyma.18G022700 | AGGTCACGTGTTGCCGTTG | AAACCACACCAATAACAACAAAGCTCT |
| Glyma.18G030200 | TGTCACAGGGCTGTTCAGAG | CATGGTCAAGCAACACAAGG |
| GmUB1 | GTGTAATGTTGGATGTGTTCCC | ACACAATTGAGTTCAACACAAACCG |