GENETIC AND METABOLOMIC CHARACTERIZATION OF WILD SOYBEAN IN RESPONSE TO SOYBEAN CYST NEMATODE (*Heterodera glycine*)

by

Neha Mittal

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

2020

Approved by:

_____
Dr. Bao-Hua Song

_____
Dr. Matthew Parrow

_____
Dr. Kenneth Piller

_____
Dr. Changbao Li

_____
Dr. Larry Leamy

_____
Dr. Jessica Shculeter

ABSTRACT

NEHA MITTAL. Genetic and metabolomic characterization of wild soybean in response to soybean cyst nematode (*Heterodera glycine*)
(Under the direction of DR. BAO-HUA SONG)

Crops harvesting ensure human survival on earth. However, sustainable crop production is challenging because of continuous biotic stresses caused by different pests (e.g., nematodes, bacteria, fungi, viruses, and insects). Soybean (*Glycine max*) is an important legume crop supplying more than half of the world's proteins, oil, and fats. It is challenged by the Soybean Cyst Nematode (SCN), *Heterodera glycines* (HG), the most devastating pest worldwide. Currently, the widely used soybean cultivars are losing resistance due to rapid evolution of SCN populations. Developing novel and diverse soybean cultivars resistant to SCN is in urgent need. In this study, we used the soybean wild relative (*Glycine soja*; *G. soja* from now onwards), to study the genetic basis of plant chemical defense to SCN.

We combined multiple approaches including metabolomics and genetics-, to investigate the molecular mechanism(s) underlying nematode resistance in wild soybean (*G. soja*) toward the long-term goal to develop new SCN resistant varieties. Specifically, for the first time, we integrated metabolomics, gene expression, and hairy root transformation analyses to identify the genetic basis conferring broad-spectrum resistance to two rarely studied SCN races (SCN5, HG type 2.5.7 and SCN2, HG type 1.2.5.7).

The outcome of the present study showed involvement of phenyl-propanoid pathway genes and metabolites in SCN2 and SCN5 resistance. Comparison of the metabolic profiles among SCN resistant (S54) and susceptible (S67) wild soybean genotypes showed clear differences, mirroring the solid part of phenolic acid conjugates

(1-O-4-hydroxybenzoyl-β-D-glucose ester, 1-O-3, 4-dihydroxybenzoyl-β-D-glucose ester, 1-O-3, 4, 5-trihydroxybenzoyl-β-D-glucose ester, 1-O- 4-hydroxy-3-methoxybenzoyl-β-D-glucose ester, N-benzoyl-L-glutamic acid and 4-hydroxybenzaldehyde) in defense. The results also indicated that many isoflavonoids (daidzein, daidzin, malonyl daidzin, genistein, formononetin, iso-formononetin) might play important roles in SCN2 and SCN5 resistance. One of the candidate genes, *hydroxyisoflavanone*/*isoflavone-4'-methyl transferase* (*HI4'OMT*/*I4'OMT*), is a small gene family and a genotype-specific allele (*H3*) was found in the resistant genotype S54. Overexpression of this *H3* allele in SCN susceptible soybean cultivar (Willimas 82) using well-developed soybean hairy-root transformation significantly inhibited the development of both SCN races, as well as decreased the SCN numbers, which suggested that this S54 genotype-specific allele could increase levels of SCN resistance. The results of this study provided a foundation toward the goal to develop novel soybean cultivars with enhanced SCN resistance.

ACKNOWLEDGMENTS

I would like to express my deep and sincere gratitude to my advisor Dr. Bao-Hua Song for the continuous support during my Ph. D. and research, for her composure, zeal, encouragement and extensive knowledge. Her dynamism, vision, sincerity and motivation have deeply inspired me & helped me in all the time of research and writing of this thesis. I would also like to thank the rest of my thesis committee: Dr. Matthew Parrow, Dr. Ken Piller, Dr. Larry Leamy, Dr. Changbao Li and Dr. Jessica Schlueter, for their thoughtful comments, interest in my research and sharing their knowledge.

My sincere thanks also go to Dr. Cory Brouwer, and Dr. Robert Reid, for offering me the summer internship opportunities in P2EP program, hence providing me experience on diverse exciting projects. I thank my fellow labmates in UNCC and NCSU: Dr. Hengyou Zhang, Dr. Jaince Kofsky, Dr. Neha Attal, Dr. Richard, Dr. Samantha and Mr. Eric Kane, for the enthisiatsic discussions, for interesting ideas and sharing their thoughts and discussing the project details together before deadlines, and for all the fun we have had in the last six years.

Also, I would like to thank my husband: Mr. Rishu Bansal, without his support, I could not have imagined my Ph. D. study. I am also heartily grateful to my family specially, my mother-in-law (Sudesh Bansal), my sisters (Megha Mittal and Varnika Mittal) for enlightening me and providing me the continuous strength.

Last but not the least, I would like to thank my parents: Neera Mittal and Yogesh Kumar Mittal, for giving birth to me at the first place and supporting me spiritually throughout my life.

TABLE OF CONTENTS

LIST OF TABLES xii

LIST OF FIGURES xiv

LIST OF ABBREVIATIONS xviii

CHAPTER 1: INTRODUCTION 1

1.1 Overview 1

1.2 Crop Wild Relatives (CWRs) in Crop Improvement 5

1.2.1 Biotic stress resistance 5

1.2.2 Abiotic stress tolerance 7

1.2.3 Improvement in yield and quality-related traits 9

1.3 Advanced Biotechnologies Accelerate the Use of Wild Relatives for Crop Improvements 11

1.4 Objective(s) of Present Study 17

CHAPTER 2: MATERIALS AND METHODS 20

2.1 Untargeted Metabolomics Analysis 20

2.1.1 Plant materials and soybean cyst nematode (SCN) stocks 20

2.1.2 Plants preparation, SCN inoculation, and tissue collection 20

2.1.3 Metabolite extraction, and data processing 22

2.1.4 Statistical analysis 22

2.2 Quantitative Expression Analysis of Phenyl-Propanoid Pathway Genes 23

2.2.1 Tissue collection and RNA isolation 23

2.2.2 Quantitative estimation of extracted RNA 24

2.2.3 Qualitative analysis of extracted RNA 24

2.2.4 Reverse transcription PCR for cDNA synthesis 25

2.2.4.1 DNase I enzyme treatment 25

2.2.4.2 First strand cDNA synthesis 25

2.2.4.3 Preparation of dilutions for PCR and qRT-PCR 26

2.2.5 Gene(s) amplification 26

2.2.5.1 Primer designing 26

2.2.5.2 PCR amplification reactions 27

2.2.5.3 Agarose gel electrophoresis 29

2.2.6 Quantitative real-time (qRT-PCR) polymerase chain reaction 30

2.2.7 Statistical analysis of gene expression data 31

2.3 cDNA and Genomic DNA Cloning of *HI4'OMT*/*I4'OMT* Candidate Gene 32

2.3.1 Tissue collection 32

2.3.2 Isolation of full-length cDNAs and genomic DNAs of *HI4'OMT*/*I4'OMT* 33

2.3.2.1 RNA and DNA extraction 33

2.3.2.2 Gene amplification 33

2.3.2.3 Agarose gel electrophoresis 34

2.3.3 Polyadenylation of purified cDNAs 35

2.3.4 Vector ligation 36

2.3.5 Transformation into the competent cells 36

2.3.6 Selection of positive transformants and plasmid extraction 37

2.3.7 Multiple sequence(s) alignment and comparison 38

2.3.8 Mutation screening of *HI4'OMT*/*I4'OMT* 38

2.4 Functional Validation of *HI4'OMT*/*I4'OMT* Candidate Gene or H1-*H4* Alleles 39

2.4.1 Plant, nematode, and bacterium sources 39

2.4.2 Amplification of *HI4'OMT*/*I4'OMT-H1-H4* (haplotype I to haplotype IV) candidate alleles 39

2.4.3 Agarose gel electrophoresis 41

2.4.4 Binary vectors construction 41

2.4.5 Transformation of binary vectors into the *Agrobacterium rhizogenes* competent cells 43

2.4.6 Generation of transgenic soybean hairy roots 44

2.4.7 SCN bioassays 47

2.4.7.1 Root inoculation with SCN 47

2.4.7.2 Analysis of SCN development in transgenic hairy roots 48

2.4.7.3 Female cyst extraction and counting 49

2.4.8 Quantitative reverse transcription PCR 50

CHAPTER 3: RESULTS 52

3.1 LC-MS Based Untargeted Metabolomics 52

3.1.1 Global analysis of metabolic changes in *G. soja* roots infected by *H. glycines* 55

3.1.2 Analysis of induced metabolites in response to *H. glycine* 1.2.5.7 (SCN2) and 2.5.7 (SCN5) infections 57

3.1.3 Metabolomic profiling for time course analysis of metabolic changes in *G. soja* roots infected with SCN2 (HG type 1.2.5.7) and SCN5 (HG type 2.5.7) 58

3.1.3.1 Heat-map for root samples infected with SCN2 (HG type 1.2.5.7) 59

3.1.3.2 Heat-map for root samples infected with SCN5 (HG type 2.5.7) 65

3.1.4 Metabolic pathways allied with HG type 1.2.5.7 (SCN2) and HG type 2.5.7 (SCN5) infections 71

3.2 Quantitative Expression Analysis of Phenyl-propanoid/Iso-flavonoid Pathway Genes 75

3.2.1 Root isoflavone and phenolic acid biosynthetic gene transcript profiling under SCN5 (HG type 2.5.7) attack 78

3.2.2 Quantification of selected genes in resistant and susceptible roots of wild soybean infected with SCN2 (HG type 1.2.5.7) and SCN5 (HG type 2.5.7) 80

3.3 Genomic and cDNA cloning of *Hydroxyisoflavanone/Isoflavone-4'-methyl transferase* (*HI4'OMT/I4'OMT*) Gene 82

3.3.1 *HI4'OMT/I4'OMT* gene variation in S54 and S67 genotypes 83

3.3.2 Haplotypes identification within *HI4'OMT/I4'OMT* gene using geographically diverse genotypes 85

3.4 Hairy root transformation of *HI4'OMT/I4'OMT* gene or *H1-H4* alleles 87

3.4.1 Production of transgenic hairy roots overexpressing *HI4'OMT/I4'OMT-H1-H4* alleles 89

3.4.2 Expression of phenyl-propanoid and iso-flavonoid pathway genes in transgenic hairy roots overexpressing *HI4'OMT/I4'OMT-H1-H4* alleles 91

3.4.3 Evaluation of SCN development and female cysts on transgenic hairy roots overexpressing *H1-H4* candidate alleles from *HI4'OMT/I4'OMT* gene 95

CHAPTER 4: DISCUSSION 99

4.1 Untargeted Metabolomics for Broad-Spectrum SCN Resistance 99

4.1.1 Diverse defense mechanism(s) in response to various environmental stresses 99

4.1.2 Resistant metabolites in response to SCN infection 102

4.1.3 Basal resistance versus pathogen specific resistance: adjusted incorporation as required 107

4.1.4 Metabolites diversity in response to various biological challenges 109

4.2 Phenyl-propanoid or Iso-flavonoid Pathway's Genes Involvement in Broad-Spectrum SCN Resistance 111

4.3 Role of *Hydroxyisoflavanone/Isoflavone-4'-methyl-transferase* (*HI4'OMT/I4'OMT*) Gene in Broad-Spectrum SCN Resistance 114

4.3.1 Selection of *H3* candidate allele from various alleles of *HI4'OMT/I4'OMT* gene 114

4.3.2	Role of *HI4'OMT*/*I4'OMT-H1-H4* alleles in broad-spectrum SCN resistance	116
REFERENCES		122
ANNEXURE I:	Composition of Different Stock Solutions	165
ANNEXURE II:	Supplemental Tables and Figures	167

LIST OF TABLES

Table 1.3.1: Representative advanced technologies that have been used in plant breeding. 13

Table 2.2.5.2.1: Selected primers used in regular PCR and real-time quantitative reverse transcription-polymerase chain reaction (qRT-PCR) for selectively amplifying defense-associated gene copies in wild soybeans (S54 and S67 genotypes). 28

Table 2.2.5.2.2: List of reagents and their concentrations used for PCR amplification. 29

Table 2.2.5.2.3: The sequential steps used for PCR amplification reactions. 29

Table 2.2.6.1: List of reagents and their concentrations used for PCR amplification. 30

Table 2.2.6.2: The sequential steps used for qRT-PCR expression reactions. 31

Table 2.3.1.1: Details of genotypes used for cDNA and genomic DNA cloning. 33

Table 2.3.2.2.1: List of reagents and their concentrations used for PCR amplification. 34

Table 2.3.2.2.2: The sequential steps used for PCR amplification reactions. 34

Table 2.3.3.1: List of reagents and their concentrations used for polyadenylation. 35

Table 2.3.4.1: List of reagents and their concentrations used for vector ligation. 36

Table 2.4.2.1: List of reagents and their concentrations used for PCR amplification. 40

Table 2.4.2.2: The sequential steps used for PCR amplification. 40

Table 2.4.4.1: The sequential steps used for double restriction digestions. 42

Table 2.4.6.1: List of reagents and their concentrations used for PCR amplifications. 46

Table 2.4.6.2: The sequential steps used for PCR amplification reactions. 47

Table 3.1.3.1a: Annotated root metabolites attained from resistant and susceptible genotypes of wild soybeans with and without SCN2 infestation 61

with exact mass, retention time and calculated *P*-values to show the statistically significant variation among different groups.

Table 3.1.3.1b: Annotated root metabolites attained from resistant and susceptible genotypes of wild soybeans with and without SCN5 infestation with exact mass, retention time and calculated *P*-values to show the statistically significant variation among different groups. 67

LIST OF FIGURES

Figure 1.1.1: Map of the known distribution of the soybean cyst nematode, in the United States and Canada from 1954 to 2017. Known infested counties are indicated in red. 2

Figure 1.1.2: The decrease in genetic diversity in modern crops during domestication due to bottleneck events. 4

Figure 1.4.1a: Geographic distribution of wild soybean in East Asia. 18

Figure 1.4.1b: Figure representing the different omics based and molecular techniques which has been selected for the present study. 19

Figure 3.1.1: Direct and indirect defense mechanisms which mainly showed by plant and crop species upon biotic stress or pathogen infections. 53

Figure 3.1.1.1: Principal component analysis of metabolites between two genotypes after 3, 5 and 8 days of **(a)** SCN2 and **(b)** SCN5 J2 nematode inoculation. Each data point in the corresponding category represents the metabolic profile of a single time point, which is averaged for three biological replicates (1-3-A). PC1: the first principal component; PC2: the second principal component. S54-C and S67-C: SCN2 and SCN5 resistant and susceptible control samples respectively; S54-T and S67-T: SCN2 and SCN5 resistant and susceptible treatment samples respectively. 56

Figure 3.1.2.1: (a) Number of expressed metabolites in SCN2 and SCN5 infected roots. Constitutively expressed metabolite means, for one race, metabolites have similarly high level of expression in infected and non-infected roots of one genotype but show significantly higher level than the other genotype. **(b)** Venn diagram showing commonly induced and race-specific induced metabolites. Upward arrows represent significant induction. 58

Figure 3.1.3.1.1: Heat map showing changes of metabolite levels in SCN2 infected and non-infected roots of resistant and susceptible genotypes at different time points. Red and blue color represents the increased and decreased metabolite levels respectively (see color scale); Highlighted black and yellow box shows the induced metabolites in resistant and susceptible infected roots correspondingly; X represents the metabolites which showed multiple hits in the database. Aut shows that these compounds compared with authentic standards. Metabolites highlighted in purple shows the clustering of commonly induced metabolites (cDEMs) among both races. 64

Figure 3.1.3.2.1: Heat map showing changes of metabolite levels in SCN5 infected and non-infected roots of resistant and susceptible genotypes at different time points. Red and blue color represents the increased and decreased metabolite levels respectively (see color scale); Highlighted black 70

and yellow box shows the induced metabolites in resistant and susceptible infected roots correspondingly; Aut shows that these compounds compared with authentic standards. Metabolites highlighted in purple shows the clustering of commonly induced metabolites (cDEMs) among both races.

Figure 3.1.4.1a and b: Summary of changed metabolite levels in roots of resistant (S54) and susceptible (S67) genotypes against SCN2 infection. X-axis represent samples, RC- non-infected resistant, RT- infected resistant, SC- non-infected susceptible, ST- infected susceptible respectively. The concetration of metabolites on Y-axis is presented as normalized and cube root transformed peak areas by Metaboanalyst software (www.metaboanalyst.ca/), and the box plots show centered means and bars show standard deviation of each metabolite. Error bars and *p*-values has been calculated by ANOVA with Fisher's LSD post-hoc analysis using Metaboanalyst program. Peak areas of each metabolite were averaged at 8 time point within each group in both genotypes using all three biological replicates. doi: 10.1371/journal.pone.0055431. 72, 74

Figure 3.2.1: Figure is showing the step by step synthesis of different iso-flavonoids and flavonoids *via* phenyl-propanoid pathway in plants. The full form of abbreviated enzymes is provided in Annexure II, Table 3.2.1. 77

Figure 3.2.1.1: Heat map showing the expression pattern of phenyl-propanoid pathway genes among resistant and susceptible genotypes at 3, 5 and 8 dpi. Three biological replicates for both genotypes at at each time point were used to generate the heat map. Genes which are highlighted in dashed green rectangle shows the induced expression. RT-resistant treatment, ST-susceptible treatment, RC-resistant control and SC-susceptible control; dpi-days post infection; Colour scale: red and green color represents the increased and decreased gene expression levels respectively. 79

Figure 3.2.2.2.1: Expression patterns of three selected candidate genes involved in phenyl-propanoid pathway. Total RNA was extracted from the roots of both genotypes and was used in qRT-PCR analysis. Expression analysis of two *PAL* (*PAL1.1* (Glyma.03g181600), and *PAL1.2* (Glyma.03g181700)) and *HI4'OMT*/*I4'OMT* (Glyma.13g173300) genes that were upregulated in S54, can observed in subsection a, b and c respectively. The bar with asterisks is significantly different (unpaired student's *t*-test $^{*}P \leq 0.05$ and $^{**}P \leq 0.01$). S54- SCN-resistant; S67- SCN-susceptible; R2- SCN2 (HG type 1.2.5.7) and R5- SCN5 (HG type 2.5.7); 80

Figure 3.3.1.1a: Isoflavonoid pathway showing the biosynthesis of formononetin metabolite via *HI4'OMT* and *I4'OMT* enzymes; **(b)** relative expression of *HI4'OMT*/*I4'OMT* gene among S54 and S67 at 8 dpi; and **(c)** expression differences of formononetin metabolite among S54 and S67 at 8 dpi; 84

Figure 3.3.1.1: (d) *HI4'OMT*/*I4'OMT* gene model and polymorphisms between S54, S67 and Williams 82; **(e)** *HI4'OMT*/*I4'OMT* predicted protein sequences showing amino acid differences (I41V and G202D) between Williams 82, S54 and S67. 85

Figure 3.3.2.1: Identified haplotypes (haplotype I; *H1* to haplotype VI; *H6*) of *HI4'OMT*/*I4'OMT* gene among wild soybean (S54, S67, S100, PW5) and soybean (SC1, Williams 82) genotypes through genomic and cDNA cloning. 86

Figure 3.4.1.1: Production of transgenic hairy roots in Williams 82 overexpressing *HI4'OMT*/*I4'OMT-H1-H4*. **(a)** Schematic representation of the construct used for overexpression of a green fluorescent protein (*GFP*) reporter gene. '35S-Pro' and 'NOS-ter' represent the CaMV 35S promoter and the NOS terminator, respectively. **(b)** Schematic representation of the construct used for coexpression of a green fluorescent protein (*GFP*) reporter gene and *HI4'OMT*/*I4'OMT-H1-H4* or *GUS*. '35S-Pro' and 'NOS-ter' represent the CaMV 35S promoter and the NOS terminator, respectively. **(c-d)** Expression of *HI4'OMT*/*I4'OMT-H1-H4* in control and expression vector containing transgenic hairy roots from the 'Williams 82' genetic background in SCN2 and SCN5 races respectively. '*HI4'OMT*/*I4'OMT-H1-H4*' represents transgenic hairy roots containing *HI4'OMT*/*I4'OMT-H1-H4* transgene. '*VC*' represents transgenic hairy roots expressing *GFP* gene only and *GUS* representing the transgenic hairy roots with *GUS* gene and *GFP* marker gene. qRT-PCR was performed with *HI4'OMT*/*I4'OMT* gene specific primers. Expression values were normalized to the expression levels of the soybean ubiquitin-1 gene (*GmUB1*) in respective samples. The level of *HI4'OMT*/*I4'OMT* expression in the '*VC*' (*GFP*) transgenic hairy roots of both races (SCN2 and SCN5) was arbitrarily set at 1.0. Each bar represents the mean relative expression level of five independent biological replicates with standard errors of mean. The bar with asterisks is significantly different (unpaired student's *t*-test $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$). 90

Figure 3.4.2.1: Relative transcript abundance of **(a)** *CHS7*, **(b)** *CHS8*, **(c)** *GmIFS1*, **(d)** *GmIFS2*, **(e)** *GmIFR* in transgenic hairy roots containing *HI4'OMT*/*I4'OMT-H1-H4* transgenes and green fluorescent protein reporter gene and GUS vector controls (*VC*s) produced in the 'Williams 82' background was analyzed. qRT-PCR was performed with *HI4'OMT*/*I4'OMT* specific primers. Expression values were normalized to the expression levels of the soybean ubiquitin-1 gene (*GmUB1*) in respective samples. The relative transcript abundance of *VC* (with GFP gene only) transgenic hairy roots without *HI4'OMT*/*I4'OMT-H1-H4* and *GUS* constructs was arbitrarily set at 1.0. Each bar represents the mean relative expression levels of five biological replicates with standard errors of mean. The bar with asterisks is significantly different (unpaired student's *t*-test $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$; $^{****}P \leq 0.0001$). 93

Figure 3.4.3.1: SCN performance in transgenic hairy roots. **(a)** Representative stages of SCN in the hairy roots. Bars = 250 μm. Six types of transgenic hairy 96

roots from susceptible cultivar 'Williams 82' genetic background was compared and analyzed. '*HI4'OMT*/*I4'OMT-H1-H4*' represents transgenic hairy roots containing the *HI4'OMT*/*I4'OMT-H1-H4* candidate gene's alleles with '*GFP*' gene for coexpression. '*VC*' represents the transgenic hairy roots produced using a control vector '*GFP*' gene only and '*GUS*' with '*GFP*' for positive control. Each bar represents the **(b-c)** average of resistance index, normalized ratio of the nematodes developed beyond J2 (J3+J4+adult) stage to total number of nematodes per transgenic hairy root for SCN2 and SCN5 races respectively and, **(d-e)** the average of total number of cysts counted on independent transgenic hairy roots in SCN2 and SCN5 races with standard error of means. Bars with asterisks are significantly different (unpaired student's t-test, $^{*}P \leq 0.05$ $^{**}P \leq 0.01$; $^{***}P \leq 0.001$).

Figure 4.1.2.1: Proposed model for broad-spectrum resistance mechanism in resistant (S54) wild soybean genotype. During resistance response conjugates might directly release at feeding sites without being used as substrate in Kreb cycle for energy production. After their release at feeding site they might either convert back into the free phenolic acid forms or use in their conjugated form which both can be toxic to the growing nematodes. 105

LIST OF ABBREVIATIONS

SCN = Soybean Cyst Nematode

RNA = Ribonucleic Acid

DNA = Deoxy ribonucleic acid

cDNA = complementary Deoxyribonucleic Acid

SNP = Single Nucleotide Polymorphism

bp = base pair

Kb = Kilo bases

°C = Celsius

PCR = Polymerase Chain Reaction

dNTP = Deoxynucleoside Triphosphate

sec = Seconds

TAE = Tris-acetate-EDTA

EDTA = Ethylene diamine tetraacetic acid

w/v = Weight by volume

v/v = Volume by volume

nm = Nanometers

gm = Milligrams

µL = Microliters

mM = Millimolar

ng = Nanograms

LB = Liquid Broth

RPM = Revolutions per minute

mL = Milliliter

dpi = days post infection

FI = Female Index

h = hour

ANOVA= Analysis of Variance

PMN = Plant metabolic network

HCL = Euclidean hierarchical clustering

mRNA = Messenger ribonucleic acid

PAL = Phenylalanine ammonia lyase

V = Volume

cm = Centimeter

qRT-PCT = Quantitative real time polymerase chain reaction

ΔCt = Delta cycle threshold

cDEG = Commonly induced genes

cDEM = Commonly differentially expressed metabolite

CHAPTER 1: INTRODUCTION

1.1 One of the most significant challenges of the 21st century is to discover innovative ways of increasing global crop production to meet the demands for food from the growing human population (Bodirsky *et al*., 2015). A major roadblock to global food sufficiency is persistent loss of staple crops to pathogen infections (Koenning *et al*., 2010). Greater efforts are needed to accelerate the buildup of a comprehensive knowledge base that should explain how plant diseases occur and how plants defend against microbial pathogens. Soybean (*Glycine max*), is a major global crop as it is vital protein source and provides essential nutrients to millions of people worldwide, which have been found to be highly beneficial for human health and aid in treating life-threatening diseases. Over 1 million people rely on soybean-based food products to fulfill their daily nutrient intake requirement in the US (Gutierrez-Gonzalez *et al*., 2010 and Chennupati *et al*., 2012). It is also a major ingredient in livestock feed. Furthermore, the Food and Agriculture Organization (FAO) predicts that soybean production needs to increase at least 2.5-fold to feed rapidly growing human population in the near future (Lin *et al*., 2013). However, meeting this demand is challenging due to climate change, anthropogenic effects, and specifically due to pest prevalence (Lin *et al*., 2013 and Zhang *et al*., 2017). My research primarily focuses on the formidable soybean cyst nematode (SCN, *Heterodera glycine*) pathogen, an adversary of soybeans'-that threaten growth and causes ~1 billion dollars of crop yield loss in U.S. (Figure 1.1.1; Koenning *et al*., 2010 and Zhang *et al*., 2016).

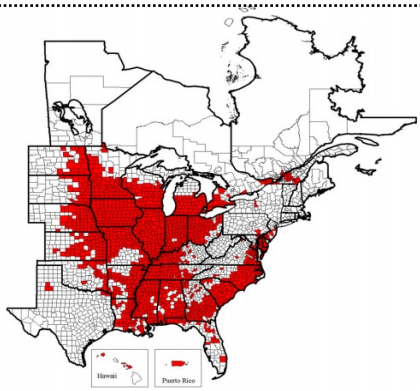


Figure 1.1.1: Map of the known distribution of the soybean cyst nematode, in the United States and Canada from 1954 to 2017. Known infested counties are indicated in red. Figure resource: Tylka and Marett, 2017;

Using SCN-resistant soybean cultivars is the most efficient strategy in managing SCN damage; however, developing them has been very challenging because: a) SCN populations evolve rapidly; b) soybean varieties often show race-specific resistance; and c) soybean cultivars possess less genome wide diversity as a result of the domestication process. Thus, it is vital to develop new soybean varieties with increased SCN resistance to enhance soybean productivity and economic viability at the local, national, and global scale.

Plant domestication is an evolutionary process in which humans have used wild species to develop new and altered forms of plants with morphological or physiological traits that meet human needs. Typically, limited numbers of individuals of progenitor species were used by early farmers and the traits selected usually were related to overall yield, harvesting, and edibility (Konishi *et al*., 2006 and Hua *et al*., 2015). As a consequence, this strong selection process produced genetic bottlenecks of varying degrees

that have resulted in a heterogeneous reduction in the level of genetic variation among annual herbaceous crops (Figure 1.1.2; Buckler *et al*., 2001; Meyer *et al*., 2012 and Miller and Gross, 2011). The domestication process has resulted in reduced diversity at both the genome and local levels. For example, more than half of the genetic variation has been lost in cultivated soybean (Hyten *et al*., 2006 and Zhou *et al*., 2015), 2–4% of maize genes experienced artificial selection (Wright *et al*., 2005), and genetic diversity has been significantly reduced in cultivated rice (Xu *et al*., 2012), as compared with its wild counterpart. In addition, positive selection on a target locus controlling a domestication trait of interest can also result in a reduction in the diversity of closely linked loci (selective sweep). In fact, many of these selective sweeps have been found at previously reported quantitative trait loci (QTL) associated with domesticated traits. Examples include an approximately 90-kb selective sweep at the promoter region of axillary-branch formation-related gene *tb1* in maize (Clark *et al*., 2004), a 600-kb sweep at kernel-related gene *Y1* in maize (Palaisa *et al*., 2004), and a 260-kb sweep at amylose-related gene *waxy* in rice (Sweeney and Mccouch, 2007).

These results suggest that domestication has reduced or eliminated genetic diversity at certain loci in modern crops including soybeans and limited their potential to develop novel varieties with improved traits. In contrast, crop wild relatives (CWRs) retain high levels of genetic diversity compared to their domesticated descendants (Figure 1.1.2). There are two ways in which CWRs have been defined. One is the gene pool (GP) concept proposed by Harlan and de Wet (1971), where CWRs were classified into groups (GP-1 to GP-3) based on the relative ease of gene exchange with cultivated crops. Gene exchange occurs relatively easily between primary (GP-1) and secondary GPs (GP-2) by crossing

(and fertile hybrids can be produced), whereas gene transfer between primary and tertiary (GP-3) groups is usually difficult. Even though the CWRs that have been used in crop improvement mostly belong to GP-1 (Munns *et al*., 2012) or GP-2 categories (Fetch *et al*., 2009 and Saintenac *et al*., 2013), there are some examples where useful alleles from distant wild relatives, such as GP-3 plants, have been successfully transferred for crop improvement (Abedinia *et al*., 2000 and Marais *et al*., 2003).

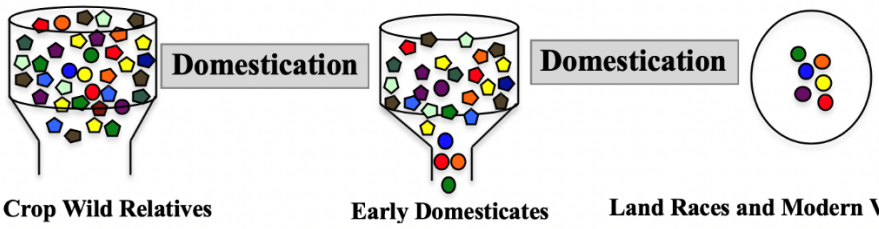


Figure 1.1.2: The decrease in genetic diversity in modern crops during domestication due to bottleneck events.

The second concept for CWRs is that of the taxon group (TG), a system that is based on the ranking of the taxonomic hierarchy of crops (Maxted *et al*., 2006). A taxon group (TG) may include a wide range of wild species that may be evolutionarily closely or distantly related to crop species within the same genus. With this concept, CWRs were defined in a range from TG1 (same species as the crop) to TG4 (different species within the same genus as the crop). Crop wild relatives are widely distributed on all continents except for Antarctica, and many are located in Vavilov centers of diversity and adjacent regions (Maxted and Kell, 2009; Larson *et al*., 2014 and Castaneda-Alvarez *et al*., 2016). The global distribution of CWRs suggests that there are ample resources to be explored for use in plant breeding. In fact, among approximately 50,000–60,000 total crop and CWR species, it has been estimated that 10,739 species (or even more) have a direct value for food security (Maxted and Kell, 2009). Although the number of publications discussing

the use of CWRs in breeding has increased over the years and the use of CWR for crop improvement has been gradually recognized (Maxted and Kell, 2009), the exploration and utilization of the genetic diversity contained in wild relatives has lagged considerably.

1.2 Crop Wild Relatives (CWRs) in Crop Improvement

Numerous efforts have been made to utilize the genetic diversity in CWRs to improve various crop species (Tanksley *et al*., 1996; Hajjar and Hodgkin, 2007; Nevo and Chen, 2010 and Maxted, Magos, and Kell, 2013). These efforts have been concentrated primarily on certain crop species, including wheat, barley, rice, and tomatoes (Xiao *et al*., 1996; Nevo and Chen, 2010; and Foolad and Panthee, 2012). Possible reasons for the greater use of CWRs in only certain crops include (i) cross-compatibilities, (ii) the taxonomic relationship between crops and their close wild species, (iii) fertility in the F1 and subsequent progeny, (iv) availability or conservation of CWRs, (v) exploration and utilization of wild germplasms, and (vi) regional financial support based on local need and geographic distribution of CWRs (Zamir, 2001). Because of these and other reasons, the use of CWRs lags far behind its potential.

1.2.1 Biotic stress resistance

Plant yields are significantly reduced by attacks from various biotic stressors such as pathogens (fungi, viruses, and bacteria), nematodes, and insect pests. Even though resistant varieties have been developed, continuous use of limited numbers of resistant resources is not a long-term strategy because pathogens and insects evolve very rapidly. To mitigate this evolutionary arms race, plant researchers and breeders have been exploiting exotic genetic resources, such as CWRs, to develop biotic stress-resistant varieties (Hajjar and

Hodgkin, 2007). These efforts include attempts to identify resistant genes in CWRs using various strategies such as metabolomics and transcriptomics. Once identified, the intent then would be to pyramid multiple exotic resistant genes into crop varieties to achieve a durable or broad-spectrum resistance.

In maize, the corn blight of 1978 reduced the yield of corn by as much as 50% in the United States (Food and Agriculture Organization of the United Nations, 2005). This was resolved by transferring blight-resistant alleles from a wild relative of Mexican maize (*Tripsacum dactyloides* L.) into commercial corn lines (Maxted and Kell, 2009). Another devastating pest of corn in the United States is rootworm. Prischmann et al. (2009) introduced genes from gamagrass (*Tripsacum dactyloides* L.), a wild relative of maize that exhibits rootworm resistance, into cultivated corn. Repeated field trials showed that the descendants from this transfer appeared tolerant to rootworm damage, showing that this exotic allele was effective in combating rootworm.

Using wild relatives to improve biotic stress tolerance in cultivated rice has been very successful. A majority of the 22 wild rice species are being explored as alternate sources for resistance to bacterial blight, blast, brown plant hopper attacks, and sheath blight (Jena, 2010). For example, bacterial leaf blight, caused by *Xanthomonas oryzae* pv. *oryzae* (*Xoo*), has been one of the most widely distributed and devastating rice diseases worldwide. The introduction of two resistant *Xa* genes (*Xa3* and *Xa*4) into rice cultivars has increased bacterial leaf blight resistance. However, the level of resistance in these cultivars has been decreasing as expected due to evolutionary changes in bacterial leaf blight. To address this, researchers have identified new bacterial blight - resistant genes (*Xa21* and *Xa23*) in wild rice (Song *et al*., 1995 and Zhou *et al*., 2011). These *Xa* genes

have been used individually or in combination in bacterial leaf blight-resistant rice breeding programs (Zhou *et al*., 2009), and this has led to significant successes in bacterial leaf blight management. Progress in combating bacterial leaf blight also has been achieved from introgression of a rice blast-resistant gene, *Pi33*, from wild rice, *Oryza rufipogon* (Ballini *et al*., 2007), into the most used rice blast resistance variety (IR64).

The improvement of biotic resistant cultivated tomatoes (*Solanum lycopersicum*) has also significantly benefitted the transfer of various traits from tomato wild relatives. These traits include resistance to bacteria, viruses, fungi, nematodes, and insect pests. The tomato wild relatives that have been used include *Solanum chilense* (Zamir *et al*., 1994), *Solanum habrochaites* (Prasanna, Kashyap, *et al*., 2015; Prasanna, Sinha, *et al*., 2015), *Solanum peruvianum* (Seah *et al*., 2004 and Lanfermeijer *et al*., 2005)*, Solanum pennellii* (Parniske *et al*., 1999), and *Solanum pimpinellifolium* (Chunwongse *et al*., 2002). In one example, five *Ty* genes exhibiting varying degrees of resistance to *tomato yellow leaf curl virus* (TYLCY) were successfully introgressed into cultivated varieties (Ji *et al*., 2007 and Menda *et al*., 2014). Pyramiding of these *Ty* genes from different wild tomatoes has contributed to durable and broad resistance to TYLCY (Vidavski *et al*., 2008 and Kumar *et al*., 2014).

1.2.2 Abiotic stress tolerance

Salinity and drought are two of the most important environmental factors limiting worldwide crop yields. The effects of both stressors have been intensely studied in primary crop species such as soybeans, tomatoes, and cereals as well as in their wild relatives (Munns *et al*., 2012; Placido *et al*., 2013 and Qi *et al*., 2014). Here, we discuss several

examples of the use of wild relatives for salt tolerance in wheat and soybeans, and for drought tolerance in barley.

In 2013, wheat was the third-most-produced cereal after maize and rice ("FAOStat," retrieved January 27, 2015), and it is significantly affected by salinity (Mujeeb-Kazi and De Leon, 2002). Intensive efforts have been made to search for salt-tolerant genes/alleles in wheat wild relatives (Colmer *et al*., 2006 and Nevo and Chen, 2010). Recently, Australian scientists have produced a salt-tolerant commercial durum wheat variety by introducing an allele from its wild relative, *Triticum monococcum*, via crossbreeding. This cultivar showed 25% greater yield in high-saline fields compared to its Tamaroi parent (Munns *et al*., 2012). The gene transferred was *TmHKT1; 5-A*, and has been found in a wild wheat ancestor, *Triticum monococcum* (James *et al*., 2006). *TmHKT1; 5-A* reduces the Na+ level in plant leaves that prevents yield losses under salinity stress. This gene transfer has provided a successful example of the use of a CWR to improve crop salt tolerance. This gene could also be transferred to other wheat cultivars, such as bread wheat, to develop salt-tolerant commercial lines. Soybean is moderately salt-sensitive, and all developmental stages can be affected by salinity stress (Munns and Tester, 2008) that can result in a decrease in yield upto 40% (Chang *et al*., 1994). A salt-tolerant gene, *GmCHX1,* was recently identified in *Glycine soja* (Sieb. &Zucc), the wild progenitor of cultivated soybean (*Glycine max* (L.) Merr.) (Qi *et al*., 2014). The discovery of *GmCHX1* and another salt-tolerant gene, *GmSALT3* (Guan *et al*., 2014), suggests that salt-tolerant alleles might have been lost in soybean during domestication. Both of these salt-tolerant genes are involved in regulating ion homeostasis and offer great potential for the development of commercial soybean varieties with improved salt tolerance.

Drought is another important factor limiting crop production, especially in the context of global climate change. Although several drought-tolerant QTL or genes have been identified in wild relatives of crops such as barley (Diab *et al*., 2004 and Suprunova *et al*., 2004, 2007), wheat (Placido *et al*., 2013), and tomatoes (Fischer *et al*., 2013), applications of these exotic genetic resources for improved drought tolerance in crops have not proven as successful as expected. This may largely be because drought tolerance is a polygenic quantitative trait presumably controlled by many QTL, each with small effects. On the other hand, the development of advanced backcross introgression libraries (ILs) provides a useful alternative method for the transfer of drought-tolerant genes. Recently, significant progress was made on barley drought tolerance using wild barley (*Hordeum spontaneum*)-introgressed ILs (von Korff *et al*., 2004; Schmalenbach *et al*., 2008; Honsdorf *et al*., 2014 and Naz *et al*., 2014). Follow-up studies demonstrated that the ILs containing certain drought-tolerant QTL exhibited drought tolerance in field trials (Honsdorf *et al*., 2014 and Arbelaez *et al*., 2015). However, construction of ILs and exclusion of "linkage drag" genes require years of backcrossing and selection. Despite these difficulties, this approach may be our best strategy if no other effective breeding alternatives are available.

1.2.3 Improvement in yield and quality-related traits

Compared with the great success of introducing biotic and abiotic stress resistance genes from CWRs for crop improvement (Colmer *et al*., 2006; Du *et al*., 2009 and Munns *et al*., 2012), applications of CWRs for crop yield traits have been relatively less successful. There are exceptions, however, in which the magnitude of CWR contributions to yield improvement may be even greater than expected. For example, the use of a small-fruited

tomato ancestor (*S. pimpinellifolium*; Eshed and Zamir, 1994) and a wild tomato species (*Solanum hirsutum*) led to a 20% increase in yield and soluble solids content, and an improvement in fruit color, in cultivated tomatoes (Tanksley *et al*., 1996). Similarly, in rice, backcrossing a low-yielding wild ancestor (*O. rufipogon*) into a Chinese hybrid strain increased its yield up to 17% (Tanksley and McCouch, 1997 and Xie *et al*., 2008). Yield-enhancing QTL have also been identified in wild relatives of diverse crops such as wheat, barley, soybeans, beans, and *Capsicum* (Swamy and Sarla, 2008), and it will be interesting to see how effective they might be when transferred into crops by traditional breeding or advanced biotechnological approaches.

It has been reported that the transfer of a QTL from *G. soja* into domesticated soybeans significantly increased the yield by about 190–235 kg/ha (Li *et al*., 2007). Although introducing *G. soja* QTL into cultivated soybean lines of diverse parentages has often resulted in some unfavorable characteristics such as lodging susceptibility (Li *et al*., 2007), the improvement in yield and/or other important traits generally has made these efforts worthwhile. The role of CWRs in the development of new varieties with improved fruit and grain quality has become increasingly important in recent years. The introduction of newly developed nutrition-rich varieties is expected to play a role in the improvement of human health. For example, glucosinolates, a class of metabolites prevalent in crucifer species, are known to reduce the risk of various cancers (Dinkova- Kostova and Kostov, 2012). These metabolites can easily be increased in some food plants such as broccoli where crosses between wild and cultivated broccoli produce a hybrid that contains three times more glucosinolates than conventional varieties (Sarikamis *et al*., 2006).

As an alternative to targeting only yield traits, researchers have incorporated abiotic and biotic stress-tolerant traits from CWRs to achieve crop yield enhancement. Typically, crop varieties introduced with wild resistant/tolerant alleles behave similarly compared to their cultivated counterparts but benefit when exposed to environmental stresses. For example, field trials of the wheat cultivar Tamaroi carrying the wild salt-tolerant gene *TmHKT1; 5-A* produced yields similar to the Tamaroi parent lacking this gene when grown in fields with less saline but produced superior yields in high-saline fields (Munns *et al*., 2012). In a recent study, tomato cultivars carrying a single wild-derived TYLCY-resistant gene, *Ty-2* or *Ty-3,* or their pyramided lines, had greater yields in the presence of the pathogen TYLCY than cultivars lacking these genes (Prasanna, Kashyap, *et al*., 2015). The yield increases in these cases mostly are attributable to the enhanced resistance/tolerance to biotic or abiotic stresses generated by introducing the resistant/tolerant alleles from their wild relatives. Thus, the recruitment of stress resistance traits from wild relatives into modern varieties is a practical alternative for the improvement of crop yields.

1.3 Advanced Biotechnologies Accelerate the Use of Wild Relatives for Crop Improvements

The surge of diverse biotechnologies has significantly facilitated modern breeding over the past two decades (Varshney *et al*., 2005). These technologies include the application of omics-scale technologies for gene discovery, and advanced techniques to transfer genes of interest from wild plant species to cultivated crops (Table 1.3.1).

Genomic approaches have been widely used to identify genes or genomic regions controlling complex traits. High-throughput next-generation sequencing technologies offer

opportunities to efficiently discover SNPs associated with important traits in both diploid (Hyten *et al*., 2010) and polyploid plant species (Akhunov *et al*., 2009). With recent significant cost reductions, scientists are now able to genotype thousands of individuals by genotyping-by-sequencing or resequencing. With the availability of increasing numbers of SNPs and phenotypic data, researchers have been able to validate and fine-map previously identified genes and to discover novel genomic regions underlying valuable agronomic traits in CWRs by association mapping (Xu *et al*., 2012; Li *et al*., 2014; Qi *et al*., 2014 and Zhou *et al*., 2015). However, the development of a high-throughput phenotyping pipeline remains challenging, especially in the field (Kelly *et al*., 2016). Some of the genomic regions associated with domestication traits have enhanced our understanding of their genetic basis and will encourage further investigation to see whether allelic variation in those regions in CWRs can additionally benefit crop improvement. Genotyping-by-sequencing of segregating populations (F_2, BC2, near-isogenic lines, and recombinant inbred lines [RILs]) allows the construction of high- resolution linkage maps that can be used to narrow QTL regions. This strategy holds the promise of mapping QTL using fewer RILs than would be necessary with DNA markers at relatively low densities. This approach also should facilitate map-based cloning of target genes, such as the salt-tolerant gene *GmCHX1* unique to *G. soja* (Qi *et al*., 2014). It also is feasible to apply genotyping-by-sequencing for heterozygous plant species using case-specific strategies (Uitdewilligen *et al*., 2013 and Hyma *et al*., 2015). Thus, it is clear that the availability of this sort of genome-wide data and efficient phenotyping approaches will continue to accelerate the discovery of genes controlling superior traits in CWRs.

Table 1.3.1: Representative advanced technologies that have been used in plant breeding

Approaches	Usages	Advantages	Shortcomings	References
Genomics	Germplasm resource evaluation and identification; heterosis prediction; linkage and association mapping; marker-assisted breeding	High throughput; time saving	Costly; bioinformatics skills required; difficulties in assembly of polyploid genomes	Brozynska et al. (2015); Langridge and Fleury (2011)
Transcriptomics and Proteomics	Quantification of expression variants response to environment stress; updating genome annotation	Generating numerous candidate genes; regulatory network identification; more useful when combined with linkage analysis	Difficult to pinpoint causal genes or proteins; high cost for proteomics	Langridge and Fleury (2011); Brozynska et al. (2015)
Metabolomics	Metabolic profiling	Quantification of target or global metabolites	Costly; limited annotation data; low heritability; requiring chemical and statistical skills	Fernie and Schauer (2009)
Advanced introgression lines	Genetic mapping; introgression breeding	Traditional breeding; introducing multigenic traits	Need supports by molecular DNA markers; cross-compatible; laborious and tedious backcrossing	Placido et al. (2013); Honsdorf et al. (2014)
Transgenesis	GM*	Transfer between non-crossable species	Subject to GMO regulations; foreign genes	Schaart, et al. (2016)
Genome editing	GM*	Precise and predefined modification	Might subject to GM regulatory regime; public acceptance	Bortesi and Fischer (2015); Schaart et al. (2016)
Cisgenesis/intragenesis	GM*	Genes from species itself or crossable species; stacking multiple genes; public acceptable; avoid linkage drag	Might require traditional breeding step	Haverkort et al. (2009); Vanblaere et al. (2011)
High-throughput phenotyping	Phenotyping	High throughput; real-time; multidimensional	High cost; mathematical and statistical skill required	Honsdorf et al. (2014); Rahaman, et al. (2015)

*GM: Genetic modification;

Functional omics approaches, including transcriptomics, proteomics, and metabolomics, have provided alternative opportunities for global analysis of regulatory genes, expressed proteins, or metabolite candidates underlying important traits in CWRs. These omics approaches also are particularly suitable for dissection of the variation in complex traits such as drought tolerance and pest resistance. By characterizing CWRs under diverse treatments using omics strategies, a number of stress-resistant genes have been identified in various wild relatives of crops. For example, the dehydrin genes in both wild barley (*H. spontaneum*) and wild tomato species (*S. chilense* and *S. peruvianum*), as well as ABA/water stress/ripening-induced (*Asr*) gene family members (*Asr2* and *Asr4*) from wild *Solanum* species, are known to be involved in drought tolerance (Suprunova *et al*., 2004 and Fischer *et al*., 2013). From these and other studies with large "omics" datasets, it generally has been proven difficult to pinpoint the causal genes, proteins, or metabolites underlying the traits of interest. However, this is possible when "omics" approaches are combined with other strategies such as linkage mapping or molecular

biology approaches. For example, by quantifying gene expression levels within target QTL, Suprunova et al. (2007) were able to identify a novel gene (*Hsdr4*) involved in water-stress tolerance in wild barley (*H. spontaneum*). With a similar strategy, two candidate genes (*KNAT3* and *SERK1*) conferring drought tolerances in wild wheat have also been identified (Placido *et al*., 2013). Further, as knowledge of transcriptome profiles under various stress conditions increases, the combination of transcriptomics, proteomics, and metabolomics with QTL analyses should prove to be a powerful tool for large-scale study of gene function at different levels. With such an approach, however, the difference in temporal transcription and translation of genes and metabolic processes involved should be taken into account in the data interpretation (Schmollinger *et al*. 2014).

Metabolomics-assisted breeding has been considered as a viable strategy for crop improvement (Fernie and Schauer, 2009). Although the high quantification cost and the relatively low levels of heritability of metabolites limit the direct application of this approach to breeding programs, several studies have revealed its potential by quantifying the variation of certain metabolites and/or metabolomics and uncovering their genetic basis with genomic approaches (Bleeker *et al*., 2011, 2012). Bleeker et al. (2011) showed that the application of 7-epizingiberene, extracted from wild tomato (*S. habrochaites*) and applied on susceptible cultivated tomatoes, was effective in repelling whiteflies (Frelichowski and Juvik, 2001). The cultivated tomatoes with 7-epizingiberene acquired from its wild relatives also showed resistance to spider mites (Bleeker *et al*., 2012).

Genome-wide association studies (GWAS) of metabolomics have also become an effective way to investigate global profiles of the thousands of metabolites typically

produced in plants (Luo, 2015). This approach is thought to be suitable for the exploitation of CWRs because these wild species have been subjected to long-term evolution in their diverse natural habitats and therefore are expected to have a greater level of variation in their metabolic profiles than their cultivated descendants. CWRs do require a higher density of genomic markers for metabolomics association studies because they typically have much lower levels of linkage disequilibrium than are found in domesticated crops. Fortunately, many markers now are publically available or can be genotyped at a reduced cost. Several laboratories recently have successfully developed high-density SNP markers for wild soybean (Song *et al*., 2015 and Zhou *et al*., 2015), wild tomatoes (Aflitos *et al*., 2014), and wild rice (Xu *et al*., 2012), all of which can be or have been used in metabolomics mapping in CWRs. These and other studies (Schauer and Fernie, 2006 and Fernie and Schauer, 2009) suggest that useful metabolites in CWRs can be identified and used in plant breeding. One difficulty is that most of the metabolites produced from the profiling platforms are unannotated and thus are unknown. Once identified, however, metabolite pathways can be traced by searching available annotations in a metabolomics database. It also has proven feasible to transfer an appropriate metabolite or metabolic pathway (e.g. the terpenoid biosynthetic pathway) from CWRs to increase resistance to biotic stress in cultivated tomatoes and other commercial plant species (Bleeker *et al*., 2012).

Genetic modification (GM) technology has been considered a revolutionary solution to transfer target genes to crop cultivars to obtain desired traits. Commercial GM crops typically produce their target product and yield as expected, and they have the advantage of not suffering from the introduction of other linked genes (linkage drag).

Genetic engineering techniques are particularly useful when the desired trait is not present in the germplasm of the crop or when the trait is very difficult to improve by conventional breeding methods. A well-known example of the use of this technology was in the production of transgenic *Bacillus thuringiensis* (*Bt*) crops (Tabashnik, 2010). However, there is some evidence that the use of GM technology to manipulate drought tolerance might not be as effective as traditional breeding methods (Gilbert, 2014), including introgression (Honsdorf *et al*., 2014). In addition, the safety of foods developed from transgenic crops continues to remain a concern to the public.

Another approach to facilitate crop production is to induce mutations in existing genes rather than introduce new genes (Lusser *et al*., 2012). This approach includes cisgenesis, intragenesis, genome editing, RNA-dependent DNA methylation, and oligo-directed mutagenesis techniques, some of which have been used for crop improvement using CWR species (Table 1.3.1). Cisgenesis, for example, refers to the GM of crop plants with genes from the crop plant itself or from a sexually compatible donor such as a CWR. This technique has been successfully used to confer resistance to late blight in potatoes (Haverkort *et al*., 2009) and scab resistance in apples (Vanblaere *et al*., 2011). It is important to note that because cisgenesis only transfers a gene from a native or cross-compatible species, this results in plants with a performance comparable to that possible from conventional breeding (Krens *et al*., 2015). However, transfer of a single desired gene by cisgenesis avoids any linkage drag and reduces the time involved compared with traditional breeding strategies. Although these new techniques outperform traditional breeding in some respects and are easily adopted by the industry, their application to the production of a wide range of commercial crops depends on many factors such as the

technical efficiency of some processes, the extent of social acceptance, and worldwide regulatory restrictions. Nevertheless, these technologies are useful in accelerating gene introduction and producing plants with desired traits and could be used as an alternative strategy for crop improvement in the absence of other efficient approaches.

High-throughput genomic approaches have become an important tool in efficient use of crop genetic resources, including CWRs, deposited in gene banks. High-throughput genotyping of stored accessions allows the examination of genetic relationships, which enable breeders to effectively select the accessions of interest based on their genetic background (Kadam *et al*., 2016). Phylogenetic trees generated from these genomic data also allow the construction of core germplasm collections (Brown, 1989) representing the allelic richness of the gene bank. Although large-scale phenotyping remains time-consuming and costly, high-throughput phenotyping of the representative core collection for the traits of agronomic importance is less laborious than characterizing the entire collection (Honsdorf *et al*., 2014). Consequently, genomic estimated breeding values for the core collection facilitate genomic selection of superior CWR accessions (Xavier *et al*., 2016).

1.4 Objective(s) of present study

Wild soybean (*Glycine soja*) is the soybean wild progenitor. It harbors a much higher level of genetic diversity (Figure 1.1.2) then soybean because wild soybeans have been challenged in natural environments for thousands of years (Figure 1.4.1a). For some reasons, however, most contemporary investigations into soybean-SCN interactions at the molecular level devote on one type of SCN race using soybean cultivar and relatively little

effort has been made to understand the common resistance mechanism for multiple SCN strains using its wild relatives. In this project, for the first time, we had used multiple approaches including omics and molecular biology techniques (Figure 1.4.1b) in wild soybean genotypes to investigate the broad-spectrum resistance mechanism(s) for two SCN races (SCN5, HG type 2.5.7 and SCN2, HG type 1.2.5.7), which are less studied and prevalent in the Southeastern region.

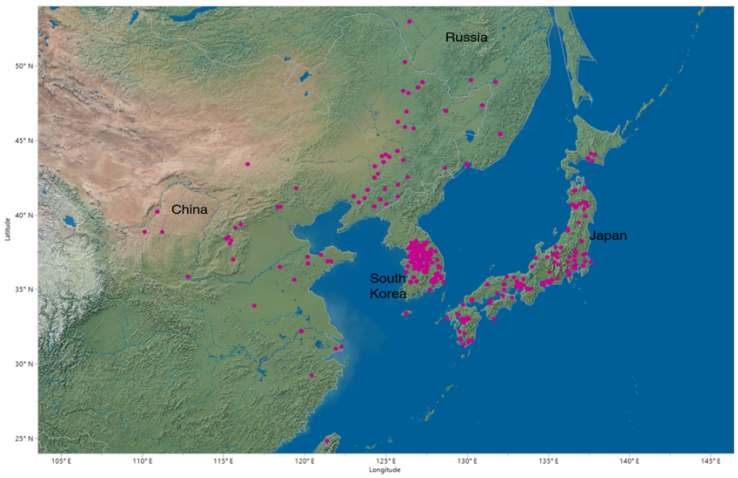


Figure 1.4.1a: Geographic distribution of wild soybeans in East Asia; Figure resource: Kofsky et al. (2018).

To accomplish the long-term goal, the following objectives have been selected for the present study:

- Analysis of metabolomic variation in wild soybean SCN resistant (S54) and non-resistant (S67) genotypes for identification of significant metabolites and pathways involved in broad-spectrum nematode resistance.
- Selection of significant candidate genes which might confer SCN resistance, based on metabolomics comparison, gene expression and molecular analysis.

- Functional validation of identified SCN resistant gene(s) through hairy root transformation to show the role of selected genes for broad spectrum nematode resistance.

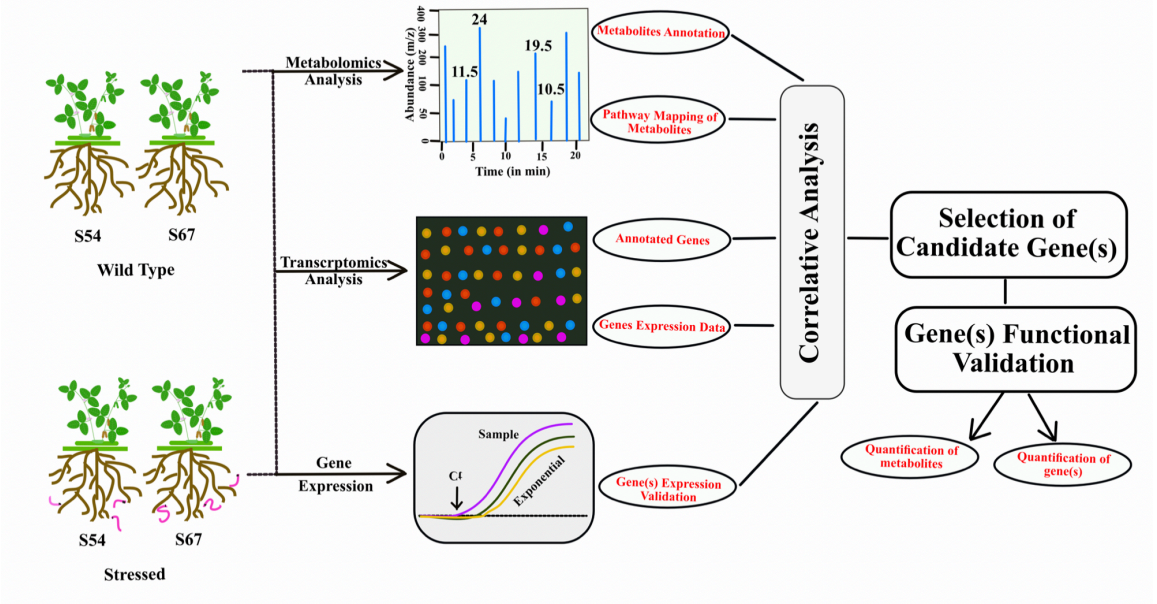


Figure 1.4.1b: Figure representing the different omics based and molecular techniques which has been selected for the present study;

CHAPTER 2: MATERIALS AND METHODS

2.1 Untargeted Metabolomics Analysis

2.1.1 Plant materials and soybean cyst nematode (SCN) stocks

Metabolomics experiments have been carried out on two previously identified wild soybean (*G. soja*) genotypes in our laboratory: S54 (FI < 10%) and S67 (FI > 100%), which were found highly resistant and susceptible respectively to two SCN races (Zhang *et al*., 2016) i.e., SCN2 (HG type 1.2.5.7) and SCN5 [(HG type 2.5.7) (Jiao *et al*., 2015)]. Seeds of both genotypes (S54 and S67) were obtained from USDA Soybean Germplasm Collection- Grin Global (www.ars-grin.gov/). Both SCN races (HG type 1.2.5.7; SCN2 and HG type 2.5.7; SCN5) were used in the study. They were separately reared on soybean cv. Williams 82 grown in clay pots filled with sand in the greenhouse under controlled conditions (27 °C, 16h light/ 8h dark) for more than 30 generations. These races were originally obtained from Dr. Erik (Rick) Davis, Department of Plant Pathology, North Carolina State University, Raleigh, NC, USA in the year of 2015.

2.1.2 Plants preparation, SCN inoculation, and tissue collection

Seed preparation, germination, transplanting, and SCN inoculation were performed as described by Zhang et al. (2017) and Zhang et al. (2017). Briefly, seeds of S54 and S67 genotypes were surface sterilized with 0.5% sodium hypochlorite for 60 seconds, rinsed with autoclaved water, and then placed on a piece of wet sterile filter paper in a petri dish for germination at 27 °C in the growth chamber (Intellus Control System, Percival Scientific Inc., Perry, Iowa). After 2-3 days, each healthy seedling was transplanted into a

cone-tainer (Greenhouse Megastore, Danville, IL, USA) filled with sterile sand. Three days after transplantation, healthy seedlings were used for inoculation and sample collection for untargeted metabolomics. A randomized complete block design was used for cone-contained seedlings arrangement.

For nematodes preparation, cysts were collected from stock roots that were maintained in the greenhouse at UNC Charlotte. Briefly, SCN cysts were harvested by massaging the stock roots in water and sieving the solution through nested 850 and 250 µm test sieves (Fisher Scientific, Suwanee, GA, USA). The collected cysts were crushed with a rubber stopper in a 250-µm sieve, the released eggs were collected in a 25-µm mesh sieve (Klink *et al*., 2007). The eggs were then purified by sucrose flotation protocol (Matthews *et al*., 2003) with some modifications. For nematode hatch, purified eggs were placed on a wet paper tissue in a plastic tray with an appropriate level of water, which was covered with aluminum foil in an incubator maintaining at 27 °C for 3 days. Hatched second-stage juvenile nematode (J2) were collected and suspended in 0.09% liquid agarose at a final concentration of 1,800 J2/ml. For inoculation, 1 ml of J2 inoculum was added on each root as treatment, and seedlings inoculated with 0.09% agarose were used as non-infected controls. Three days post-inoculation, three roots of randomly selected seedlings were stained with acid fuchsin to validate the successful inoculation and investigate the growth of nematode in the roots as previously described Bybd et al. (1983) and Klink et al. (2007). Root tissues were separately prepared for LC-MS and gene expression analysis and sampled at 3, 5, and 8 days post-inoculation (dpi). For gene expression analysis [Zhang et al. (2017) and Zhang et al. (2017)], root tissues from three individuals were pooled as one biological replicate and three replicates were collected for each condition. For LC-MS

analysis, roots from 3-5 plants per condition were collected, weighed, and analyzed individually. All samples were immediately freeze in liquid nitrogen and stored at -80 °C until use.

2.1.3 Metabolite extraction, and data processing

Metabolites were extracted from each root tissue and the resulting data were processed as previously described by Strauch et al. (2015). Briefly, root tissues were extracted with 50% (v/v) methanol at 60 °C water bath for 30 min. A tissue-to-solvent ratio (w/v) of 1:10 was constantly used across all of the samples. Extracts were filtered using 0.2-µm filter prior to LC-MS profiling on a G6530A Q-TOF LC/MS (Agilent Technologies, USA). Peaks consistently detected in at least three biological replicates within each group at each time point were used for downstream analysis. Few metabolites identities including daidzein, daidzin, genistein, genistin, formononetin, 3-hydroxybenzoic acid, 4-hydroxybenzoic acid, 2, 3-dihydroxybenzoic acid, and 4-hydroxybenzaldehyde, in untargeted metabolomics were confirmed by using a pool of pure standard compounds (Annexure I; Table 2.1.3.1). All the remaining metabolites from untargeted metabolomics were identified and annotated according to the mass match with Kyoto Encyclopedia of Genes and Genomes [(KEGG) (Kanehisa *et al*., 2016)], Plant Metabolic Network (PMN) at SoyCyc database (https://www.plantcyc.org/), and an in-house database.

2.1.4 Statistical analysis

Untargeted metabolomics data were normalized and analyzed using MetaboAnalyst program (an R based online tool for metabolomics data analysis; Xia *et al*., 2015). Briefly, cube root transformation was conducted for all numerical data transformed from the peak areas for data normalization and Pareto scaling in an effort to reduce the undesired

biasedness and to show the biologically relevant differences for each compound among different groups (Alonso *et al*., 2015). One-way ANOVA with Fisher's LSD post-hoc analysis (FDR < 5%) was performed using normalized peak area to determine the expression of metabolite(s) among four groups (S54_T, S54_C, S67_T, and S67_C) at each time point. The metabolites significantly inducted in S54_T compared with the other three groups with the challenges by both races (SCN2 and SCN5) were considered as the metabolites showing potential broad-spectrum resistance. The metabolites that were strongly induced in both treatments and controls of one genotype at all three-time points but not occurred in the other genotype were regarding as constitutively expressed metabolites. For data visualization, principal component analysis (PCA) was conducted with R package *princomp* using all detected high-confident metabolites. Heat maps illustrating the comparison of expressed metabolites in the form of fold change were made using JMP pro 13 (SAS Institute Inc., Cary, NC). Biological databases including KEGG (http://www.genome.jp/kegg/), Metacyc (https://metacyc.org/) and Soycyc on plant metabolic network (PMN; www.plantcyc.org) were used as a reference for the mapping of different metabolites into their corresponding pathways.

2.2 Quantitative Expression Analysis of Phenyl-Propanoid Pathway Genes

2.2.1 Tissue collection and RNA isolation

Root tissues were sampled after 3, 5, and 8 days of SCN post-inoculation (dpi) from resistant (S54) and susceptible (S67) wild soybean genotypes in liquid nitrogen and stored at -80 °C as described in the section 2.1.2. Total RNA was extracted from different roots by means of RNeasy Plant Mini Kit® (QIAGEN, CA, USA). Liquid nitrogen and

autoclaved pestle were used to crush the roots and to make powder in the collection tube. Buffer RLC and β-mercaptoethanol was added to the powder and transferred to a QIAshredder spin column and centrifuged (Eppendorf, Hamburg, Germany). Rest of the RNA extraction was carried out using manufacturer's instruction (RNeasy Plant Mini Kit® QIAGEN, CA, USA).

2.2.2 Quantitative estimation of extracted RNA

The concentration of the extracted RNA from susceptible and resistant genotypes was quantified in NanoDrop™ 2000 spectrophotometer (Thermo Scientific, MA, USA). The purity ratio of the extracted RNA was measured at A_{260}/A_{280} nm wavelength and the concentration was measured in ng/µL using spectrophotometer built in program. The purity ratio of A_{260}/A_{280} from 1.8-2.0 and the concentrations above 100 ng/µL were retained for the further analysis.

2.2.3 Qualitative analysis of extracted RNA

The quality of extracted RNA was checked on 1.2% (w/v) agarose (GeneMate™ Low Melt Agarose, Kaysville, UT) gel using 1X TAE buffer (Annexure I; Table 2.2.3.1) stained with SYBR® safe DNA gel stain (1X, 40 mM Tris-acetate, 1 mM EDTA, pH ~8.3, Thermo Scientific, MA, USA). The isolated RNA (3 µL; total RNA 200 ng) with 3 µL of 6X loading dye (bromophenol blue, xylene cyanol FF, Orange G, Ficoll 400, Tris-HCl and EDTA; Promega, Madison, WI, USA) was loaded on to the B1A Mini Gel Electrophoresis system (Thermo Scientific, MA, USA). The gel was run at the constant voltage of 5 V/cm of the length of gel till the bands separated. After 1/4th running of the gel it was photographed with Biospectrum AC® imaging system (UVP, Upland, CA, USA). The

size of the amplified band was compared with 1 kb DNA molecular weight marker (Thermo Scientific Waltham, MA, USA), which was used as marker to know the size of the extracted RNA fragments (28s rRNA and 18s rRNA).

2.2.4 Reverse Transcription PCR for cDNA Synthesis

2.2.4.1 DNase I enzyme treatment

DNase I enzyme treatment was carried out in 10 µL reaction mixture, containing 3 µL nuclease free water, 1.0 µL of 10X reaction buffer with $MgCl_2$ (100 mM Tris-HCl (pH 7.5 at 25 °C), 25 mM $MgCl_2$, 1 mM $CaCl_2$), 1 µL RNase free DNase I enzyme [(1U/µL) (Thermo Scientific, MA, USA)] and 5 µL template RNA. Enzyme treatment was performed in a PCR thermocycler (Eppendorf, Hamburg, Germany) at 37 ˚C for 30 minutes. After 30 minutes of incubation, 1 µL of 50 mM EDTA was added to the reaction mixture. The samples were then incubated for an additional 10 minutes at 65 ˚C. After enzyme treatment, the DNA contamination free RNA samples were used for first strand cDNA synthesis.

2.2.4.2 First strand cDNA synthesis

The cDNA was synthesized by using high capacity cDNA reverse transcription kit (RevertAid RT kit, Thermo Scientific, MA, USA) following manufacturer's instructions. cDNA synthesis was carried out in a 20 µL reaction mixture, containing 11 µL of DNase I treated template RNA, 1 µL of random hexamer primer, 4 µL of 5X reaction buffer (250 mM Tris-HCl (pH 8.3), 250 mM KCl, 20 mM $MgCl_2$, 50 mM DTT), 1 µL of RiboLock RNase Inhibitor (20 U/µL), 2 µL of 10 mM dNTP mixture, and 1 µL of RevertAid RT enzyme (200 U/µL). The reagents were mixed by centrifugation and incubated in a PCR thermocycler (Eppendorf, Hamburg, Germany) at 25 ˚C for 5 minutes followed by 42 ˚C

for 60 minutes. The reaction was then terminated by heating at 70 ˚C for 5 minutes. The prepared reverse transcription products (cDNA) were quantified as described in the section 2.2.2. The cDNA samples with the A_{260}/A_{280} ratio of 1.7-1.9 were retained for qRT-PCR analysis.

2.2.4.3 Preparation of dilutions for PCR and qRT-PCR

The volume of cDNA needed for dilutions preparation was calculated using the following formula $N_1 XV_1 = N_2 XV_2$ equaling to final concentration of 25 ng/µL and final volume of 100 µL.

N1 = cDNA Stock concentration

V1 = Volume required from cDNA stock

N2 = Required concentration of dilutions

V2 = Required volume of dilutions

The prepared cDNA dilutions concentration was further quantified as described in section 2.2.2 using NanoDrop™ 2000 Spectrophotometer (Thermos Scientific, MA, USA). The cDNA dilutions with a concentration of 25 ng/µL were retained for PCR and qRT-PCR analysis.

2.2.5 Gene(s) amplification

2.2.5.1 Primer designing

Phenyl-alanine ammonia lyase, and *2'-hydroxyisoflavone reductases* (*PAL* and *IFR*; Chu *et al*., 2014), *chalcone synthase* (*CHS*; Clough *et al*., 2004), *2-hydroxyisoflavanone synthase*/*isoflavone synthase* (*IFS*; Subramanian *et al*., 2004), *Hydroxyisoflavanone*/*isoflavone-4'-O-methyl-transferase* (*HI4'OMT*/*I4'OMT*; Li *et al*., 2016) are multigene families. Primer pairs for *PAL*, *CHS*, *IFS*, *IFR* and *CPY93A* families'

member genes (specific primer pair to the unique gene copy) were used from Cheng et al. (2015); Chu et al. (2014); Gutierrez-Gonzalez et al. (2010) and Chen et al. (2009). Primer pairs for gene copies from *HI4'OMT*/*I4'OMT* gene family was designed based on soybean mRNA sequences published on phytozome12 database (the plant genomics resource; (https://phytozome.jgi.doe.gov/pz/portal.html). The primer pairs for corresponding gene were designed using primer3web version 4.1.0 [(http://primer3.ut.ee) (Rozen and Skaletsky, 2000)], and the properties of all designed primers were checked using the online primer blast tool of national center for biotechnology information (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primer3 software was able to locate primer sets that were specific to a unique gene copy, when taking into account all parameters required for primer design such as primer length, amplicon length, primer guanine–cytosine content, melting temperatures, no hairpin and primer dimer formation. Eukaryotic elongation factor 1-alpha (*ELFα*, GenBank accession, TC203954; Gutierrez-Gonzalez *et al.*, 2010) and soybean ubiquitin-1 [*GmUB1* (Glyma.10g251900) GenBank accession, D28123; Zhang *et al.*, 2017] were used as housekeeping genes for internal normalization.

2.2.5.2 PCR amplification reactions

All designed primers were first tested using regular PCR on prepared cDNAs from S54 and S67 samples to check their specificity and amplification efficiencies. Based on amplification capability, one primer pair was selected for each gene (Table 2.2.5.2.1). Polymerase Chain Reaction (PCR) was carried out for the selected genes from *PAL*, *CHS*, *IFS*, *IFR, CPY93A1* and *HI4'OMT*/*I4'OMT* gene families using the conditions as outlined in Table 2.2.5.2.2. All the reagents were used in the amplification reactions were obtained

from the Promega (Promega, Madison, WI, USA) except the primers. The primers were synthesized commercially by the Eurofins Genomics (Eurofins MWG Operon LLC company, 12701 Plantside Dr, Louisville, KY 40299).

Table 2.2.5.2.1: Selected primers used in regular PCR and real-time quantitative reverse transcription-polymerase chain reaction (qRT-PCR) for selectively amplifying defense-associated gene copies in wild soybeans (S54 and S67 genotypes).

Target gene(s)[Π]	Primer name	Forward and reverse primer sequence (5' to 3')	Primer Tm (°C) [‡]	PCR product size and location of primers in transcript	Gene ID
PAL1.1	PAL-1.1-F	AACAGCATGAACAACGGCAC	62.4	91, position at 295 to 385	Glyma.03g181600
	PAL-1.1-R	GCAGAGCACCACCTTGTTTG	62.4		
PAL1.2	PAL-1.2-F	AGCAACACAACCAGGATGTCAA	60.8	120, position at 1542 to 1561	Glyma.03g181700
	PAL-1.2-R	CAATTGCTTGGCAAAGTGCA	58.4		
PAL1.3	PAL-1.3-F	AGCAGCATAACCAGGATGTGAA	60.8	120, position at 1439 to 1558	Glyma.19g182300
	PAL-1.3-R	CAATCGTTGGCAAAGTGCA	60.4		
PAL2.1	PAL-2.1-F	TCCATGGATAACACCCGTTT	58.4	278, position at 1222 to 1499	Glyma.13g145000
	PAL-2.1-R	TTCCTTGATGAAATCAGCCC	58.4		
CHS7	CHS-7-F	AACCCACCAAA**Y**[#]CGTGTTGAT	59.6	111, position at 76 to 186	Glyma.01g228700
	CHS-7-R	**Y**[#]TTGTCACACATGCGCTG**R**[#]AAT	60.8		
CHS8	CHS-8-F	ATGGAGCTGCTGCTGTCATTG	62.6	132, position 650 to 781	Glyma.11g110500
	CHS-8-R	CCTCACGAAGGTGTCCATCAA	62.6		
IFS1	IFS-1-F	AGAATTCCGTCCCGAGAGGTT	62.6	148, position 1233 to 1380	Glyma07g202300
	IFS-1-R	TGCCATTCCTGAAGTAGCCAA	60.6		
IFS2	IFS-2-F	AATGTGCCCTGGAGTCAATCTG	62.7	121, position 1341 to 1461	Glyma13g173500
	IFS-2-R	GGCGTCACCACCCTTCAATAT	62.6		
GmIFR	IFR-F	AGATGGAAATGTGAAAGGAGCG	60.8	101, position 573 to 673	Glyma.01g172600
	IFR-R	TGTGCACGGCTTTGTTCAAG	60.4		
CYP93A1	CYP93A1-F	ATGTGTTGGAGAAGGCAAGG	60.4		Glyma.03g143700
	CYP93A1-R	CCCTACCAATAGCCCAAACA	60.4	234, position 980 to 1213	
HI4'OMT/I4'OMT	I4'OMT-F	GGCAGGATTCAGCAACTAC	60.2	104, position 1104 to 1207;	Glyma.13g173300
	I4'OMT-R	GTACACAACGATTACAACAGC	58.7		
GmUB1	GmUB1-F	GTGTAATGTTGGATGTGTTCCC	60.8	108, position 1159 to 1266	Glyma.20g141600
	GmUB1-R	ACACAATTGAGTTCAACACAAACCG	61.3		

[Π]**Target genes:** *PAL*, *phenyl-alanine ammonia lyase*; *CHS*, *chalcone synthase*; *IFS*, *2-hydroxyisoflavanone/isoflavone synthase*; *IFR*, *2'-hydroxy isoflavone reductase*; *HI4'OMT/I4'OMT*, *Hydroxyisoflavanone/isoflavone-4'-O-methyl transferase*; *CYP93A1*, *3, 9-dihydroxypterocarpan 6a-monooxygenase*;

Degenerate primers symbol sequences: **Y[#]:** C + T and **R[#]:** A + G; ‡Primer melting temperature (Tm) curve is provided as Tm = 2(A + T) + 4(C + G).

Table 2.2.5.2.2: List of reagents and their concentrations used for PCR amplification

Reagents	Stock Concentration	Volume (µL)	Final Concentration
5X Green Go *Taq*® Flexi Buffer	5X	4.0	1X
Magnesium chloride ($MgCl_2$)	25 mM	4.0	2.0 mM
dNTPs	2.5 mM each; Total 10 mM	1.0	0.1 mM each
Forward Primer (5'-3')	20 µM	1.0	0.16 µM
Reverse Primer (5'-3')	20 µM	1.0	0.16 µM
Go *Taq* DNA *Polymerase* Enzyme	5 U/ µL	.25	0.125 U/ µL
Sterilized Nuclease Free Water	-------	6.75	----------
cDNA	25 ng/ µL	4 µL	100 ng/ µL
Final Volume	**20 µL**		

The PCR amplification reactions were carried out in a Mastercycler gradient PCR system (Eppendorf, Hamburg, Germany). PCR was carried out for 30 cycles, under the following conditions (Table 2.2.5.2.3)

Table 2.2.5.2.3: The sequential steps used for PCR amplification reactions

Steps	Temperature (°C)	Time duration (sec)	Number of cycles
Initial denaturation	95	600	1
Denaturation	95	30	40
Annealing	60	30	40
Extension	72	30	40
Final extension	72	300	1
Final hold	4	∞	-

2.2.5.3 Agarose gel electrophoresis

The amplified PCR products were separated on 1.0% (w/v) agarose (GeneMate™ Low Melt Agarose, Kaysville, UT) gel using 1X TAE buffer (Annexure I; Table 2.2.3.1) stained with SYBR® safe DNA gel stain (1X, 40 mM Tris-acetate, 1 mM EDTA, pH ~8.3, Thermo Scientific, MA, USA) as described earlier in the section 2.2.3. After 3/4th running of the gel it was photographed with Biospectrum AC® imaging system (UVP, Upland,

CA, USA) and the band sizes were compared with 100 bp DNA molecular weight marker (MO BIO Laboratories, INC., CA, USA).

2.2.6 Quantitative real-time (qRT-PCR) polymerase chain reaction

Quantitative real-time PCR (qRT-PCR) was carried out through Applied Biosystems 7500 Real-Time PCR system (Applied Biosystems Foster City, CA, USA). cDNA template was prepared for quantification, from the extracted RNA from the same biological replicates. Quantitative Real-Time (qRT-PCR) Polymerase Chain Reaction was carried out for the gene from *PAL*, *CHS*, *IFS*, *IFR*, *CYP93A1* and *HI4'OMT*/*I4'OMT* gene families of the phenyl-propanoid or iso-flavonoid pathway using the selected primers (Table 2.2.5.2.1) as per the conditions outlined in Table 2.2.6.1.

Table 2.2.6.1: List of reagents and their concentrations used for PCR amplification

Reagents	**Stock Concentration**	**Volume (µL)**	**Final Concentration**
2X Fast SYBR Master Mix	2X	10	1X
Forward Primer (5'-3')	10 µM	2.5	1.2 µM
Reverse Primer (5'-3')	10 µM	2.5	1.2 µM
Sterilized Nuclease Free Water	-------	---------	----------
cDNA	25 ng/ µL	5 µL	100 ng/ µL
Final Volume	**20 µL**	–	

All the reagents used in the amplification reactions were obtained from the New England Bio Labs Inc. (Ipswich, MA, USA) except the primers. The primers were synthesized commercially by the Eurofins Genomics (Eurofins Genomics Oligo LLC, Huntsville, AL, USA). The reaction was conducted in triplicates for the biological replicates and triplicates for each biological replicate for getting the technical replicates. Eukaryotic elongation factor 1-alpha (*ELFα*; Gutierrez-Gonzalez *et al.*, 2010) and soybean ubiquitin-1 (*GmUB1*; Zhang *et al.*, 2017) were used as housekeeping genes (Table 2.2.5.2.1) for comparison

between ΔCt values. DNA accumulation was measured using SYBR Green as the reference dye. Only one product was present in each reaction as indicated by the reference dye's dissociation curve of amplified products (Annexure II; Figure 2.2.6.1). qRT-PCR was programmed for 40 cycles under the following conditions (Table 2.2.6.2):

Table 2.2.6.2: The sequential steps used for qRT-PCR expression reactions

Steps	Temperature (°C)	Time duration (min)	Number of cycles
Initial denaturation	95	0.5	1
Denaturation	95	0.07	40
Annealing	60	0.25	40
Extension	72	0.17	40

PCR efficiencies for target and reference genes were equal between the target and control samples.

2.2.7 Statistical analysis of gene expression data

Three biological replicates and three technical replicates were used in qRT-PCR. The Ct values that were generated from qRT-PCR were used in the calculation of mean fold change, which is the average of the replicates. The ΔΔCT method was used for relative quantification of gene expression (Livak *et al*., 2001). The standard deviation was calculated from the same technical replicates. Susceptible control was used as the endogenous control within the calculations. Once the mean fold change and standard deviation were calculated, then a student's t-test was calculated to determine statistical significance between the samples. A $P \leq 0.05$ was considered to indicate statistical significance. Comparisons included susceptible control vs. susceptible treated, resistant control vs. resistant treated, susceptible treated vs. resistant treated, susceptible control vs.

resistant treated, and resistant control vs susceptible treated. There was special focus on susceptible treated vs resistant treated.

2.3 cDNA and Genomic DNA Cloning of *HI4'OMT*/*I4'OMT* Candidate Gene

2.3.1 Tissue collection

Four wild soybean genotypes (S54, S67, S100, and PW5) and two soybean cultivars (SC1, and Williams 82) were used in this study. Wild soybean S54 (Zhang *et al*., 2016) and S100 were found resistant for both SCN (SCN5 and SCN2) races, however, soybean cultivar (SC1) was found resistant to SCN5 only [Zhang et al. (2016) lab unpublished data; (Table 2.3.1.1)]. The seeds of each genotype were obtained from USDA Soybean Germplasm Collection- Grin Global (www.ars-grin.gov/). Ten seeds per genotype were surface sterilized and processed for germination as described in the section 2.1.2. After 2-3 days, each healthy seedling was transplanted into the pots having miracle grow soil for more growth. Approximately, after two-weeks of plant growth the leaf tissues were collected for DNA extraction and stored at -80 °C until use. Root tissues that have been collected for gene expression analysis (section 2.1.2 and 2.2.1) for genotypes S54, and S67 were used for RNA extraction and cDNA preparation. Except S54, S67 and Williams 82 (used for both genomic DNA and cDNA cloning) all other genotypes were used for genomic DNA cloning only. Available coding and genomic sequences (https://phytozome.jgi.doe.gov/pz/portal.html) of *HI4'OMT*/*I4'OMT* gene from 'Williams 82' cultivar was used as template for sequence comparisons.

Table 2.3.1.1: Details of genotypes used for cDNA and genomic DNA cloning

Genotype(s)	Wild Soybean/Soybean	SCN Resistance	Country of Origin	*Female Index (%)	
				SCN5	SCN2
S54	Wild Soybean	Resistant (SCN5 and SCN2)	Chungchong Puk, South Korea	5.2	8.0
S100	Wild Soybean	Resistant (SCN5 and SCN2)	Khabarousk, Russian Federation	3.3	3.3
PW5	Wild Soybean	--------	Liaoning, China	---	---
S67	Wild Soybean	Susceptible (SCN5 and SCN2)	Shanxi, China	149	80
SC1	Soybean Cultivar	Resistant (SCN5)	China	2.4	3.0
Williams-82	Soybean Cultivar	Susceptible (SCN5 and SCN2)	Illinois, USA	116.8	120
Total	**06**				

*Data source for female index: SCN 5 female index data source Zhang et al. (2016); SCN 2 female index data source Zhang et al. (2019; under publication) and Dr. Song lab's unpublished data;

2.3.2 Isolation of full-length cDNAs and genomic DNAs of *HI4'OMT*/*I4'OMT*

2.3.2.1 RNA and DNA extraction

Total RNA was isolated from the SCN infected root tissues of wild soybeans (S54, S67) and soybean (Williams 82) using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The DNase treatment and first strand cDNA synthesis has been carried out as described in the section 2.2.4.1 and 2.2.4.2 respectively. Total genomic DNA extraction of all six genotypes was carried out using CTAB protocol as described by Mittal et al. [(2010) (Annexure I; Table 2.3.2.1.1)] with some modifications. The purity ratio and concentration of extracted RNA, DNA and cDNA was measured by NanoDrop™ 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA) as described in the section 2.2.2.

2.3.2.2 Gene amplification

Full length cDNAs and genomic DNAs of *HI4'OMT*/*I4'OMT* gene were amplified via PCR using the following two gene specific primers: the forward primer 5'-ATATCCTTCGAAAACCACTCACTC-3' and the reverse primer 5'-CCGTAGTACACAACGATTACAACA-3' and *Q5 high-fidelity DNA* polymerase enzyme (Table 2.3.2.2.1). All the reagents except the primers were used in the

amplification reactions were obtained from the New England Bio Labs Inc. (Ipswich, MA, USA). The primers were synthesized commercially by the Eurofins Genomics (Eurofins Genomics Oligo LLC, Huntsville, AL, USA).

Table 2.3.2.2.1: List of reagents and their concentrations used for PCR amplification

Reagents	**Stock Concentration**	**Volume (µL)**	**Final Concentration**
5X Q5 Reaction Buffer	5X	5.0	1X
dNTPs	10 mM	0.5	200 µM
Forward Primer (5'-3')	10 µM	1.25	0.5 µM
Q5 High-Fidelity DNA Polymerase	5 U/ µL	0.25	0.02 U/ µL
Sterilized Nuclease Free Water	-------	6.75	----------
cDNA or genomic DNA	25 ng/ µL	4 µL	100 ng/ µL
Final Volume	**20 µL**		

The PCR amplification reactions were carried out in a Mastercycler gradient PCR system (Eppendorf, Hamburg, Germany). PCR was carried out for 35 cycles, under the following conditions (Table 2.3.2.2.2):

Table 2.3.2.2.2: The sequential steps used for PCR amplification reactions

Steps	**Temperature (°C)**	**Time duration (sec)**	**Number of cycles**
Initial denaturation	98	30	1
Denaturation	98	10	35
Annealing	65	30	35
Extension	72	50	35
Final extension	72	120	1
Final hold	4	∞	-

2.3.2.3 Agarose gel electrophoresis

The amplified PCR products were separated on 1.0% (w/v) agarose (GeneMate™ Low Melt Agarose, Kaysville, UT) gel using 1X TAE buffer (Annexure I; Table 2.2.3.1) stained with SYBR® safe DNA gel stain (1X, 40 mM Tris-acetate, 1 mM EDTA, pH ~8.3, Thermo Scientific, MA, USA) as described earlier in the section 2.2.2. After 3/4th running of the

gel, it was photographed with Biospectrum AC® imaging system (UVP, Upland, CA, USA) and the band sizes were compared with 1 Kbp DNA molecular weight marker (Thermo Scientific Waltham, MA, USA). After, visualization the accurate gene product was cut and excised from the agarose gel. Excised gene products were purified using gel extraction purification kit (QIAquick gel extraction kit, Qiagen, Hilden, Germany) following manufacturer's instruction. The quantity of the purified gene product was measured using NanoDrop™ 2000 spectrophotometer (Thermo Scientific, MA, USA). The purity ratio of the refined gene product (purified gene product) was measured at A_{260}/A_{280} nm wavelength and the concentration was measured in ng/µL using spectrophotometer built in program. Samples having the purity ratio (A_{260}/A_{280} ratio) above 1.7 and the concentration above 20 ng/µL were retained for the further steps.

2.3.3 Polyadenylation of purified cDNAs

The amplified and purified cDNAs and genomic DNAs of *HI4'OMT*/*I4'OMT* gene from various genotypes was further used for polyadenylation. The poly-A tail was added at 3'-end of cDNA or genomic DNA using below mentioned protocol (Table 2.3.3.1) to make the 3'-ends stickies:

Table 2.3.3.1: List of reagents and their concentrations used for polyadenylation

Reagents	**Stock Concentration**	**Volume (µL)**	**Final Concentration**
5X *Go Taq* Colorless Buffer	5X	5.0	1X
Magnesium Chloride ($MgCl_2$)	25 mM	3.0	1.5 mM
dATP	10 mM	1.0	0.2 mM
Go Taq-Flexi DNA Polymerase	5 U/ µL	4.0	0.4 U/ µL
Sterilized Nuclease Free Water	-------	9.0	----------
Purified PCR product	20-60 ng/ µL	28 µL	~500-1200 ng/ µL
Final Volume	**50 µL**		

All the reactions were mixed properly by pipetting and incubated in the Mastercycler gradient PCR system (Eppendorf, Hamburg, Germany) at 70 °C for 30 minutes. After incubation each polyadenylated reaction was purified again using gel extraction purification kit (QIAquick gel extraction kit, Qiagen, Hilden, Germany). The concentration of the purified reactions was quantified in ng/µL using NanoDrop™ 2000 spectrophotometer (Thermo Scientific, MA, USA) built in program. The reactions having concentrations above 20 ng/µL were only retained for the ligation.

2.3.4 Vector ligation

Purified and polyadenylated gene products were ligated into the pGEM-*T4* Easy Vector (Promega Corporation, Madison, WI, USA) following manufacturer's instructions (Table 2.3.4.1).

Table 2.3.4.1: List of reagents and their concentrations used for vector ligation

Reagents	**Stock Concentration**	**Volume (µL)**	**Final Concentration**
2X Rapid Ligation Buffer	2X	5.0	1X
pGEM®-T Easy Vector	50 ng	1.0	5.0 ng
T4-DNA Ligase	3 Weiss U/ µL	1.0	0.3 Weiss U/ µL
Purified PCR product	20-30 ng/ µL	3.0	~60-100 ng/ µL
Final Volume	**10 µL**		

The reactions were mixed properly by pipetting and incubated overnight (~18-20 hours) at 4 °C in Mastercycler gradient PCR system (Eppendorf, Hamburg, Germany) to get the maximum number of transformants.

2.3.5 Transformation into the competent cells

After overnight incubation, the reactions were further incubated on ice for 10 minutes and carefully 20-30 µL of stellar competent cells [(*E. coli* cells) (Takara Bio USA Inc.,

Mountain View, CA, USA)] were transferred into each reaction. Tubes were flicked gently and again incubated at ice for 30 minutes. Then reactions were heat shock in water bath (Benchmark Scientific Inc., Sayreville, NJ, USA) at 42 °C for exactly 60 seconds and immediately return on ice for 2 minutes. Further ~1000 µL liquid LB medium (Annexure I; Table 2.3.5.1a) were added into each tube and shake for 1.5 hours at 37 °C onto the benchtop shaking incubator (Labnet International Inc., Edison, NJ, USA) and then centrifuged (Eppendorf, Hamburg, Germany) for 8 minutes at 10, 000 rpm. Transformants pellets were mixed with 100 µL liquid LB media and the culture was spread onto the LB/ampicillin/IPTG/X-Gal (Annexure I; Table 2.3.5.1b-e) plates. The plates were incubated at 37 °C in incubator (Biorad, Hercules, CA, USA) for 24 hours.

2.3.6 Selection of positive transformants and plasmid extraction

After overnight incubation, single bacterial colonies (single clones) from each genotype had been selected and shake overnight into the 1000 µL LB liquid media with 50 mg/mL ampicillin [(1 µL) (Annexure I; Table 2.3.5.1a and Table 2.3.5.1c)] at 37 °C onto the benchtop shaking incubator (Labnet International Inc., Edison, NJ, USA). Further, after overnight shake PCR was carried out using transformant culture as template via protocol described into the section 2.3.2.2.1-2.3.2.2.2. The positive clones (amplified targeted gene only) were then selected and sub cloned into the 4 mL LB liquid media with 50 mg/mL ampicillin [(4 µL) (Annexure I; Table 2.3.5.1a and Table 2.3.5.1c)] and shake overnight again at 37 °C onto the benchtop shaking incubator (Labnet International Inc., Edison, NJ, USA). The shake cultures were further used for plasmid extraction using the Qiagen MiniSpin Prep Kit (Qiagen, Hilden, Germany). The quantity and quality of the extracted

plasmid was measured using NanoDrop™ 2000 spectrophotometer (Thermo Scientific, MA, USA). The plasmids having purity ratio (A_{260}/A_{280} ratio) above 1.6 and the concentrations ≥ 100 ng/µL were only used for the sequencing. Sequencing of the selected plasmids has been carried out by Eurofins Genomics (Eurofins Genomics, Huntsville, AL, USA). Approximately 10-12 clones' plasmid DNA have been sequenced for each genotype. T7 (forward primer: 5′-TAATACGACTCACTATAGGG-3′) and SP6 (reverse primer: 5'-ATTTAGGTGACACTATAGAAT-3') universal primers were used for the sequencing.

2.3.7 Multiple sequence(s) alignment and comparison

The quality of sequences was checked by SnapGene Viewer (SnapGene software, GSL Biotech; snapgene.com) and FinchTV version 1.4 (http://www.geospiza.com/Products/finchtv.shtml). Sequences with peak quality of more than 60 (Annexure II; Figure 2.3.7.1a) with no noise were used for the alignment and comparison analysis. A total number of 8-10 clone sequences from each genotype were used for the alignment and the analysis of nucleotide variations for both cDNA and genomic DNA comparisons. Sequence alignment was carried out using Bioedit software and the aligned sequence(s) were compared to identify the nucleotide variations and to identify the different haplotypes among various genotypes (Annexure II; Figure 2.3.7.1b). Sequences of coding region were protein translated and the amino acid comparisons were carried out.

2.3.8 Mutation screening of *HI4'OMT*/*I4'OMT*

Protein variation effect analyzer (PROVEAN) score predictions were performed on identified amino acid mutations. PROVEAN predicts whether an amino acid

substitution affects protein function based on sequence homology and the physical properties of amino acids. PROVEAN score with prediction cutoff of ≤ -2.5 are considered confident. Changes with a PROVEAN score, ≤ -2.5 are predicted to be damaging to the protein (Choi *et al.*, 2012).

2.4 Functional Validation of *HI4'OMT/I4'OMT* Candidate Gene or *H1-H4* alleles

2.4.1 Plant, nematode, and bacterium sources

Soybean genotype: 'Williams 82' was used in this study. Williams 82 is the susceptible soybean cultivar for SCN races 2 and 5. SCN race 2 and 5 both were used for SCN bioassays and the maintenance of SCN was carried as previously described in the section 2.1.1. *Agrobacterium rhizogenes* strain K599 (obtained from North Carolina State University, Raleigh, NC, USA) was utilized for the generation of the soybean hairy roots.

2.4.2 Amplification of *HI4'OMT/I4'OMT-H1-H4* (haplotype I to haplotype IV) candidate alleles

The *HI4'OMT/I4'OMT-H1-H4* haplotype's plasmid DNA (from the cDNA cloning, section 2.3.2 to 2.3.6) was amplified using *Q5 high-fidelity DNA polymerase* enzyme and the following two gene and vector specific primers: the forward primer (T-P-1-F) 5'-GGACTCTAGAGGATCCATGGTTTTCTGTGGCAAC-3' and the reverse primer (T-P-1-R) 5'- GATCGGGGAAATTCGAGCTCTTAAGGATAAACTTCAATGAGAG-3' (Table 2.4.2.1). All the reagents except the primers were used in the amplification reactions were obtained from the New England Bio Labs Inc. (Ipswich, MA, USA). The primers were designed by In-Fusion Cloningtools(http://www.takarabio.com/US/Products/Cloning_andCompetent_Cells/Sele

ction_Guides/Online_In-Fusiob_Tools) and synthesized commercially by the Eurofins Genomics (Eurofins Genomics Oligo LLC, Huntsville, AL, USA).

Table 2.4.2.1: List of reagents and their concentrations used for PCR amplification

Reagents	**Stock Concentration**	**Volume (µL)**	**Final Concentration**
5X *Q5* Reaction Buffer	5X	5.0	1X
dNTPs	2.5 mM each; 10 mM	0.5	200 µM each
Forward Primer (5'-3')	10 µM	1.25	0.5 µM
Reverse Primer (5'-3')	10 µM	1.25	0.5 µM
Q5 DNA Polymerase	5 U/ µL	0.25	0.02 U/ µL
Sterilized Nuclease Free Water	-------	15.75	----------
Plasmid DNA	50-60 ng/ µL	1 µL	50-60 ng/ µL
Final Volume	**20 µL**		

The PCR amplification reactions were carried out in a Mastercycler gradient PCR system (Eppendorf Hamburg, Germany). PCR was carried out for 35 cycles, under the following conditions (Table 2.4.2.2):

Table 2.4.2.2: The sequential steps used for PCR amplification reactions

Steps	**Temperature (°C)**	**Time duration (sec)**	**Number of cycles**
Initial denaturation	98	30	1
Denaturation	98	10	35
Annealing	58	30	35
Extension	72	40	35
Final extension	72	120	1
Final hold	4	∞	-

2.4.3 Agarose gel electrophoresis

The amplified PCR products were separated on 1.0% (w/v) agarose (GeneMate™ Low Melt Agarose, Kaysville, UT) gel using 1X TAE buffer (Annexure I; Table 2.2.3.1) stained with SYBR® safe DNA gel stain (1X, 40 mM Tris-acetate, 1 mM EDTA, pH ~8.3, Thermo Scientific, MA, USA) as described earlier in the section 2.2.2. After 3/4th running of the gel it was photographed with Biospectrum AC® imaging system (UVP, Upland, CA, USA) and the band sizes were compared with 1 Kbp DNA molecular weight marker (Thermo Scientific Waltham, MA, USA). After, visualization the accurate gene product was cut and excised from the agarose gel. Excised gene products were purified using gel extraction purification kit (QIAquick gel extraction kit, Qiagen, Hilden, Germany) following manufacturer's instruction. The quantity of the purified gene product was measured using NanoDrop™ 2000 spectrophotometer (Thermo Scientific, MA, USA). The purity ratio of the refined gene product (purified gene product) was measured at A_{260}/A_{280} nm wavelength and the concentration was measured in ng/µL using spectrophotometer built in program. Samples having the purity ratio (A_{260}/A_{280} ratio) above 1.8 and the concentration above 100 ng/µL were retained for further steps.

2.4.4 Binary vectors construction

The pCAMBIA-1302 plant expression vector with green fluorescent protein (*GFP*): 35S-*GFP*-NOS terminator (Marker Gene Technologies, Inc, Eugene, OR, USA) was used as a negative control vector. Next, the cassette of 35S::*GUS*/NOS terminator was excised from a pBI121 vector by digestion with *EcoR*I and *Hind*III [Table 2.4.4.1; (New England Bio Labs Inc., Ipswich, MA, USA)] and inserted into the *EcoR*I and *Hind*III sites of pCAMBIA 1302-*GFP* through *T4* DNA ligation as described in the section 2.3.4, which led to a new

vector named pCAMBIA 1302-*GFP*-35S::*GUS*/NOS terminator, which was used as a positive control.

Table 2.4.4.1: The sequential steps used for double restriction digestions

Reagents	Stock Concentration	Volume (µL)	Final Concentration
10X Cut Smart Reaction Buffer	10X	5.0	1X
*Eco*RI (Restriction enzyme#1)	20,000 U/ mL	0.25	5.0 units
*Hind*III (Restriction enzyme#2)	20,000 U/ mL	0.25	5.0 units
pCAMBIA 1302 plasmid DNA	100 ng/ µL	50	5.0 µg (5,000 ng)
pBI121 plasmid DNA	100 ng/ µL	50	5.0 µg (5,000 ng)
Final volume	110 µL		
The reactions were mixed properly by pipetting and incubated overnight at 37 °C in Master cycler gradient PCR system (Eppendorf, Hamburg, Germany)			

Further the *GUS* part of binary vector (pCAMBIA 1302-*GFP*-35S::***GUS***/NOS) was removed through double digestion with *BamH*I and *Sac*I restriction enzymes (New England Bio Labs Inc., Ipswich, MA, USA) as per Table 2.4.4.1 and the linearized vector without *GUS* (Annexure II; Figure 2.4.4.1a) was further purified using gel extraction purification kit (QIAquick gel extraction kit, Qiagen, Hilden, Germany) following manufacturer's instruction. Then, purified PCR products of *H1-H4* candidate genes (from section 2.4.2 to 2.4.3) were further inserted into the *BamH*I and *Sac*I sites of linearized and purified pCAMBIA-1302-*GFP*-35S::---NOS vector via In-Fusion HD Cloning Kit (Takara Bio USA, Inc., Mountain View, CA, USA) to replace the *GUS* gene, which resulted in the pCAMBIA-1302-*GFP*-35S::*I4'OMT-H1-H4*/NOS terminator construct individually. The ligated construct was further transferred into the *E. coli* competent cells and grown overnight at 37 ˚C on LB medium plate containing 50 mg/mL kanamycin (Annexure I; Table 2.3.5.1b-c). After ~24 hours of growth, single bacterial colonies (single clones) were

further sub-cultured and shaked overnight at 37 ˚C onto the benchtop shaking incubator (Labnet International Inc., Edison, NJ, USA) into the LB liquid medium with 50 mg/mL kanamycin [1 µL; (Annexure I; Table 2.3.5.1a and Table 2.3.5.1c)]. Shaked broth was used for PCR to identify the positive clone(s). The PCR has been carried out using Go *Taq* Flexi buffer and Go *Taq* DNA *polymerase* as described in the section 2.4.2. Four positive clones were selected (Annexure II; Figure 2.4.4.1b), and plasmid extraction was carried out as described in the section 2.3.6. Plasmid DNA concentration and purity was checked using NanoDrop™ 2000 spectrophotometer (Thermo Scientific, MA, USA) built in program. The reactions having concentrations above 150 ng/µL were only used for sequencing. Validation of pCAMBIA1302-*GFP*-35S::*I4'OMT-H1-H4*/NOS terminator construct was further carried out via sequencing by Eurofins Genomics (Eurofins Genomics, Huntsville, AL, USA) using two gene specific primers: the forward primer 5'-GCAAGCATGTGTTCGAGGGA-3' and the reverse primer 5'-CAGTGCATTTCAAGTGAGGGA -3' (Annexure II; Figure 2.4.4.1c-d). The primers were designed using program primer3web version 4.1.0 (http://primer3.ut.ee) as per the section 2.2.5.1.

2.4.5 Transformation of binary vectors into the *Agrobacterium rhizogenes* competent cells

Identified positive clones having constructs including pCAMBIA1302-*GFP* and *GUS* (control vectors) and pCAMBIA 1302-*GFP*-35S::*I4'OMT-H1-H4*/NOS terminator [(expression vectors) (from section 2.4.4)] purified plasmid DNA were used separately and introduced into the *A. rhizogenes* strain K599 competent cells using the freeze thaw method (Höfgen and Willmitzer, 1988 modified protocol; modified by David Neece, Nov 2013) and the transformed cells culture were streak on to the LB medium plate containing

50 mg/mL kanamycin (Annexure I; Table 2.3.5.1b-c). The plates were kept in incubator (Biorad, Hercules, CA, USA) at 28 ˚C for approximately 2-3 days. After, ~48 hours of incubation the single bacterial colonies were generated and 10 single clones from both the controls and *H1-H4* genes were chosen and shake overnight into the 1000 µL LB liquid media with 50 mg/mL kanamycin [(1 µL) (Annexure I; Table 2.3.5.1a and Table 2.3.5.1c)] at 28 °C onto the benchtop shaking incubator (Labnet International Inc., Edison, NJ, USA). The shake broth (200 µL) was further sub-cultured in liquid media and used for plasmid extraction as described in the section 2.3.6, and the concentration of extracted plasmid was quantified using NanoDrop™ 2000 spectrophotometer (Thermo Scientific, MA, USA) built in program. The extracted plasmid DNA was further used for positive construct validation using Go *Taq* Flexi buffer and Go *Taq* DNA *polymerase* as described in the section 2.4.2. One positive clone from *H1-H4* genes was further selected (Annexure II; Figure 2.4.5.1). Broth of positive clones (*H1-P27.3*, *H2-2*, *H3-4*, and *H4-6.3*) and prepared pCAMBIA-1302 control vectors (*GFP* and *GUS*) was streak all over the LB medium plate containing 50 mg/mL kanamycin (Annexure I; Table 2.3.5.1b-c) and incubated for 24-36 hours at 28 °C. The generated single clones were sub-cultured into LB liquid media and shake for ~16 h as described above. The shake broth was centrifuged for 5 minutes at 5,000 rpm and the pellets were suspended into the supernatant. The suspended mixture (200 µL on one plate) was spread on to the LB medium plate with 50 mg/ mL kanamycin and kept for incubation at 28 °C for 24-36 hours.

2.4.6 Generation of transgenic soybean hairy roots

Soybean cultivar (Williams 82) seeds were first transferred to autoclaved filter paper moistened with sterile distilled water for germination. The germinated soybean seeds were

sowed in sterile vermiculite, and five seeds were grown in each cell of an 18-cell germination tray [(~Total 90 seeds sowed per tray) (Burpee seed trays)]. The seedlings were kept in the growth chamber (Intellus Control System, Percival Scientific Inc., Perry, Iowa) under controlled conditions (27 °C, 16 h light/ 8 h dark and 75% humidity). Seedlings were constantly monitored and regularly sprayed with water for the maintenance of humidity. The incubated bacteria lawn (from section 2.4.5) of *A. rhizogenes* K599 with binary vector pCAMBIA1302 (*GFP* and *GUS*) and pCAMBIA-1302-*GFP*-35S::*I4'OMT-H1-H4*/NOS terminator from an LB medium plate containing 50mg/ mL kanamycin (Annexure I; Table 2.3.5.1b-c) was collected for soybean transformation. Soybean hairy roots were generated as previously reported (Kereszt *et al*., 2007). The soybean plants were kept in the growth chamber (Intellus Control System, Percival Scientific Inc., Perry, Iowa) on to the following growth conditions: 75% humidity, 16 h light/8 h dark at 27 °C with irradiance from cool white fluorescent bulbs at 150-200 μmol m^{-2}/s. After about 3-4 weeks, the hairy roots grew approximately 10 cm in length, the soybean transgenic roots were detected based on *GFP* expression, using an epifluorescent microscope (Olympus stereo microscope model SZX7, Olympus America, Center Valley, PA) with a GFP filter set and Olympus U-TV1X-2 imaging software (Annexure II; Figure 2.4.6.1a). The *GFP*-positive transgenic roots harboring pCAMBIA1302 (*GFP* and *GUS*) and pCAMBIA-1302-*GFP*-35S::*I4'OMT-H1-H4*/NOS terminator were further analyzed by PCR as per below protocol using two gene and vector specific primers: the forward primer (35S-promoter from vector) 5'-GATTCCATTGCCCAGCTATC -3' and the reverse primer (T-P-1-R from gene) 5'-GATCGGGGAAATTCGAGCTCTTAAGGATAAACTTCAATGAGAG-3' for pCAMBIA-1302-*GFP*-35S::*I4'OMT-H1-H4*/NOS transgenic roots and the forward primer

(35S-promoter from vector)) 5'-GATTCCATTGCCCAGCTATC-3' and the reverse primers (*GFP*-R) 5'-GTCGTCCTTGAAGAAGATGGTC-3'for pCAMBIA1302-*GFP* negative and the GUS-R 5'-CAATAACATACGGCGTGACATC-3'for pCAMBIA-1302-*GFP*-35S::*GUS*/NOS for positive control vector transgenic roots:

Table 2.4.6.1: List of reagents and their concentrations used for PCR amplification

Reagents	**Stock Concentration**	**Volume (µL)**	**Final Concentration**
5X Green Go *Taq* Flexi Buffer	5X	4.0	1X
$MgCl_2$	25 mM	4.0	2 mM
dNTPs	2.5 mM each; 10 mM	1.0	0.1 mM each
Forward Primer (5'-3')	10 µM	1.0	0.5 µM
Reverse Primer (5'-3')	10 µM	1.0	0.5 µM
Go *Taq DNA Polymerase*	5 U/ µL	0.25	0.0625 U/ µL
Sterilized Nuclease Free Water	-------	0.75	----------
Genomic DNA	20-30 ng/ µL	8 µL	150-250 ng/ µL
Final Volume	**20 µL**		

The PCR amplification reactions were carried out in a Mastercycler gradient PCR system (Eppendorf Hamburg, Germany). PCR was carried out for 35 cycles, under the following conditions (Table 2.4.6.2):

Table 2.4.6.2: The sequential steps used for PCR amplification reactions

Steps	Temperature (°C)	Time duration (sec)	Number of cycles
Initial denaturation	98	300	1
Denaturation	98	10	35
Annealing	58	30	35
Extension	72	120	35
Final extension	72	120	1
Final hold	4	∞	-

The amplified PCR products were separated on 1% (w/v) agarose (GeneMate^{TM} Low Melt Agarose, Kaysville, UT) gel using 1X TAE buffer (Annexure I; Table 2.2.3.1) stained with SYBR®safe DNA gel stain (1X, 40 mM Tris-acetate, 1 mM EDTA, pH ~8.3, Thermo Scientific, MA, USA) as described earlier in the section 2.2.2 (Annexure II; Figure 2.4.6.1b-d).The soybean plants containing *GFP*-positive transgenic roots harboring pCAMBIA1302 (*GFP* and *GUS*), and pCAMBIA-1302-*GFP*-35S::*I4'OMT-H1-H4*/NOS terminator, the tap root, *GFP*-negative hairy roots and all were excised from the wounding site and the soybean plants with a single positive transgenic hairy root were further subjected to SCN bioassays. From the same positive transgenic root, two pieces (for PCR validation, and gene expression analysis) were collected and stored at -80 °C until use.

2.4.7 SCN bioassays

2.4.7.1 Root inoculation with SCN

SCN bioassays using positive transgenic hairy roots were conducted following the protocol described by Melito et al. (2010). Two races of SCN (SCN5; HG type 2.5.7 and SCN2; HG type 1.2.5.7) were used for the SCN bioassays. Positive transgenic hairy roots of forty to forty-five soybean plants harboring identical constructs were loaded horizontally in an

8.5 X 8.5 cm cell filled with sterilized sand, and the shoots were left out from the inoculating cell (one plant/per cell). After approximately 1 week of plants growth, the nematodes were prepared as described in the section 2.1.2. About 1500 µL (1.5 mL) of inoculum, which contained about 700-800 second-stage juveniles (J2's), was added to each root system.

2.4.7.2 Analysis of SCN development in transgenic hairy roots

Four to five roots were allowed to be inoculated with J2 nematodes for infection for ~24 hours. Roots were then washed to remove extra juvenile nematodes that had not penetrated the root tissue and transferred into the new cell having fresh sterile sand. The plants were kept in the growth chamber for further growth on the similar conditions as described in the section 2.4.6. After 5 days of inoculation, infected root samples were washed to remove sand using tap water and cleared by 20% (v/v) bleach for no more than 4 min, and then stained by acid fuchsin (Annexure I; Table 2.4.7.2.1) according to the procedure of Byrd et al. (1983). Identified nematodes at developmental stages, J2 (thin), J3 (sausage shaped), elongated male and J4 or adult female nematodes (fat lemon shaped) were counted on each infected root sample. The number of SCN was quite variable among individual roots for all the transgenic and control lines among both races. The ratio of the number of SCNs that developed beyond J2 stage (J3+J4+adult) to the total number of nematodes per root was used as an index to determine the resistance difference among the hairy roots harboring their respective constructs. Then the ratios were normalized to the mean for control transgenic lines. Data were based on three independent biological replicate experiments when used for assays. The photographs of SCN were taken to show the representative stage using a digital microscope equipped with a DFC295 digital color camera (Leica

Microsystems Inc., Illinois, USA). Resistance index (normalized proportion of nematodes developed beyond J2 stage over total nematodes) data among different types of transgenic soybean hairy roots were used to plot the graphs. Unpaired student's *t*-test ($P \leq 0.05$) was carried out to indicate the statistical significance among groups. Comparison was carried out between transgenic roots with control vectors and with *H1-H4* candidate genes individually.

2.4.7.3 Female cyst extraction and counting

To capture female cysts variation responses to HG type 2.5.7 (SCN5) and HG type 1.2.5.7 (SCN2) among positive transgenic roots (controls and *H1-H4* alleles containing positive transgenic roots), the ~15 infected roots (per construct and per race) were washed after approximately 25 days of J2 inoculation. Female cysts were collected from the infected roots by washing and massaging in water and sieving the solution through nested 850- and 250- μm sieves (Klink *et al*., 2007). The extracted female cysts were collected in 5 mL blue screw cap falcon tube and counted under a stereo master microscope. The average number of counted female cysts between 10-12 individual biological replicates among different types of transgenic soybean hairy roots was used to plot the graphs (Narayanan *et al*., 1999). Unpaired student's *t*-test ($P \leq 0.05$) was carried out to indicate the statistical significance among groups (Liu *et al.*, 2017). Comparison was carried out between transgenic roots with control vectors and with each candidate genes. Female index was also calculated using the mean of female cysts counted on *HI4'OMT*/*I4'OMT-H1-H4* candidate genes and control vectors containing transgenic roots and normalized to percent as per Zhang et al. (2016).

2.4.8 Quantitative reverse transcription PCR

Quantitative reverse transcription PCR (qRT-PCR) was performed as previously reported by (Lin *et al*., 2013) to determine gene expression of *HI4'OMT*/*I4'OMT-H1-H4* (haplotype I to haplotype IV) and five genes from iso-flavonoid pathway (*CHS7*, *CHS8*, *GmIFS1*, *GmIFS2* and *GmIFR*) in transgenic hairy roots. The soybean ubiquitin-1 gene [*GmUB1* (Glyma.10g251900) GenBank accession, D28123; forward primer sequence: 5'-GTGTAATGTTGGATGTGTTCCC-3' and reverse primer sequence: 5'-ACACAATTGAGTTCAACACAAACCG-3'; Zhang *et al*., 2017] was used as a reference gene. The sequences of gene specific primers were used (Table 2.2.5.2.1). Two types of transgenic hairy roots in the 'Williams 82' background was used to check the over expression of the *HI4'OMT*/*I4'OMT-H1-H4* genes and to compare the expression of other genes. Transgenic hairy roots with control vectors and with *HI4'OMT*/*I4'OMT-H1-H4* gene construct were used for comparisons. Total RNA was isolated from five independent biological replicates of the respective hairy root tissues using RNeasy Plant Mini Kit® (QIAGEN, CA, USA) as described in the section 2.2.1. DNase treatment to remove genomic DNA, and reverse transcription to synthesize cDNA was further carried out as per the protocol discussed in the section 2.2.4. First strand cDNA was diluted and used for qRT-PCR analysis as per section 2.2.4.3 and DNA accumulation was measured using SYBR green as the reference dye. Only one product was present in each reaction as indicated by the reference dye's dissociation curve of amplified product. PCR efficiencies for target and reference genes were equal between the target and control samples. Ct values were calculated using software (7500 HT Fast Real-Time PCR systems v. 2.3) supplied with the Applied Biosystems 7500 HT Fast Real-Time PCR systems. Each biological

replicate was used thrice (to generate three technical replicates) for qRT-PCR assay and relative abundance analysis was further conducted as per the section 2.2.7.

CHAPTER 3: RESULTS

3.1 LC-MS Based Untargeted Metabolomics

There are different ways by which plants can defend themselves from pathogens or pests (Tenenboim and Brotman, 2016). They can either develop physical barriers (like waxy cuticle layer or rigid cell walls) or can exhibit phytochemical resistance (Nurnberger and Lipka, 2005; Bektas and Eulgem, 2015). Nevertheless, phytochemical resistance over physical defense is found prevalent in plants challenged by various pathogens (Strauch *et al.,* 2015). For phytochemical resistance, plants synthesize numerous natural products (metabolites) against pathogens and herbivores during stress conditions (War *et al*., 2012; Figure 3.1.1). Information regarding the involvement of some metabolites, such as plant phenolic compounds, in soybeans defending against soybean cyst nematode (SCN) is sparse, and much of the present knowledge of root metabolism comes from the efforts done by transcriptomics-based pathways analysis (Klink *et al*., 2007 and Wan *et al*., 2015) and targeted metabolomics studies (Huang *et al*., 1991; Durner *et al*., 1997 and Nguyen *et al*., 2013). Though, targeted metabolomics can be used to study the changes that occur in metabolic profiles as a result of organismal response to the external stimulus or stressors (Hong *et al.,* 2012 and Jorge *et al.,* 2016). However, it is an emerging tool to be used in detection of specific compounds and metabolic pathways which are previously known for their direct involvement in plant-pathogen interactions (Hofmann *et al.,* 2010). Therefore, understanding of genome-wide novel complex mechanisms underlying in plant-pathogen interactions, in particular, the metabolites and pathways involved in broad-spectrum resistance remains largely unclear by targeted metabolomics. Linking plant defense with these unknown metabolites requires a comprehensive analysis of genome-wide

metabolome profile, also known as untargeted metabolomics (Tenenboim and Brotman, 2016).

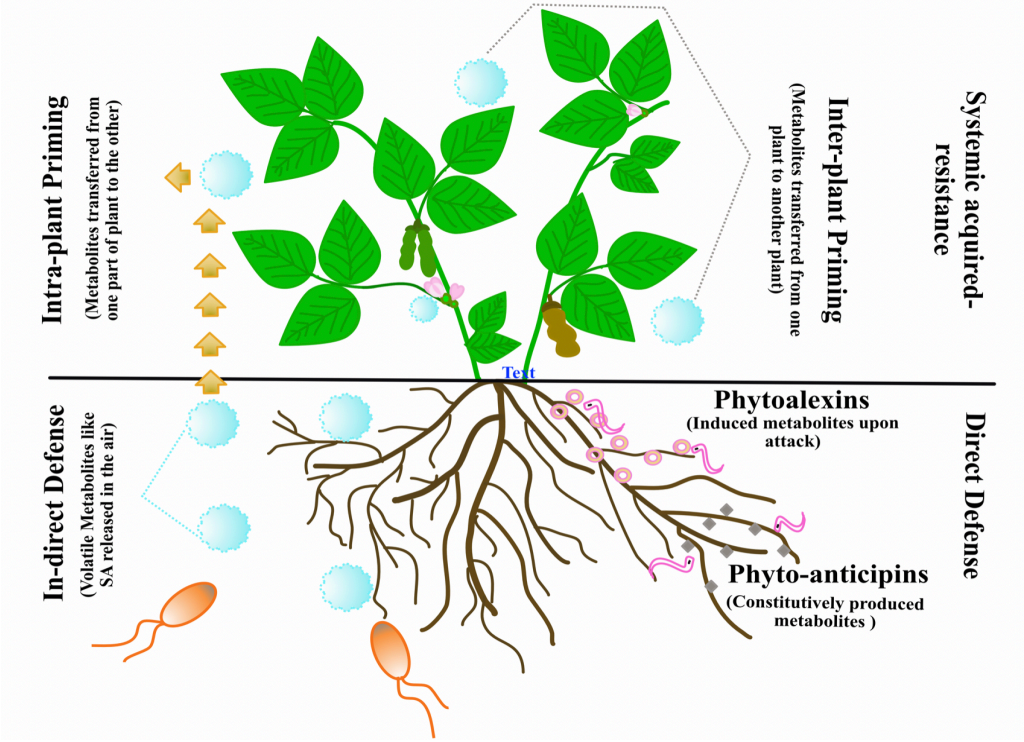


Figure 3.1.1: Direct and indirect defense mechanisms which mainly showed by plant and crop species upon biotic stress or pathogen infections;

Liquid chromatography-mass spectrometry (LC-MS) is an innovative approach capable of performing non-targeted measurement of hundreds of compounds in complex biological samples (Strauch *et al*., 2015). There are few studies that have been reported among different plant species which show successful determination of compounds and metabolic pathways involved in pathogen mediated defense responses using non-targeted metabolomics. There is one report in Arabidopsis roots, which shows induction of primary metabolites with high accumulation of 1-kestose, raffinose and α, α-trehalose sugars due to colonization of parasitic beet cyst nematode *Heterodera schachtii* (Hofmann *et al.,*

2010) when using untargeted metabolomics. Also, other reports detected similar results in white clover and soybean roots against clover cyst nematode (*Heterodera trifolii*) and soybean cyst nematode (*Heterodera glycine*; HG type 0, SCN3) respectively. These studies showed the induction of volatile plant hormones (salicylic acid, methyl salicylic acid) and other secondary metabolites (Zhu and Park, 2005 and Lin *et al.,* 2016) based on the non-targeted metabolomics results.

Conversely, genome-wide non-targeted metabolomic studies in response to newly evolving SCN races for identification of resistant-related (RR) metabolites and concomitant metabolic pathways are limited. HG type 1.2.5.7 (SCN2) and HG type 2.5.7 (SCN5) are becoming the dominant threats for soybeans in South Eastern US and profoundly damaging the crop (Niblack *et al.,* 2003 and Mitchum *et al.,* 2007). Therefore, elucidating the key metabolites and the associated pathways underlying SCN defense responses is critical in understanding the molecular mechanisms of soybean chemical defense against SCN as well as for developing SCN-resistant soybean varieties. Thus, under the present aim we performed a non-targeted genome-wide metabolomic analysis among two wild soybean genotypes (SCN-resistant; S54 and SCN-susceptible; S67) against SCN2 (HG type 1.2.5.7) and SCN5 (HG type 2.5.7) pathogens using LC-MS, to reveal underlying resistance mechanisms. Following objectives were chosen to conduct this study: **(1)** elucidation of key metabolites linked to SCN HG type 1.2.5.7 and SCN HG type 2.5.7 resistance **(2)** mapping of significant metabolites into corresponding biosynthetic pathways to intricate the resistant metabolic pathways, and **(3)** identification of common metabolites and pathways involved in broad-spectrum resistance for two SCN races.

In order to reveal the role of chemical complexes in host (wild soybean) - pathogen (SCN) interaction, global untargeted metabolomic analysis was performed in SCN resistant (S54) and susceptible (S67) wild soybean genotypes. Germinated seedlings of S54 and S67 genotypes were inoculated with ~1,800 hatched second stage juvenile (J2) nematodes. Non-infected plants without SCN inoculation were also included. To determine the role of duration (time) on metabolites expression levels, genotypes were analyzed at 3, 5, and 8 days' post inoculation (dpi) with three biological replicates for each sample. Totally, 18 infected and 18 non-infected root samples were collected for the analysis. All 36 root samples per race were subjected to LC-MS with MassHunterProfinder (Agilent) for peak extraction. Peak areas of processed metabolites were normalized, and cube root transformed prior to statistical analysis.

3.1.1 Global analysis of metabolic changes in *G. soja* roots infected by *H. glycines*

Untargeted strategy using LC-MS metabolomics profiling detected a total of 606 mass signals across all samples, with an average of 590 per sample. Of which, 572 and 532 high quality peaks for SCN2 and SCN5, respectively, were selected for further analysis. PCA of these expressed metabolites showed that the untargeted metabolomic data was able to distinguish both the effect of SCN infection and genotype-specific difference by principal components 1 (PC1) and PC2, respectively. The first principal component analysis (PC1) accounts for 38.7% and 34.2% variation and the second principal component (PC2) accounted 24.7% and 19.7% of variation among two genotypes against SCN2 and SCN5 respectively (Figure 3.1.1.1a and b).

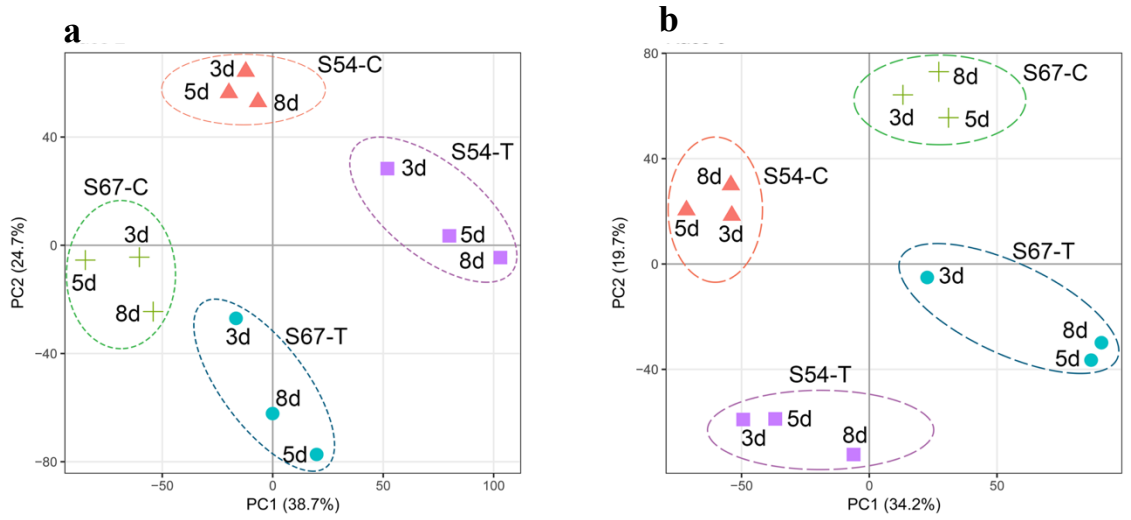


Figure 3.1.1.1: Principal component analysis of metabolites between two genotypes after 3, 5 and 8 days of **(a)** SCN2 and **(b)** SCN5 J2 nematode inoculation. Each data point in the corresponding category represents the metabolic profile of a single time point, which is averaged for three biological replicates (1-3-1). PC1: The first principal component; PC2: the second principal component. S54-C and S67-C: SCN2 and SCN5 resistant and susceptible control samples respectively; S54-T and S67-T: SCN2 and SCN5 resistant and susceptible treatment samples respectively;

Moreover, metabolites closely clustered all three testing time points in each genotype per condition while clearly separate them with other conditions. For resistant infected roots, at 5 and 8 dpi time point samples expressed differentially as compare to the 3 dpi time points when infected with SCN2. Similarly, in SCN5 infected resistant roots, replicates at 8 dpi time points was different as compare to the 3 and 5 dpi time points (Figure 3.1.1.1a and b). Clear separation between control and treatment within genotypes suggests that distinct metabolism of chemical compounds naturally exists between S54 and S67 roots, which leads to different metabolic reprogramming between them in response to *H. glycines* infection.

3.1.2 Analysis of induced metabolites in response to *H. glycine* 1.2.5.7 (SCN2) and 2.5.7 (SCN5) infections

One-way analysis of variance (ANOVA) has been performed to test the statistical significance of metabolites variation between SCN infected and non-infected roots of both resistant (S54) and susceptible (S67) genotypes. Response of each metabolite among four groups (S54-C, S54-T, S67-C and S67-T) was compared using normalized cube root transformed peak areas (Wu *et al*., 2013). The peak areas were averaged using all biological replicates for the particular group and time point. As expected, ANOVA between conditions per race across all time points indicated a strong induction of the root metabolome by *H. glycines* infection. In total, 71 metabolites were strongly induced in SCN2 treated S54 roots while 34 metabolites were significantly induced in SCN5 infected S54 roots with 17 of them being commonly induced by both SCN races (Figure 3.1.2.1a). A close investigation of the induced metabolites indicated most of the commonly differentially expressed (cDEMs) metabolites were mainly belongs to phenolic acid, isoflavonoid and flavonoid pathways (Figure 3.1.2.1b). Notably, one of the cDEM i.e. formononetin, an isoflavonoid produce in legumes (Catford *et al*., 2006 and Tsao *et al*., 2006) has been shown the involvement in disease/pest resistance previously (He *et al*., 2000).

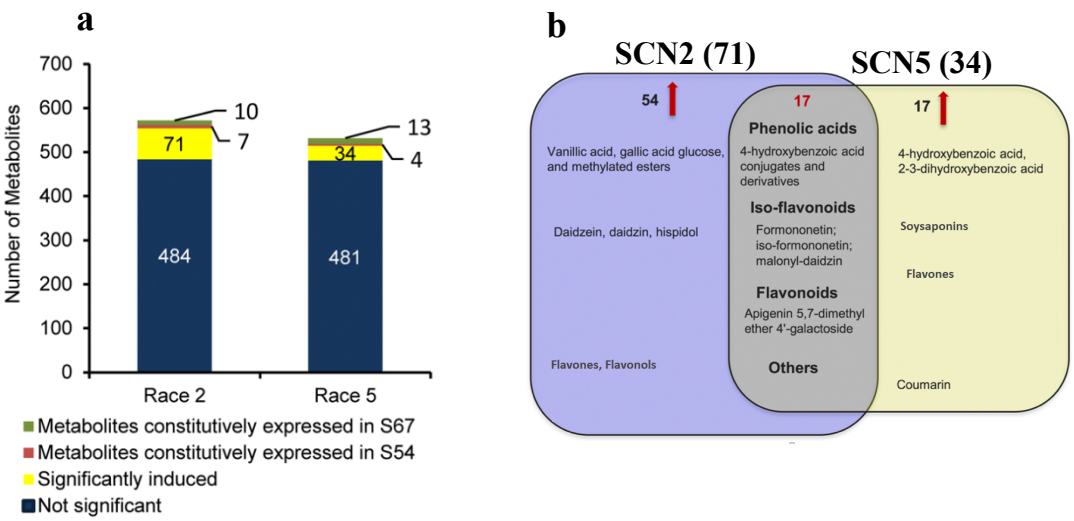


Figure 3.1.2.1: (a) Number of expressed metabolites in SCN2 and SCN5 infected roots. Constitutively expressed metabolite means, for one race, metabolites have similarly high level of expression in infected and non-infected roots of one genotype but show significantly higher level than the other genotype. **(b)** Venn diagram showing commonly induced and race-specific induced metabolites. Upward arrows represent significant induction.

3.1.3 Metabolomic profiling for time course analysis of metabolic changes in *G. soja* roots infected with SCN2 (HG type 1.2.5.7) and SCN5 (HG type 2.5.7)

To investigate the metabolite expression patterns among S54 and S67 genotypes in response to time with SCN2 and SCN5 infection, Euclidean hierarchical clustering (HCL) was performed. Out of all the detected peaks in each root extract, fifty-four (three with multiple annotation) and fifty-six metabolites were selected, which either annotated by internal standards or mass-based search in SCN2 and SCN5 root samples respectively. The annotated metabolites from both SCN races were subjected for the multivariate statistical analysis with significant test (Table 3.1.3.1a and b). A total of 36 and 34 metabolites exhibited statistically significant variations for transformed peak areas among SCN2 and SCN5 races respectively, between infected and non-infected roots of S54 and S67

genotypes at different time points or specifically at 8 time point ($p \leq 0.05$; Table 3.1.3.1a and b; Annexure II; Figure 3.1.3.1a and b).

3.1.3.1 Heat-map for root samples infected with SCN2 (HG type 1.2.5.7)

All the selected and annotated metabolites were processed for heat-map analysis and categorized in to four different groups based on their expression differences in heat map (Figure 3.1.3.1.1). Largest group (group 4) contained twenty-five metabolites of total and nineteen showed statistical significance between groups (Figure 3.1.3.1.1; Table 3.1.3.1a). Most of the metabolites found linked with conjugates of phenolic acids and iso-flavonoids. Phenolic acids' (4-hydroxybenzoic acid, 3, 4-dihydroxybenzoic acid and 3, 4, 5-trihydroxybenzoic acid) glucose and methyl esters, daidzein, daidzin, malonyl-daidzin, genistein, iso-formononetin, formononetin, and (-)-glycinol were the identified key metabolites within this group (Figure 3.1.3.1.1; Group-4). These metabolites showed a sharp increase in fold change within resistant infected root and indicate their association with SCN2 resistance. Most of the phenolic acid ester and methyl conjugates including 1-O-4-hydroxybenzoyl-β-D-glucose ester, 1-O-4-hydroxy-3-methoxybezoyl-β-D-glucose ester, 1-O-3, 4-dihydroxybenzoyl-β-D-glucose ester, and 1-O-3, 4, 5-trihydroxybenzoyl-β-D-glucose ester featured a continued upregulation from 3 dpi to 8 dpi time points and were clustered together (Figure 3.1.3.1.1; Group-4). Additionally, N-benzoyl-L-glutamate, daidzein and its malonyl (malonyl-daidzin) and methyl and glucoside derivatives (iso-formononetin, formononetin and formononetin-7-O-glucoside-6"-malonate) and pterocarpan ((-)-glycinol) also showed increased expression levels at 8 dpi time point. On the other hand, some isoflavonoids like daidzin, and its formate adduct and 2'-hydroxygenistein showed more expression at 3 and 5 dpi, as compare to 8 dpi time points (Figure 3.1.3.1.1; Group-4).

The second largest group (group-2) was found packed with fifteen metabolites included triterpenoids, phenolic acids, flavonoids and iso-flavonoids. Accumulation of hydroxy, glucoside-flavonoids and other conjugates like apigeninidin 5-O-glucoside, 8C-hexosylchrysin, and 8C-glucosyl-2-hydroxynaringenin with iso-flavonoids (glycitein, 2'-hydroxypseudobaptigenin, sojagol), phenolic acid (2, 3-dihydroxybenzoic acid) and triterpenoids (soysaponin II, soysaponin V) were found in the infected roots of both genotypes (S54 and S67). The levels of metabolites comprising apigeninidin-5-O-glucoside, 8-C-hexosylchrysin, 8C-glucosyl-2-hydroxynaringenin, furaneol-glucopyranoside were profusely raised at 5 dpi time point in infected roots of both genotypes. However, soysaponins (soysaponin V, soysaponin II), iso-flavonoids (glycitein, sojagol, 2'-hydroxypseudobaptigenin) induced more at 5 and 8 dpi time point. However, 4-hydroxybenzaldehyde induced in both genotypes at initial time point (5 dpi) but at later time point (8 dpi) it was induced more in S54 treatment as compare to the S67 treated roots (Figure 3.1.3.1.1; Group-2).

Table 3.1.3.1a: Annotated root metabolites attained from resistant and susceptible genotypes of wild soybeans with and without SCN2 infestation with exact mass, retention time and calculated *P*-values to show the statistically significant variation among different groups.

Metabolite(s)*	*P*-value at all time points F-test	*P*-value at 8 time point T-test	Retention Time (in minutes)	Exact Mass	Class of Metabolites
Group - 1					
Phlorizin	0.0014985	-	7.852	436.1368	Dihydrochalcone
Soyasaponin I	0.0007681	-	12.468	942.5218	Triterpenoid
6C-Glucosyl-2-hydroxynaringenin	0.000193	-	7.769	450.1153	Flavanone conjugate
Chalconaringenin 2'-rhamnosyl-(1-4)-glucoside	0.00046339	-	8.58	580.1778	Chalcone conjugate
Salicylic acid (AUT)	0.0000425	-	9.26	138.0312	Phenolic acid
Group - 2					
4-Hydroxybenzaldehyde (AUT)	0.00020711	-	7.09	122.0361	Phenolic acid
Furaneol glucopyranoside	N/S	N/S	6.45	306.094	Phenolic acid
Apigeninidin 5-O-glucoside	0.039926	-	7.05	416.1134	Flavonoid
8C-Hexosyl chrysin	0.017631	-	7.902	416.1103	Flavonoid conjugate
2, 3-dihydroxybenzoic acid (AUT)	0.0022216	-	6.77	154.026	Phenolic acid
8C-Glucosyl-2-hydroxynaringenin	0.00015616	-	7.77	450.1149	Flavanone conjugate
Glycitein (AUT)	0.00085626	-	10.05	284.0674	Iso-flavonoid
Soyasaponin V	0.0000402	-	10.87	958.5121	Triterpenoid
Aralidiosde	0.0084792	-	5.47	448.1205	Not Known
Asperulodisic acid	0.00086107	-	6.64	432.1257	Not Known
X	N/S	N/S	X	X	X
2'-Hydroxypseudobaptigenin	N/S	N/S	11.02	298.0459	Iso-flavonoid
Soyasaponin II	0.001674	-	12.79	912.5097	Triterpenoid
Genistin (formate adduct; AUT)	0.021175	-	8.43	478.1476	Iso-flavonoid
Sojagol	0.032077	-	14.51	336.0996	Iso-flavonoid
Group-3					
Glycitin (AUT)	N/S	N/S	7.57	446.1202	Iso-flavonoid conjugate
Glycitin (formate adduct; AUT)	N/S	N/S	7.57	492.126	Iso-flavonoid conjugate
2'-Hydroxyformononetin	N/S	N/S	7.57	284.0678	Iso-flavonoid conjugate
Glycitein 6"-O-acetylglucoside	N/S	N/S	8.42	488.1302	Iso-flavonoid conjugate
D-Phenylalanine	N/S	N/S	2.59	165.0786	Amino acid
Indole-3-acetyl-alanine	N/S	N/S	8.26	246.1001	Amino acid
Genistin (AUT)	N/S	N/S	8.29	432.1064	Iso-flavonoid conjugate
2'-Hydroxydaidzein	N/S	N/S	9.11	270.0524	Iso-flavonoid conjugate
Genistin 6"-O-acetate	N/S	N/S	9.11	474.1176	Iso-flavonoid conjugate
Group-4					
N-Benzoyl-L-glutamate	0.0019866	-	7.21	251.0789	Phenolic acid conjugate
Apigenin 5,7-dimethyl ether 4'-galactoside	0.00058847	-	7.78	460.1357	Flavonoid conjugate
Malonyl-daidzin	0.000109	-	7.18	502.1108	Iso-flavonoid conjugate

X	N/S	N/S	X	X	X
6,7-Dehydrobaicalein	N/S	N/S	9.30	268.0362	Flavonoid
Formononetin-7-O-glucoside-6''-malonate	0.0080207	-	8.25	516.1255	Iso-flavonoid conjugate
1-O-Feruloyl-β-D-glucose ester	0.0036735	-	6.1	356.1088	Phenolic acid conjugate
1-O-4-hydroxy-3-methoxybenzoyl-β-D-glucose ester	0.0120527	-	4.73	330.0942	Phenolic acid conjugate
1-O-4-hydroxybenzoyl-β-D-glucose ester	0.0006376	-	6.086	300.0843	Phenolic acid conjugate
1-O-3,4-dihydroxybenzoyl-β-D-glucose ester	0.0050486	-	5.168	316.0789	Phenolic acid conjugate
1-O-3,4,5-trihydroxybenzoyl-β-D-glucose ester	0.014431	-	1.22	332.0735	Phenolic acid conjugate
1-O-3,4-dihydroxybenzoyl-β-D-glucose ester	0.01016	-	4.44	316.0790	Phenolic acid conjugate
Daidzin (AUT)	0.016409	-	7.43	416.1115	Iso-flavonoid conjugate
Daidzin (formate adduct; AUT)	0.021016	-	7.43	462.1159	Iso-flavonoid conjugate
Hispidol	0.025519	-	8.14	254.0569	Aurone
Daidzein 6''-O-acetate	0.013099	-	8.14	458.1204	Iso-flavonoid conjugate
Apigeninidin	N/S	N/S	8.14	254.0569	Flavonoid
X	N/S	X	X	X	X
Formononetin (AUT)	0.0003498	-	12.16	268.0722	Iso-flavonoid conjugate
Isoformononetin	0.003938	-	10.12	268.0724	Isoflavonoid conjugate
Genistein (AUT)	0.033434	-	11.08	270.0518	Iso-flavonoid
(-)-Glycinol	0.0089145	-	10.4	272.0674	Pterocarpan
Daidzein (AUT)	0.00027199	-	9.83	254.0585	Iso-flavonoid
2'-Hydroxygenistein	N/S	N/S	9.54	286.0459	Iso-flavonoid
5-Amino-6-(5-phospho-D-ribitylamino) uracil	N/S	N/S	5.34	356.0734	Nucleobase

Total Significant Metabolites: 36

* All mentioned metabolites were categorized into three groups based on their expression pattern and presentation in the heat map (Figure 1.3.2.1.1). N/S: non-significant metabolites; X- metabolites with multiple annotation, showed multiple hits in the database; Highlighted metabolites are the commonly induced compounds (cDEMs) among both the races.

The first group (group-1; having 5 metabolites of total) and group-3 (with 9 metabolites of total) were mainly composed of chalcones and iso-flavonoids respectively. Metabolites including phlorizin, soysaponin I, 6C-glucosyl-2-hydroxynaringenin, chalconaringenin-2'-rhamonsyl-(1-4)-glucoside and salicylic acid (2-hydroxybenzoic acid) showed increased levels in susceptible infected roots versus resistant infected roots and non-infected roots of both genotypes at all time points (Figure 3.1.3.1.1; Group-1). Some additional metabolites mainly expressed at 3 dpi time point in the infected and non-infected roots of both resistant and susceptible genotypes were clustered in group 3. Higher levels of iso-flavanoids (glycitin, glycitin (formate adduct), glycitein-6"-O-acetylglucoside, genistin, genistin-6"-O-acetate, 2'-hydroxydaidzein) were detected in susceptible infected roots but showed less expression among other groups at 3 dpi time point. However, these metabolites showed an intense decrease at 5 and 8 dpi time points in all four groups, which indicates their trade-off mechanism in response to SCN2 infection at later days of infection (Figure 3.1.3.1.1; Group-3).

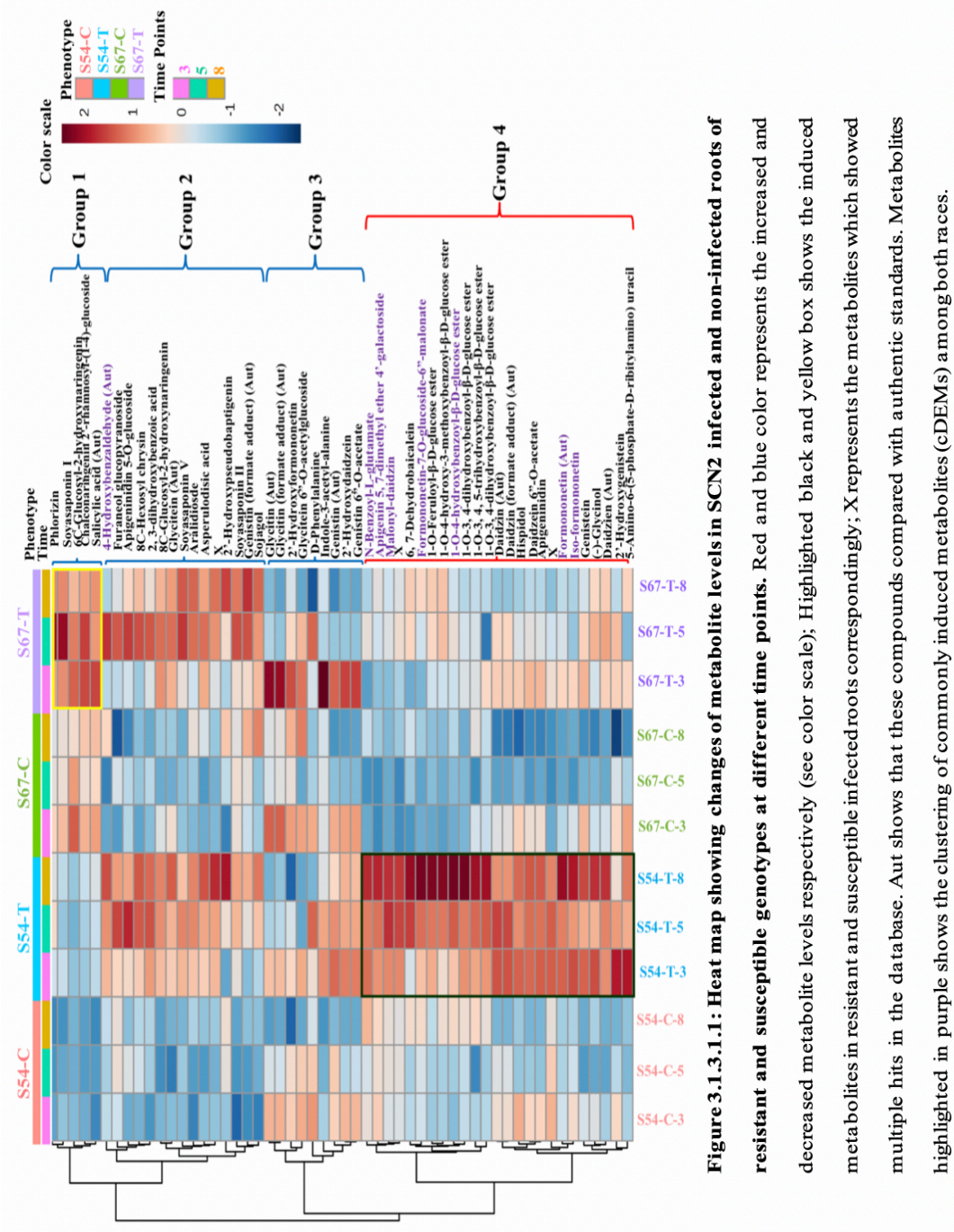


Figure 3.1.3.1.1: Heat map showing changes of metabolite levels in SCN2 infected and non-infected roots of resistant and susceptible genotypes at different time points. Red and blue color represents the increased and decreased metabolite levels respectively (see color scale); Highlighted black and yellow box shows the induced metabolites in resistant and susceptible infected roots correspondingly; X represents the metabolites which showed multiple hits in the database. Aut shows that these compounds compared with authentic standards. Metabolites highlighted in purple shows the clustering of commonly induced metabolites (cDEMs) among both races.

3.1.3.2 Heat-map for root samples infected with SCN5 (HG type 2.5.7)

The selected fifty-six metabolites infested with SCN5 were processed for multi-variate analysis in the form of heat map. Based on the expression differences among compounds they mainly differentiated in to the three groups, including two major groups i.e. group-2 and group-3, which were found packed with twenty and twenty-four metabolites respectively (Figure 3.1.3.2.1; Table 3.1.3.1b). Out of detected twenty metabolites in group 2, twelve showed statistical significance between groups and majority of them were found induced in the infected root samples of resistant (S54) genotype. The induced metabolites were either annotated phenolic acids, iso-flavonoids or their conjugates (Figure 3.1.3.2.1, Group-2; Table 3.1.3.1b). Iso-flavonoids (isoformononetin, formononetin, phlorizin), their glucose and malonyl conjugates (formononetin-7-O-glucoside-6”-malonate and malonyl-daidzin), flavonoid conjugate (apigenin 5, 7-dimethylether-4’-galactoside) and phenolic acids (4-hydroxybenzaldehyde and 2, 3-dihydroxybenzoic acid) showed a sharp increase in fold change within resistant infected roots at 8 dpi, which indicates their association with SCN5 resistance specifically at later days of infection. On the other hand, phenolic acids (N-benzoyl-L-glutamate, and 4-hydroxybenzoic acid) and their glucose and ester conjugate i.e. 1-O-feruloyl-β-D-glucose ester, 1-O-4-hydroxybenzoyl-β-D-glucose ester and 1-O-3, 4-dihydroxybenzoyl-β-D-glucose ester showed consistent inducement in resistant infected roots as compare to its non-infected and susceptible root samples. However, D-phenylalanine and soyasaponin V showed more expression at 3 and 5 dpi time points as compare to the 8dpi time point (Figure 3.1.3.2.1; Group-2).

The second largest group (group-3) was found packed with twenty-four metabolites included flavonoids, iso-flavonoids, their conjugates and soyasaponins. Accumulation of iso-flavonoids (daidzein) and their hydroxy (2'-hydroxygenistein, and 2'-hydroxyformononetin), acetyl (genistein 6"-O-acetate) and glucoside (genistin, genistin (formate adduct), glycitein 6"-O-acetylglucoside, glycitin (formate adduct)) conjugates, with flavonoid conjugates (6C and 8C-glucosyl-2-hydroxynaringenin) and soyasaponins (soyasaponin I, II and III) were found more in the susceptible (S67) infected roots as compare to their non-infected roots. Also, these metabolites showed decrease levels in both infected and non-infected roots of resistant genotype (Figure 3.1.3.2.1; Group-3). However, daidzein conjugates including daidzin, daidzin (formate adduct), and other metabolites (hispidol, 6, 7 dehydrobaicalein and apigeninidin 5-O-glucoside) mainly showed higher levels in the non-infected roots of susceptible genotype. All other groups including susceptible infected roots showed decrease levels of these compounds (Figure 3.1.3.2.1; Group-3).

Table 3.1.3.1b: Annotated root metabolites attained from resistant and susceptible genotypes of wild soybeans with and without SCN5 infestation with exact mass, retention time and calculated *P*-values to show the statistically significant variation among different groups.

Metabolite(s)*	***P*-value at all time points F-test**	***P*-value at 8 time point T-test**	**Retention Time (in minutes)**	**Exact Mass**	**Class of Metabolites**
Group - 1					
2'-Hydroxypseudobaptigenin	N/S	N/S	11.02	298.0459	Iso-flavonoid
Sojagol	0.085885	-	14.51	336.0996	Iso-flavonoid
Aralidioside	N/S	N/S	5.47	448.1205	Not Known
Soyasaponin V	0.0500000	-	10.92	958.511	Triterpenoid
Indole-3-acetyl-alanine	0.0035423	-	8.26	246.1001	Amino acid
Asperulosidic acid	N/S	N/S	6.64	432.1257	Not Known
Glycitein	0.006846	-	10.05	284.0674	Iso-flavonoid
(-)-Glycinol	0.028802	-	10.4	272.0674	Pterocarpan
Genistein	0.04865	-	11.08	270.0518	Iso-flavonoid
(S)-Furanopetasitin	N/S	N/S	7.04	432.199	Not Known
1-O-3, 4, 5-trihydroxybenzoyl-β-D-glucose ester	0.00027485	-	1.22	332.0735	Phenolic acid conjugate
Salicylic Acid (Aut)	0.00020479	-	9.26	138.0312	Phenolic acid
Group-2					
Iso-Formononetin (Aut)	0.54043	0.000502	10.12	268.0724	Iso-flavonoid
Formononetin (Aut)	0.94196	0.011800	12.16	268.0722	Iso-flavonoid
Formononetin-7-O-glucoside-6''-malonate	N/S	N/S	8.25	516.1255	Iso-flavonoid conjugate
Phlorizin	0.0010404	-	7.86	436.1364	Dihydrochalcone
Phlorizin	N/S	N/S	6.93	436.1365	Dihydrochalcone
Malonyl-daidzin	N/S	N/S	7.18	502.1108	Iso-flavonoid conjugate
4-hydroxybenzaldehyde (Aut)	0.034589	-	7.09	122.0361	Phenolic acid
Apigenin 5,7-dimethylether 4'-galactoside	0.065645	0.00138	7.78	460.1357	Flavonoid conjugate
2, 3-dihydroxybenzoic acid (Aut)	0.045866	-	6.77	154.026	Phenolic acid
1-O-3, 4-dihydroxybenzoyl-β-D-glucose ester	N/S	N/S	4.44	316.079	Phenolic acid conjugate
D-Phenylalanine	N/S	N/S	2.59	165.0786	Amino acid
Soyasaponin V	0.049226	-	12.38	958.5143	Triterpenoid
1-O-Feruloyl-β-D-glucose ester	N/S	N/S	6.1	356.1088	Phenolic acid conjugate
N-benzoyl-L-glutamate	0.0050287	-	7.21	251.0789	Phenolic acid conjugate
5-amino-6-(5-phospho-D-ribitylamino) uracil	N/S	N/S	6.36	356.074	Nucleobase
8C-hexosyl chrysin	N/S	N/S	7.9	416.1102	Flavonoid conjugate
Soyasaponin II	0.011076	-	12.79	912.5097	Triterpenoids
1-O-3, 4-dihydroxybenzoyl-β-D-glucose ester	0.037122	-	5.168	316.0789	Phenolic acid conjugate

1-O-4-hydroxybenzoyl-β-D-glucose ester	0.00077871	-	6.08	300.0851	Phenolic acid conjugate
4-hydroxybenzoic acid (Aut)	0.00011068	-	6.08	138.031	Phenolic acid
Group-3					
Daidzin (Aut)	N/S	N/S	7.43	416.1115	Iso-flavonoid
Daidzin [(formate adduct) (Aut)]	N/S	N/S	7.43	462.1159	Iso-flavonoid
Daidzein-6"-O-acetate	0.027832	-	8.14	458.1204	Iso-flavonoid conjugate
Hispidol	0.032900	-	8.14	254.0569	Aurone
6, 7-dehydrobaicalein	N/S	N/S	9.3	268.0362	Flavonoid
Apigeninidin 5-O-glucoside	N/S	N/S	7.05	416.1134	Flavonoid conjugate
Furaneol glucopyranoside	0.035924	-	6.06	306.944	Phenolic acid
Daidzein-6"-O-acetate	0.026541	-	8.12	458.1204	Iso-flavonoid conjugate
Apigeninidin	N/S	N/S	8.14	254.0569	Flavonoid
Genistin (Aut)	0.023503	-	8.29	432.1064	Iso-flavonoid conjugate
2'-hydroxydaidzein	0.014795	-	9.11	270.0524	Iso-flavonoid conjugate
Genistin 6"-O-acetate	0.0058534	-	9.11	474.1176	Iso-flavonoid conjugate
Daidzein (Aut)	N/S	N/S	9.83	254.0585	Iso-flavonoid
Soyasaponin II	0.021076	-	12.79	912.0579	Triterpenoids
8C-glucosyl-2-hydroxynaringenin	N/S	N/S	7.12	450.1161	Flavonoid conjugate
Genistin [(formate adduct) (Aut)]	0.028461	-	8.43	478.1476	Iso-flavonoid conjugate
2'-hydroxyformononetin	0.025137	-	7.57	284.0678	Iso-flavonoid conjugate
Glycitein 6"-O-acetylglucoside	N/S	N/S	8.42	488.1302	Iso-flavonoid conjugate
6C-glucosyl-2-hydroxynaringenin	0.00031987	-	7.77	450.1153	Flavonoid conjugate
Soyasaponin I	0.00054305	-	12.48	942.5218	Triterpenoids
Glycitin [(formate adduct) (Aut)]	0.0000176	-	7.57	492.126	Iso-flavonoid conjugate
2'-hydroxygenistein	N/S	N/S	9.54	286.0459	Iso-flavonoid conjugate
Chalconaringenin 2'-rhamnosyl-(1-4)-glucoside	N/S	N/S	7.66	580.1779	Chalcone conjugate
Soysaponin III	0.0000023	-	12.84	796.4616	Triterpenoids
Total Significant Metabolites: 34					

*All mentioned metabolites were categorized into three groups based on their expression pattern and presentation in the heat map (Figure 1.3.2.2.1); N/S: non-significant metabolites; Highlighted metabolites are the commonly induced compounds (cDEMs) among both the races.

The first group (group-1) was having twelve metabolites in total, which were mainly belongs to the iso-flavonoid, phenolic acids, pterocarpans or terpenoids. Metabolites including 2'-hydroxypseduobaptigenin, sojagol, aralidioside, indole-3-acetyl-alanine, asperulosidic acid, glycitein, (-)-glycinol, genistein and salicylic acid were all showed increased levels in infected roots versus non-infected roots among both genotypes. However, most of the metabolites showed intense expression in susceptible infected roots than the resistant infected roots at 5 dpi and 8 dpi time points specifically (Figure 3.1.3.2.1; Group-1). This indicates that both resistant and susceptible genotype might induce similar compounds initially, however later S54 has possibly trade-off these compounds to incorporate energy in production of other significant metabolites like induced compounds of group-2 (Figure 3.1.3.2.1; Group-2). This dissimilarity in the metabolites expression can be linked to the susceptibility of susceptible (S67) genotype as compare to the resistant (S54) genotype against SCN5 nematode pathogenicity.

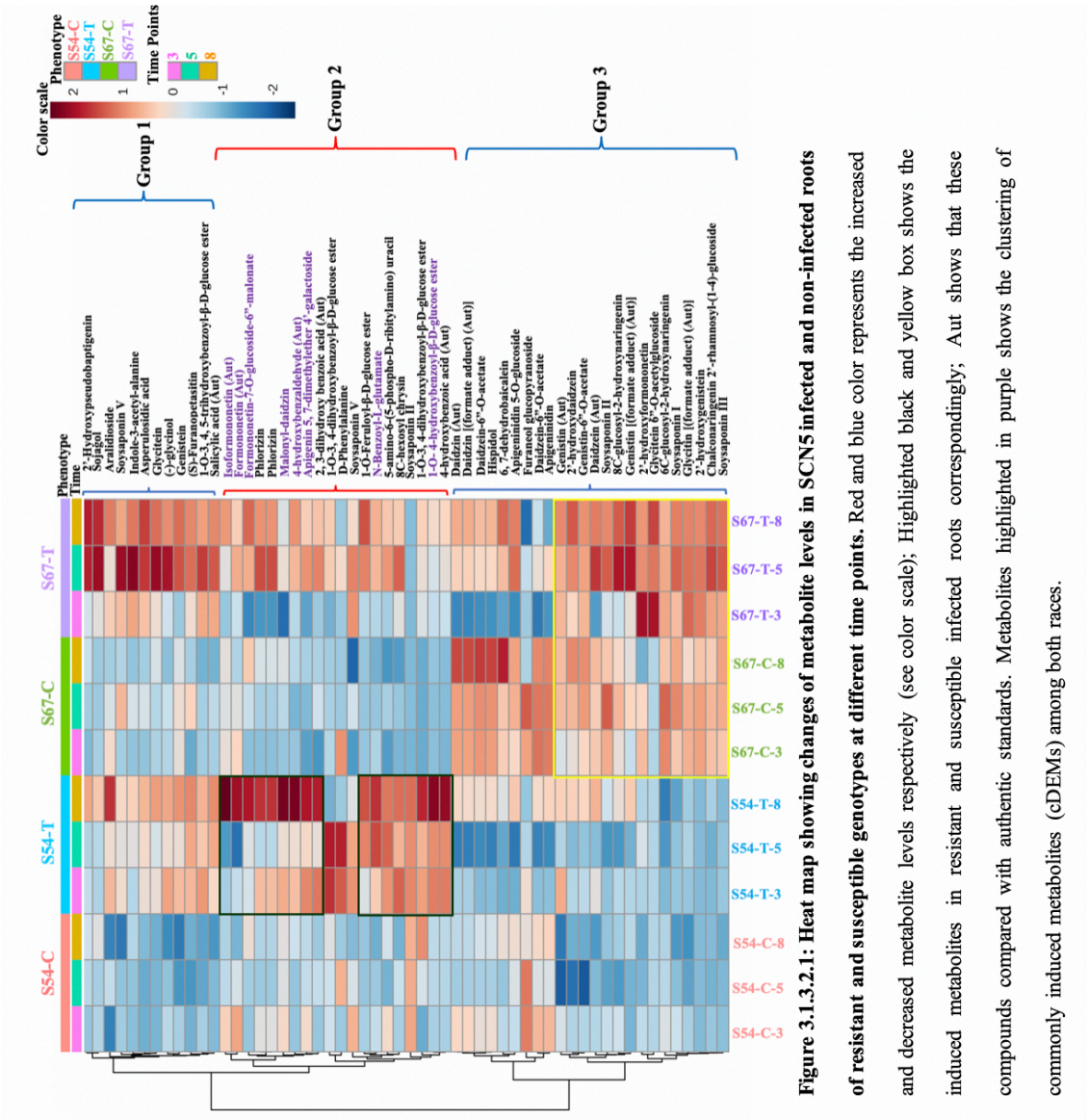


Figure 3.1.3.2.1: Heat map showing changes of metabolite levels in SCN5 infected and non-infected roots of resistant and susceptible genotypes at different time points. Red and blue color represents the increased and decreased metabolite levels respectively (see color scale); Highlighted black and yellow box shows the induced metabolites in resistant and susceptible infected roots correspondingly; Aut shows that these compounds compared with authentic standards. Metabolites highlighted in purple shows the clustering of commonly induced metabolites (cDEMs) among both races.

3.1.4 Metabolic pathways allied with HG type 1.2.5.7 (SCN2) and HG type 2.5.7 (SCN5) infections

SCN2 and SCN5 confrontation in resistant (S54) genotype was mainly found associated with the accumulation of phenolic acid, isoflavonoids and their conjugates (Figure 3.1.3.1.1; Group-4 and Figure 3.1.3.2.1; Group-2). These metabolites are mainly produced by chorismic acid via either isochorismic acid or D-phenylalanine and belong to the two major pathways i.e. phenolic acid pathway and isoflavonoid pathway (Pieterse *et al.,* 2012; An and Mou, 2014 and Shine *et al.,* 2016). Isoflavonoid biosynthesis was differentially affected by both SCN races infection among resistant and susceptible genotypes. Isoflavonoids like daidzein (against SCN2 infection); genistein derived from liquiritigenin and naringenin flavanones respectively, increased in roots of resistant infected genotype. Also, glucoside, malonyl and methylated derivatives of daidzein such as the daidzin, malonyl-daidzin, iso-formononetin, and formononetin were accumulated more in root of resistant infected genotype. Phenolic acid (2, 3-dihydroxy benzoic acid induce against SCN5 infection) and their methylated and glucoside conjugates like N-benzoyl-L-glutamate, 1-O-4-hydroxybenzoyl-β-D-glucose ester, 1-O-4-hydroxy-3-methoxybezoyl-β-D-glucose ester, 1-O-3, 4-dihydroxybenzoyl-β-D-glucose ester, and 1-O-3, 4, 5-trihydroxybenzoyl-β-D-glucose ester, can either be produced by positional isomerization of salicylic acid (2-hydroxybenzoic acid) to 4-hydroxybenzoic acid and then conjugation with glucose or methyl functional groups or can be produced via benzaldehyde to 4-hydroxybenzoic acid and then further get conjugated (Figure 3.1.4.1a and b).

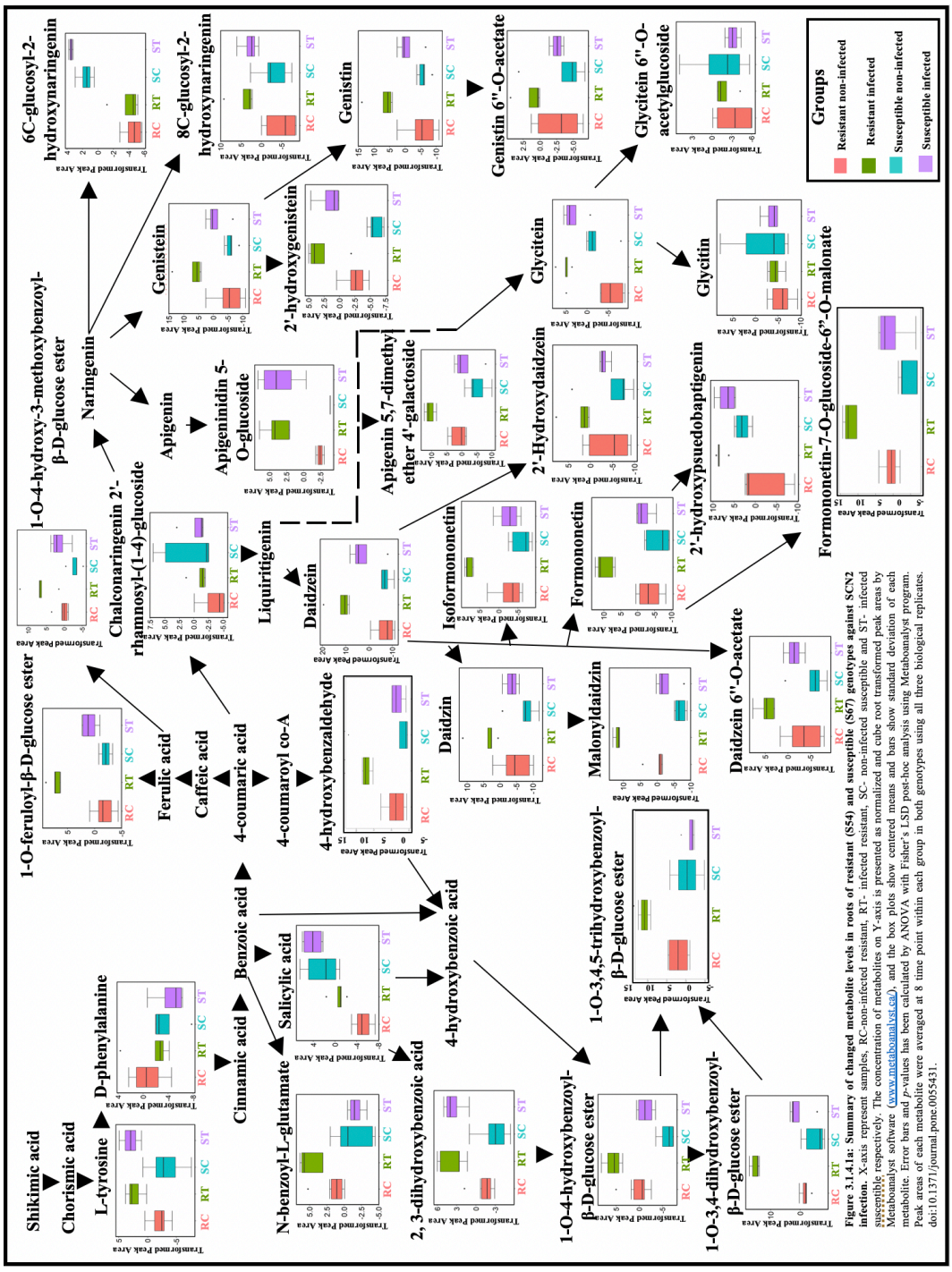


Figure 3.1.4.1a: Summary of changed metabolite levels in roots of resistant (S54) and susceptible (S67) genotypes against SCN2 infection. X-axis represent samples, RC-non-infected resistant, RT- infected resistant, SC- non-infected susceptible and ST- infected susceptible respectively. The concentration of metabolites on Y-axis is presented as normalized and cube root transformed peak areas by Metaboanalyst software (www.metaboanalyst.ca), and the box plots show centered means and bars show standard deviation of each metabolite. Error bars and *p*-values have been calculated by ANOVA with Fisher's LSD post-hoc analysis using Metaboanalyst program. Peak areas of each metabolite were averaged at 8 time point within each group in both genotypes using all three biological replicates. doi:10.1371/journal.pone.0055431.

In contrast, some other iso-flavonoids (like glycitein, glycitin, and 2'-hydroxypseudobaptigenin) synthesized in iso-flavonoid pathway, produce about similar concentration in infected roots of both S54 and S67 genotypes but low levels were detected in non-infected groups when tested against SCN2. However, iso-flavonoids like daidzein, genistein and other conjugated compounds such as glycitein-6"-O-acetylglucoside, genistein-6"-O-acetate, 2'-hydroxydaidzein, 2'-hydroxyformononetin, and 2'-hydroxygenistein were induced in the infected roots of susceptible genotype when inoculated with HG type 2.5.7 (SCN5). Further, Apigenin 5, 7-dimethyl ether 4'-galactoside (methylated and glucoside form of apigenin flavonoid) was the only flavonoid which induces in resistant infected roots against both SCN infection. On the other hand, flavanones derivatives (glucosyl derivatives of naringenin) 8C-glucosyl-2-hydroxynaringenin and 6C-glucosyl-2-hydroxynaringenin were either induced or produce equally (to infected roots of resistant genotype) in infected roots of susceptible genotype (Figure 3.1.4.1a and b). These findings indicate the more association of 2- or 4-hydroxybenzoic acid (a branch of phenolic acid pathway) derived metabolic pathway and iso-flavonoid pathways in broad-spectrum (against two SCN races; SCN2 and SCN5) resistance as compare to the flavonoid and other phenolic acid pathways.

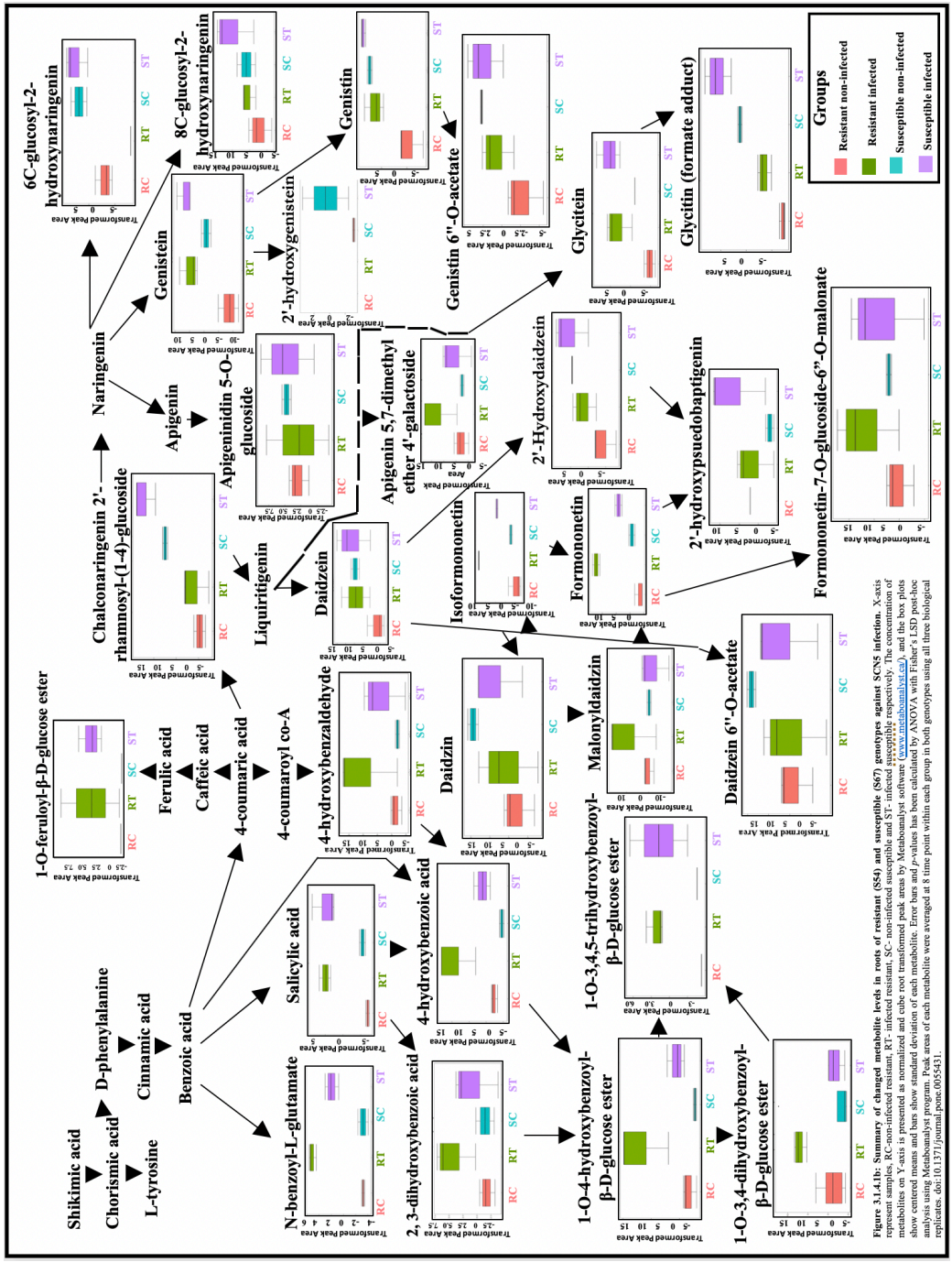


Figure 3.1.4.1b: Summary of changed metabolite levels in roots of resistant (S54) and susceptible (S67) genotypes against SCN5 infection. X-axis represent samples, RC-non-infected resistant, RT- infected resistant, SC- non-infected susceptible and ST- infected susceptible respectively. The concentration of metabolites on Y-axis is presented as normalized and cube root transformed peak areas by Metaboanalyst software (www.metaboanalyst.ca/), and the box plots show centered means and bars show standard deviation of each metabolite. Error bars and *p*-values has been calculated by ANOVA with Fisher's LSD post-hoc analysis using Metaboanalyst program. Peak areas of each metabolite were averaged at 8 time point within each group in both genotypes using all three biological replicates. doi:10.1371/journal.pone.0055431.

3.2 Quantitative Expression Analysis of Phenyl-propanoid/Iso-flavonoid Pathway Genes

Plants' responses to stress conditions constitute the activation of genetic cascades, which result in adaptive responses such as the up regulation of metabolic compounds. In the present study, defense response of SCN resistant wild soybean (S54) genotype was characterized by higher accumulation of iso-flavonoids, phenolic acids and their derivatives & conjugates (Figure 3.1.3.1.1 and 3.1.3.2.1) from the phenyl-propanoid pathway when expose to nematodes. These results indicate the defensive role of phenyl-propanoid compounds in SCN stress, which further suggests the involvement of their biosynthetic genes in nematode resistance. Expression analysis is one of the best approaches to determine the role of a biosynthetic gene under certain conditions (Zhao *et al*., 2013). Transcriptomic gene expression studies have become an important discipline for functional genomics, systems biology, and biotechnology. Global transcriptional changes have been examined in several plant species during various pathogen infections (Moy *et al*., 2004; Zabala *et al*., 2006; Soria-Guerra *et al*., 2010; Mazarei *et al*., 2011; Ishiga *et al*., 2015 and Xu *et al*., 2015). Recently, our laboratory has also reported the involvement of calcium/calmodulin-mediated defense signaling, jasmonic acid (JA)/ethylene (ET) and salicylic acid (SA)-involved signaling, the MAPK signaling cascade, and WRKY-involved transcriptional regulation in wild soybean against SCN infection (Zhang *et al*., 2017).

Though transcriptomic analysis can reveal the complexity of biosynthetic genes expressed globally by a given plant species in a specific circumstance, but the relative expression of different candidate genes is more useful to determine their direct involvement

in biotic stress (Zhao *et al*., 2013). The analysis of the variation in gene expression between different plant genotypes can provide insight into the genes, proteins, and linked pathways during a pathogen infection, which can provide a valuable readout of the molecular processes underlying resistance and susceptibility (Jadhav *et al*., 2013). Phenyl-propanoid metabolism triggers a cascade of biochemical reactions upon pathogen infections which lead to the changes in expression level of several linked genes. Some of the recent studies have shown the differential expression of phenyl-propanoid pathway genes during biotic and abiotic stresses (Chen *et al*., 2009; Soybean - *Rhizoctonia solani*; Sepiol *et al*., 2017; Soybean - *Phytophthora sojae* and Gutierrez-Gonzalez *et al*., 2010; Soybean – Drought stress). However, there is no report on expression analysis of phenyl-propanoid/iso-flavonoid pathway genes during wild soybean – SCN interaction. Therefore, the present investigation was conducted to study the expression profile of key genes of the phenyl-propanoid/iso-flavonoid pathway which can be responsible for the high accumulation of defense metabolites belongs to above pathways during SCN attack. We quantified the transcript (mRNA) abundance of selected key genes from phenyl-propanoid/iso-flavonoid pathway (Figure 3.2.1) in the roots of resistant (S54) and susceptible (S67) genotypes of wild soybeans and investigated their relative expression differences after 3, 5 and 8 days of SCN infection. The key genes used for the expression analysis either has been selected based on previous literature (*Phenylalanine ammonia lyase* (*PAL*); Zhang *et al*., 2016; *Chalcone synthase* (*CHS7* and *CHS8*); Yi *et al*., 2010; *2'-hydroxyisoflavone reductase* (*IFR*); Cheng *et al*., 2015 and 3, 9-*dihydroxypterocarpan 6a-monooxygenase* (*CYP93A1*); Kinzler *et al*., 2016), which showed the role of phenyl-propanoid pathway genes in plant-pathogen interaction, or based on the expression differences in roots of resistant (S54) and

susceptible (S67) wild soybean genotypes determined by transcriptomic analysis (*Hydroxyisoflavanone/isoflavone-4'-methyl-transferase* (*HI4'OMT/I4'OMT*)) when infected with SCN pathogen (Annexure II; Table 3.2.1).

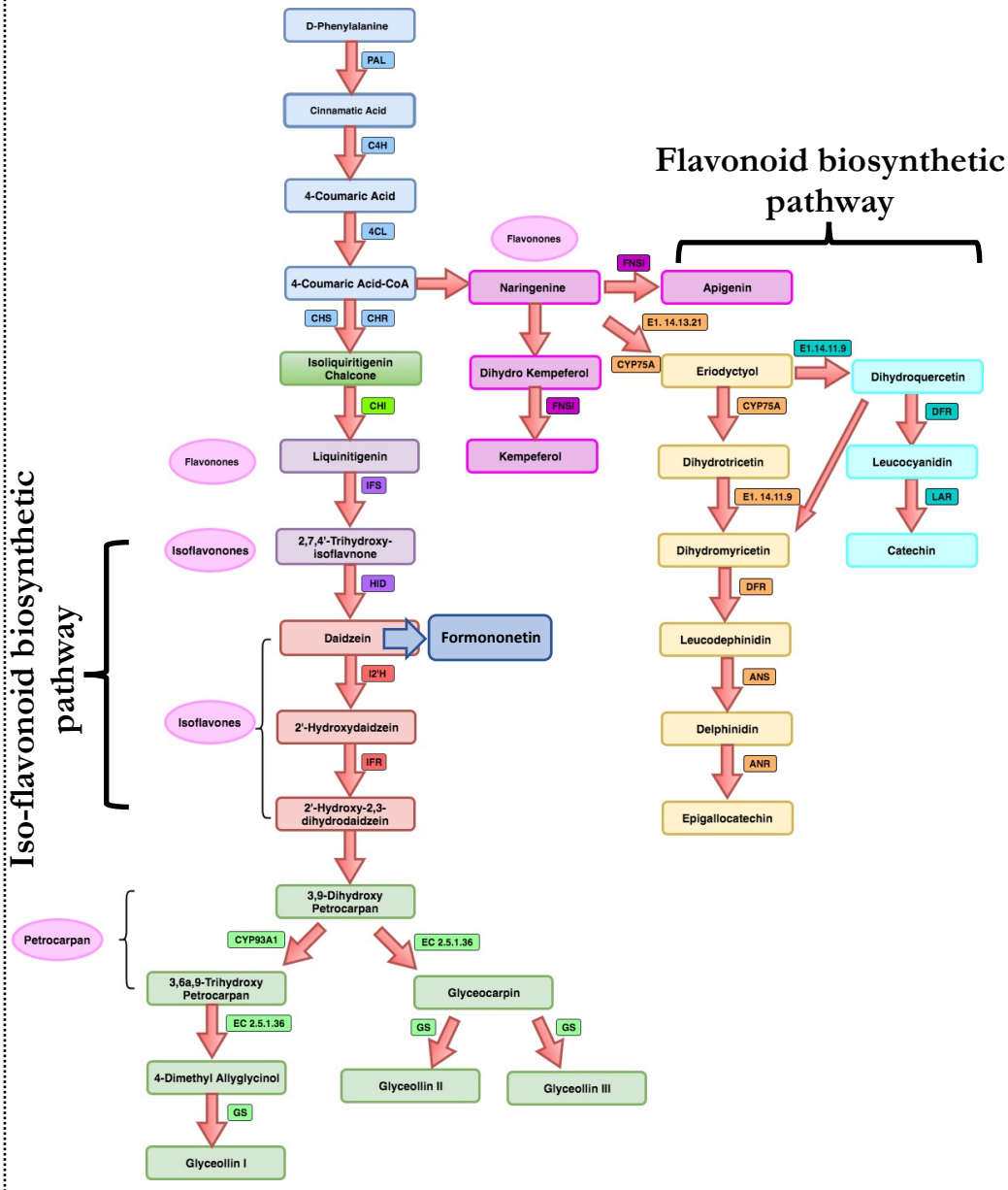


Figure 3.2.1: Figure is showing the step by step synthesis of different iso-flavonoids and flavonoids *via* phenyl-propanoid pathway in plants. The full form of abbreviated enzymes is provided in Annexure II, Table 3.2.1.

Differences in phenolic acid and isoflavone metabolic contents between SCN-resistant (S54) and SCN-susceptible (S67) genotypes suggested that possibly expression differences of phenyl-propanoid pathway genes influenced this metabolic variation. To test this hypothesis, roots of S54 and S67 genotypes were initially inoculated with SCN5 (HG type 2.5.7) juvenile (J2 stage) nematodes and with distilled water (no nematodes). Root samples were collected after 3, 5, 8 days of nematode inoculation to conduct qRT-PCR with primers targeting the most important genes for iso-flavonoid and phenolic acid biosynthesis (Table 2.2.5.2.1). The phenyl-propanoid pathway genes for quantitative real-time PCR (qRT-PCR) analysis were either selected based on previous literature (Figure 3.2.1) or gene expression variation in resistance (S54) and susceptible (S67) genotypes through genome-wide RNAseq (transcriptomic) analysis (Annexure II; Table 3.2.1). We pursued three objectives: (i) development of a profile of phenyl-propanoid pathway genes under both nematodes inoculated and control conditions; (ii) examination of the relationships between time variation of SCN inoculation with gene expression; and (iii) investigation of possible correlations of expression of phenyl-propanoid pathway genes with phenolic acid and iso-flavonoids accumulation.

3.2.1 Root isoflavone and phenolic acid biosynthetic gene transcript profiling under SCN5 (HG type 2.5.7) attack

Relative quantification of different genes belonging to phenyl-propanoid pathway revealed increased expression of many genes including *PAL*, *CHS* and *HI4'OMT*/*I4'OMT*. However, results showed that induced genes have differential expression pattern depending on the SCN infection duration (Figure 3.2.1.1). Two highly induced *PAL* (*PAL1.1* (Glyma.03g181600), and *PAL1.2* (Glyma.03g181700)) genes that encode *phenylalanine*

ammonia lyase enzyme, which catalyse the first key step of the phenyl-propanoid pathway, were showed significant increase in the expression at 3 and 5 dpi as compare to 8 dpi. On the other hand, genes including *CHS7* (Glyma.01g228700), *CHS8* (Glyma.11g110500) and *HI4'OMT/I4'OMT* (Glyma.13g173300) had lower expression at early stages of SCN infection (3 dpi), as compare to later days (5 and 8 dpi) of infection (Figure 3.2.1.1). A significant increase of *HI4'OMT/I4'OMT* gene at 8 dpi agreed with metabolic data, as this gene encodes for *Hydroxyisoflavanone/isoflavone-4'-O-methyl-transferase* enzyme, which catalyses the synthesis of formononetin compound, the metabolite which also highly accumulated at later days of nematode infection (8 dpi; Figure 3.1.3.1.1and 3.1.3.2.1).

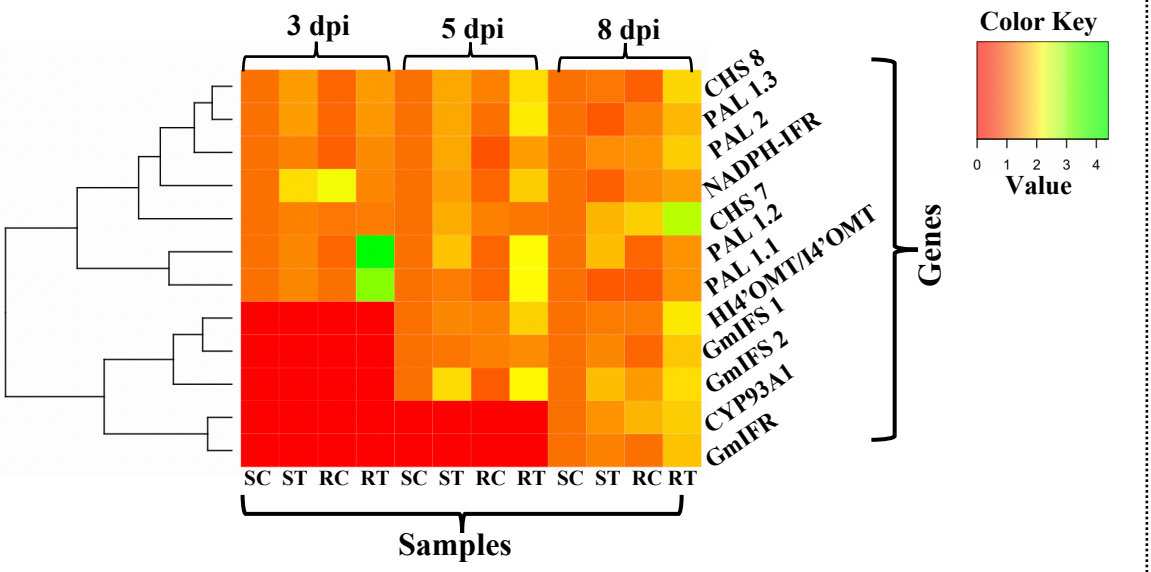
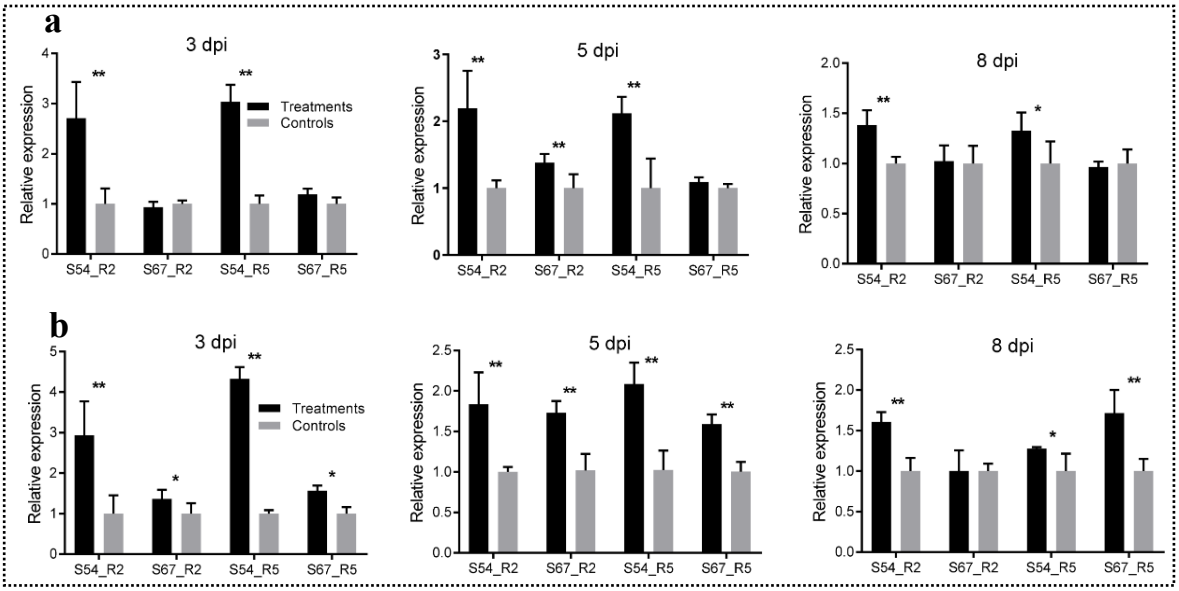


Figure 3.2.1.1: Heat map showing the expression pattern of phenyl-propanoid pathway genes among resistant (S54) and susceptible (S67) genotypes at 3, 5 and 8 dpi. Three biological replicates for both genotypes at each time point were used to generate the heat map. RT-resistant treatment (S54 treatment), ST- susceptible treatment (S67 treatment), RC-resistant control (S54 control) and SC-susceptible control (S67 control); dpi- days post infection; Colour scale: green and red color represents the increased and decreased gene expression levels respectively;

3.2.2 Quantification of selected genes in resistant and susceptible roots of wild soybean infected with SCN2 (HG type 1.2.5.7) and SCN5 (HG type 2.5.7)

Higher expression of *PAL* (*PAL1.1* (Glyma.03g181600), and *PAL1.2* (Glyma.03g181700)) and *HI4'OMT*/*I4'OMT* (Glyma.13g173300) genes in resistant roots than in the susceptible ones suggested that these genes might be responsible for higher accumulation of iso-flavonoid and phenolic acid metabolites under SCN stress in resistant (S54) genotype. So, we subsequently performed *q*RT-PCR to check the expression of these genes in root tissues of resistant and susceptible genotype infected with SCN2. The *PAL* genes expression was compared at all three time points but *HI4'OMT*/*I4'OMT* was compared only at later days of infection (8 dpi) the time at which the strongest increase in the abundance of the associated metabolites had been observed.



Consistent with the SCN5 gene expression data, we found that all three selected genes showed a significantly higher expression in the infected S54 roots than in the infected S67 roots. A significant increase in the expression of two *PAL* genes was observed at 3 &

5 dpi, and *HI4'OMT/I4'OMT* gene's higher expression at 8 dpi (Figure 3.2.2.1 a-c). The similarity in the expression of these pathway genes and associated metabolites after both SCN races infection shows the robustness of integration of gene expression and metabolomics data, and suggests that these commonly induced genes (cDEGs) can be important in the diversity of the commonly induced metabolites (cDEMs) that enable S54 resistant to *H. glycines*.

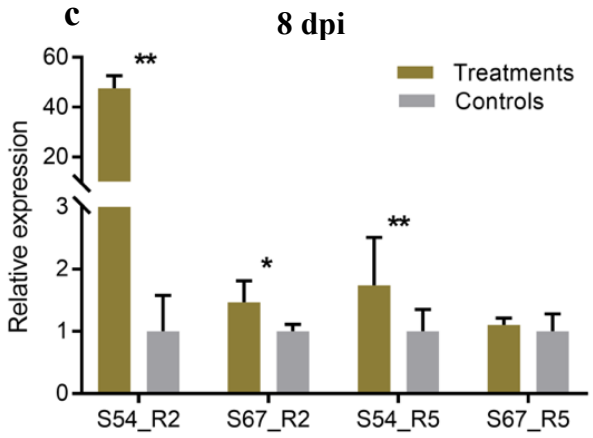


Figure 3.2.2.1: Expression patterns of three selected candidate genes involved in phenyl-propanoid pathway. Total RNA was extracted from the roots of both genotypes and was used in qRT-PCR analysis. Expression analysis of two *PAL* (*PAL1.1* (Glyma.03g181600), and *PAL1.2* (Glyma.03g181700)) and *HI4'OMT/I4'OMT* (Glyma.13g173300) genes that were upregulated in S54, can observed in subsection a, b and c respectively. The bar with asterisks is significantly different (unpaired student's *t*-test $^{*}P \leq 0.05$ and $^{**}P \leq 0.01$). S54- SCN-resistant; S67- SCN-susceptible; R2- SCN2 (HG type 1.2.5.7) and R5- SCN5 (HG type 2.5.7);

3.3 Genomic and cDNA cloning of *Hydroxyisoflavanone/Isoflavone-4'-methyl transferase* (*HI4'OMT/I4'OMT*) Gene

In recent years, exploration of evolutionary processes has been emerged as a substantial mechanism to identify gene(s) or allele(s), which influences the complex trait variation in nature (Prasad *et al*., 2012). This process also gives a conception on how genetic changes can modify plant gene expression and function in diverse environmental conditions. Genetic changes including SNPs (single nucleotide polymorphisms), indels (insertion/deletion polymorphisms), diversification within regulatory regions, alternative splicing and copy number variations are some of the key evolutionary mechanisms that have been used by several plant species for ecological adaptation (Saxena *et al*., 2014 and Marroni *et al*., 2014). One of the recent studies revealed the role of rhg-1-b (at *Rhg1* quantitative trait locus) mediated nematode resistance in SCN resistant soybean lines by copy number variation. A 31-kilobase segment at rhg-1-b contains genes for an amino acid transporter (*GmAAT*), an *alpha-SNAP* (*GmSNAP18*) protein, and a *GmWI12* (wound-inducible domain) protein, all of which contribute to resistance. However, only one copy of the 31-kilobase segment per haploid genome was found in susceptible genotype, and 10 tandem copies were detected in resistant soybean genotype. This shows the association of SCN resistance with overexpression of a set of dissimilar genes together in a repeated multigene segment (Cook *et al*., 2012). Similarly, map-based cloning of the *SHMT* (*serine hydroxymethyltransferase*) gene at *Rhg4* (one of the other major quantitative trait loci contributing resistance to *Heterodera glycines* 4) locus showed two genetic polymorphisms (SNPs) that shows amino acid changes, which alter the key regulatory property of the enzyme and was found linked with the resistance of SCN (*Heterodera glycines* 4; Liu *et*

al., 2012). So far, mainly two resistant soybean genotypes are widely used against soybean cyst nematode (SCN) pathogen. One of the resistant soybean genotypes is Peking, whose resistance requires the genes or alleles from both *rhg1-a* and *Rhg4* locus and the other type is PI88788 which needs multiple copies of *rhg1-b* locus genes' (*GmSNAP18*, *GmAAT*, and *GmWI12*; Liu *et al*., 2017). However, in the transcriptomic (RNA-seq) analysis we haven't detected the higher expression of any identified gene from *rhg1-a*, *b* and *rhg4* locus within SCN resistant genotype (S54). Also, there is no other study available which shows the clear mechanism of wild soybean resistance against SCN pathogen.

In the present study, metabolomics and gene expression analysis from section 3.1 and section 3.2 showed the higher expression of *hydroxyisoflavanone/isoflavone-4'-methyl-transferase* (*HI4'OMT*/*I4'OMT*) gene and its associated metabolites from the phenyl-propanoid pathways in resistant wild soybean (S54) genotype for SCN races 2 and 5 at intense time of SCN infection (8 dpi). In view of this, it is hypothesized that significantly increased expression of *HI4'OMT*/*I4'OMT* gene can confer non-race specific SCN resistance by enhancing the biosynthesis of significant metabolites such as iso-flavonoids and phenolic acids. Therefore, to understand which evolutionary mechanism(s) can control the *HI4'OMT*/*I4'OMT* gene expression variation among resistant and susceptible wild soybean genotypes, we carried out the molecular (cDNA and genomic DNA) cloning among them and compared their sequences in the present section.

3.3.1 *HI4'OMT* /*I4'OMT* gene variations in S54 and S67 genotypes

Hydroxyisoflavanone-4'-methyl-transferase (*HI4'OMT*) or *isoflavone-4'-methyl-transferase* (*I4'OMT*) gene from the phenyl-propanoid or iso-flavonoid pathway, have shown the higher expression in the SCN resistant genotype (S54) upon infection at 8 dpi

(Figure 3.3.1.1a-b). This gene regulates the production of formononetin metabolite, which also induces in S54 upon infection as shown in Figure 3.3.1.1c. A comparison of *HI4'OMT/I4'OMT* cDNA and genomic DNA sequences between S54, S67 and Williams 82 identified a total of 6 nucleotide differences [(two single nucleotide polymorphisms (SNPs) and two insertions/deletions (indels)]. Two of the SNP variations were resulted in an amino acid change in the predicted protein sequences [(I41V; 121 bp, A/G) and (G202D; 605 bp, G/A); Figure 3.3.1.1d-e]. However, both mutations (I41V and G202D) were predicted to be not deleterious (neutral; protein variation effect analyzer (PROVEAN) score = -0.932 and 3.157 respectively) to the protein (Figure 3.3.1.1d).

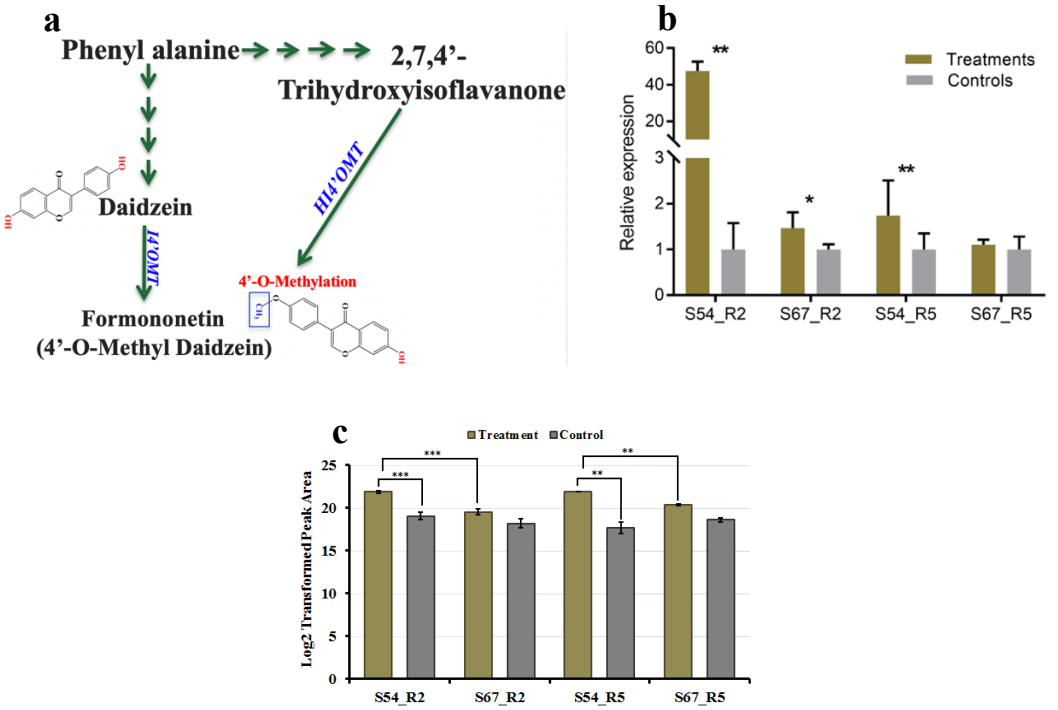
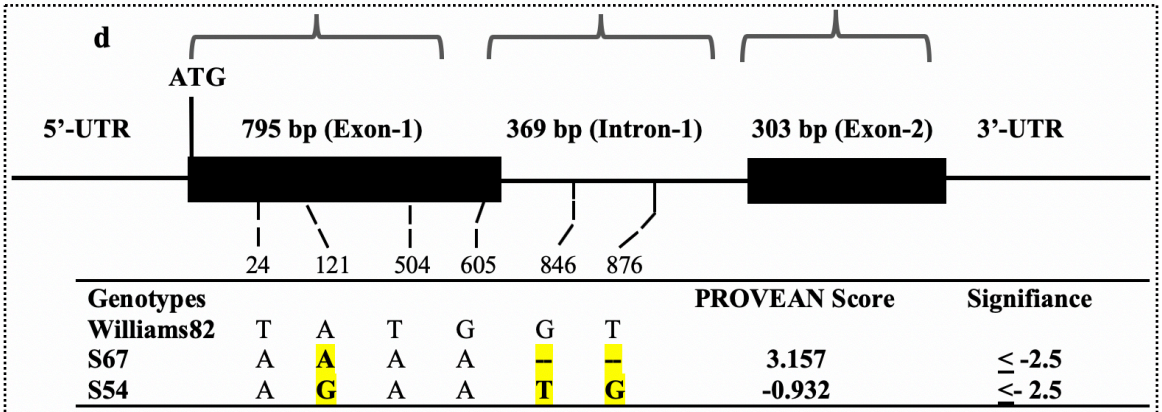


Figure 3.3.1.1: (a) Isoflavonoid pathway showing the biosynthesis of formononetin metabolite via *HI4'OMT* and *I4'OMT* enzymes; **(b)** relative expression of *HI4'OMT/I4'OMT* gene among S54 and S67 at 8 dpi; and **(c)** expression differences of formononetin metabolite among S54 and S67 at 8 dpi. (unpaired student's *t*-test $^{*}P \leq 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$);



Genotypes							PROVEAN Score	Significance
Williams82	T	A	T	G	G	T		
S67	A	A	A	A	–	–	3.157	≤ -2.5
S54	A	G	A	A	T	G	-0.932	≤- 2.5

e

Williams 82
MVFCGNSTEESELHHAQIHLYKHVYNFVSSMALKSAMELGIADVIHSHGKPMTISELSSA 60
V(S54) and I(S67)
LKLHPSKVSVLQRFLRLLTHNGFFAKTILPSKNGVEGGEEIAYALTPPSKLLIRNKSICLAPIVK 125
GALHSSSLDMWHSSKKWFSEDKELTLYESATGESFWDFLNKTTESDTLGMFQDAMAADSKV 186
FKLALEECKHVFEGLGSLVDVGGGTGVVTRLISETFPHLKCTVFDQPQVVANLTGNENLNF 247
D (S54 and S67)
VGGDMFKSIPSADAVLLKWVLHDWNDELSVKILKNCKEAISGKGKEGKVIIIDIAIDEVGDDR 310
EMTELKLDYDLVMLTMFNGKEREKKEWEKLIYEAGFSNYKIIPICGFKSLIEVYP 365

Figure 3.3.1.1: (d) *HI4'OMT/I4'OMT* gene model and polymorphisms between S54, S67 and Williams 82; **(e)** *HI4'OMT/I4'OMT* predicted protein sequences showing amino acid differences (I41V and G202D) between Williams 82, S54 and S67.

3.3.2 Haplotypes identification within *HI4'OMT/I4'OMT* gene using geographically diverse genotypes

To determine the effect of ecological diversity on *HI4'OMT/I4'OMT* gene, six geographically diverse genotypes (including both wild soybeans and soybeans) have been selected. The chosen genotypes were also diverse for SCN resistance (Table 2.3.1.1). The gene was fully sequenced in all the genotypes and a total number of six haplotypes were determined (Haplotype I to Haplotype VI). Haplotype I and haplotype II were detected in all SCN resistant and non-resistant genotypes, except the Williams-82 (this genotype was

carrying only haplotype-I). On the other hand, haplotype III, haplotype IV, haplotype V and haplotype VI were found specific for S54, S100 and PW5 respectively (Figure 3.3.2.1).

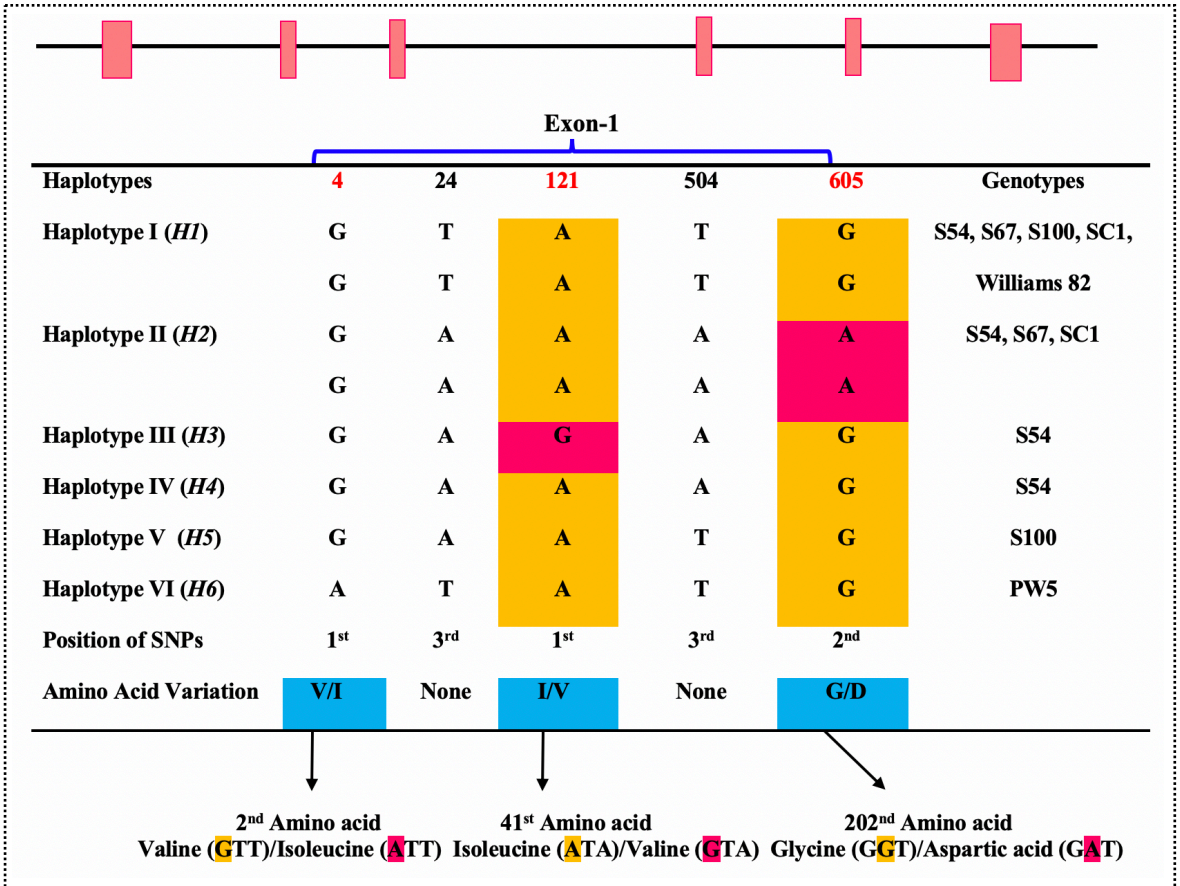


Figure 3.3.2.1: Identified haplotypes (haplotype I; *H1* to haplotype VI; *H6*) of *HI4'OMT/I4'OMT* gene among wild soybean (S54, S67, S100, PW5) and soybean (SC1, Williams 82) genotypes through genomic and cDNA cloning.

In total, 5 nucleotide differences (SNPs) were identified in all haplotypes, which resulted in to three amino acid changes i.e. V2I (4 bp, G/A), I41V (121 bp, A/G) and G202D (605 bp, G/A). However, one amino acid change at position 41st (isoleucine to valine) was only found in haplotype III, which was carried by resistant wild soybean S54 only.

3.4 Hairy root transformation of *HI4'OMT*/*I4'OMT* gene or *H1-H4* alleles

The iso-flavonoids are valuable secondary metabolites, which exhibit a range of biological activities and are predominantly produced in legumes via phenylpropanoid pathway (He and Dixon, 2000). In response to attack by pathogens, soybean also majorly accumulates iso-flavonoids and their conjugates (Murakami *et al*., 2014 and Cheng *et al*., 2015). Methylated iso-flavonoids such as formononetin show significant role against biotic stresses (for example bacteria and fungus; Pandey *et al*., 1997 and Liu *et al*., 2006) with medicinal effects (like anti-cancer (Kumar *et al*., 2011) and neuroprotective agent (Tian *et al*., 2013)). *HI4'OMT* (*hydroxyisoflavanone-4'-methyl-transferase*) or *I4'OMT* (*isoflavone-4'-methyl-transferase*) acts as a key-control enzyme(s), which catalyzes the methylation of 2, 7, 4'-trihydroxyisoflavanone, and daidzein-isoflavone substrates respectively in the synthesis of formononetin (4'-*O*-methyldadzien) using SAM (S-adenosyl-$_{L}$-methionine) as the methyl donor (Akashi *et al*., 2003 and Li *et al*., 2016). In soybean, *HI4'OMT*/*I4'OMT* is encoded by a small gene family containing three members (*Glyma.13g173300*, *Glyma.13g173600* and *Glyma.15g220700*; Chu *et al*., 2014). Though molecular cloning of *HI4'OMT*/*I4'OMT* gene(s) has been done in past among various legume species including Chinese licorice (*Glycyrrhiza echinata*; Akashi *et al*., 2003), birdsfoot trefoil (*Lotus japonicus*; Asamizu *et al*., 2000), barrel medic (*Medicago truncatula*; accession no. XP_013441548.2), and Asian arrowroot (*Pueraria lobata*, Li *et al*., 2016), however, the gene function against biotic stress specifically for disease resistance is still unclear. Therefore, to understand the function of *HI4'OMT*/*I4'OMT* gene and its regulatory mechanism is vital to exploring the genetic control of the iso-flavonoid pathway in pathogen resistance.

One of our previous study (gene expression) showed a member (*Glyma.13g173300*) of *HI4'OMT*/*I4'OMT* gene family significantly induced in S54, the resistant wild soybean genotype, after 8 days post inoculation (dpi) with SCN races 2 and 5. In contrast, its expression in S67, the susceptible genotype, was not significantly changed by SCN infection (Figure 3.3.1.1b). Using a molecular cloning (cDNA) approach, we detected a total number of four haplotypes (haplotype I; *H1* to haplotype IV; *H4*) of *Glyma.13g173300* between S54 and S67 genotypes. Haplotype I (*H1*) and haplotype II (*H2*) were detected in both genotypes (S54 and S67), however, haplotype III (*H3*) and haplotype IV (*H4*) were found specific to SCN resistant S54 genotype, and haplotype III (*H3*) was found with one amino acid change at 41st position [(isoleucine to valine; 121bp, A/G) (Figure 3.3.2.1)]. Based on this knowledge, we hypothesized that haplotype III (*H3*) of *hydroxyisoflavanone*/*isoflavone-4'-methyl-transferase* (*HI4'OMT*/*I4'OMT*) can confer broad spectrum SCN resistance by modulating the expression pattern of phenylpropanoid and iso-flavonoid pathway genes and related metabolites. Further to explore the role of identified other haplotypes including IV (*H4*) of S54 and common haplotypes (haplotype I; *H1* and II; *H2*) from S54 and S67 on non-resistant Williams 82 soybean cultivar we checked the detailed function of all alleles of *HI4'OMT*/*I4'OMT* gene using overexpression method via hairy root transformation.

Each allele of *HI4'OMT*/*I4'OMT* candidate gene was amplified using gene specific primers from S54 resistant wild soybean genotype and transferred into soybean 'Williams 82' cultivar by means of *Agrobacterium rhizogenes* K599 competent cells under the control of 35S promoter. The defense pattern of transgenic soybeans overexpressing *HI4'OMT*/*I4'OMT-H1-H4* gene under SCN stresses were examined using nematode

development analysis and female cysts counting. The possible mechanistic aspects of SCN resistance can be conferred by overexpression of *HI4'OMT*/*I4'OMT* which was determined by analyzing the expression of each alleles (*H1-H4*) of *HI4'OMT*/*I4'OMT* candidate gene.

3.4.1 Production of transgenic hairy roots overexpressing *HI4'OMT*/*I4'OMT-H1-H4* alleles

A transgenic hairy root system was used to assay whether overexpression of *HI4'OMT*/*I4'OMT-H1-H4* can confer resistance to SCN races 2 and 5 in soybean cultivar. To perform vigorous SCN bioassays, it is vital to nondestructively identify transgenic hairy roots overexpressing candidate gene *HI4'OMT*/*I4'OMT-H1-H4*. To this end, we chose to coexpress *HI4'OMT*/*I4'OMT-H1-H4* with a green fluorescent protein (*GFP*) marker gene [Baranski *et al*., 2006 (Annexure II; Figure 2.4.6.1a)]. We used two binary vectors one was having *GFP* marker gene only (obtained from Marker Gene Technologies, Inc, Eugene, OR, USA; Figure 3.4.1.1a) act as negative control and the other one was carrying two cassettes first overexpressing *HI4'OMT*/*I4'OMT-H1-H4* (expression vector) or *GUS* (positive control) candidate gene and the second *GFP* marker gene (constructed binary vector, section 2.4.4; Figure 3.4.1.1b). All vectors were used to produce transgenic hairy roots in cultivar 'Williams 82'. No apparent differences in the efficiency of hairy root generation was observed among controls and *HI4'OMT*/*I4'OMT-H1-H4* construct containing plants. Approximately 90% of the *A. rhizogenes*-infected plants produced transgenic hairy roots. The expression levels of *HI4'OMT*/*I4'OMT-H1-H4* in both types (control and with *HI4'OMT*/*I4'OMT-H1-H4* candidate genes) of transgenic hairy roots, which were initially selected by the presence of *GFP* signal, were determined using qRT-PCR. There was a significant difference in the relative expression of *HI4'OMT*/*I4'OMT-H3* allele between control and transgenic roots for both SCN races (SCN2, 4.53 times and

SCN5, 3.09 times respectively) was observed (Figure 3.4.1.1c-d). However, other alleles *H1*, *H2* and *H4* didn't show the significant differences as compare to the other control transgenic roots (*GFP* and *GUS*; Figure 3.4.1.1c-d).

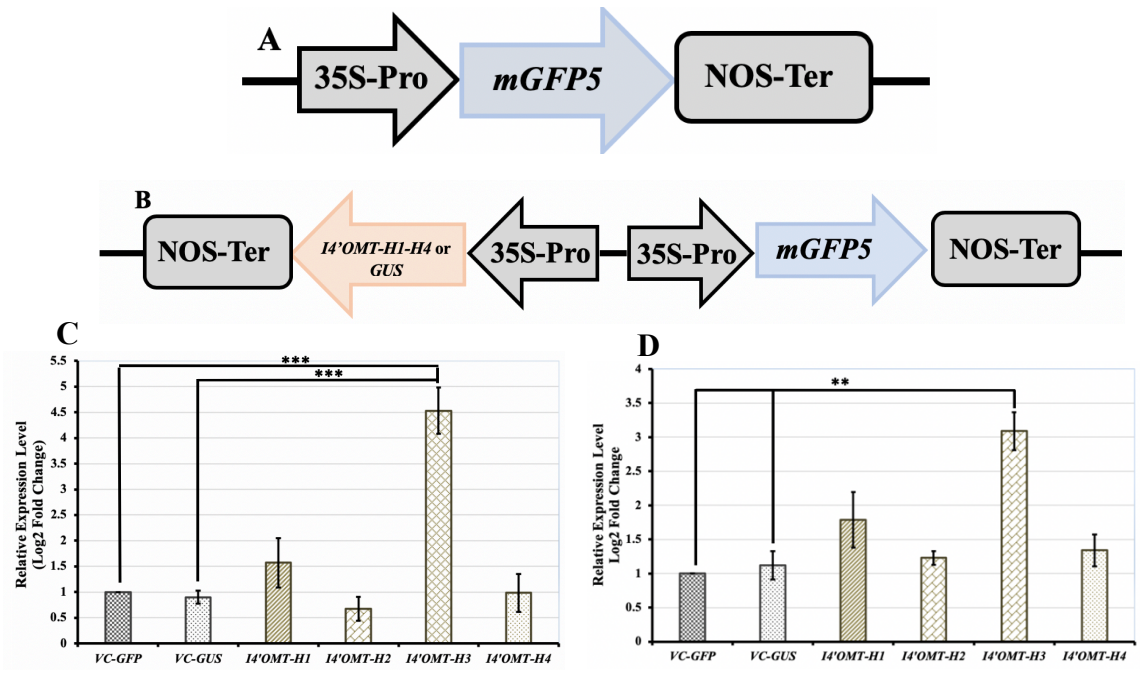


Figure 3.4.1.1: Production of transgenic hairy roots in Williams 82 overexpressing *HI4'OMT/I4'OMT-H1-H4*. (a) Schematic representation of the construct used for overexpression of a green fluorescent protein (*GFP*) reporter gene. '35S-Pro' and 'NOS-ter' represent the CaMV 35S promoter and the NOS terminator, respectively. **(b)** Schematic representation of the construct used for coexpression of a green fluorescent protein (*GFP*) reporter gene and *HI4'OMT/I4'OMT-H1-H4* or *GUS*. '35S-Pro' and 'NOS-ter' represent the CaMV 35S promoter and the NOS terminator, respectively. **(c-d)** Expression of *HI4'OMT/I4'OMT-H1-H4* in control and expression vector containing transgenic hairy roots from the 'Williams 82' genetic background in SCN2 and SCN5 races respectively. '*HI4'OMT/I4'OMT-H1-H4*' represents transgenic hairy roots containing *HI4'OMT/I4'OMT-H1-H4* transgene. '*VC*' represents transgenic hairy roots expressing

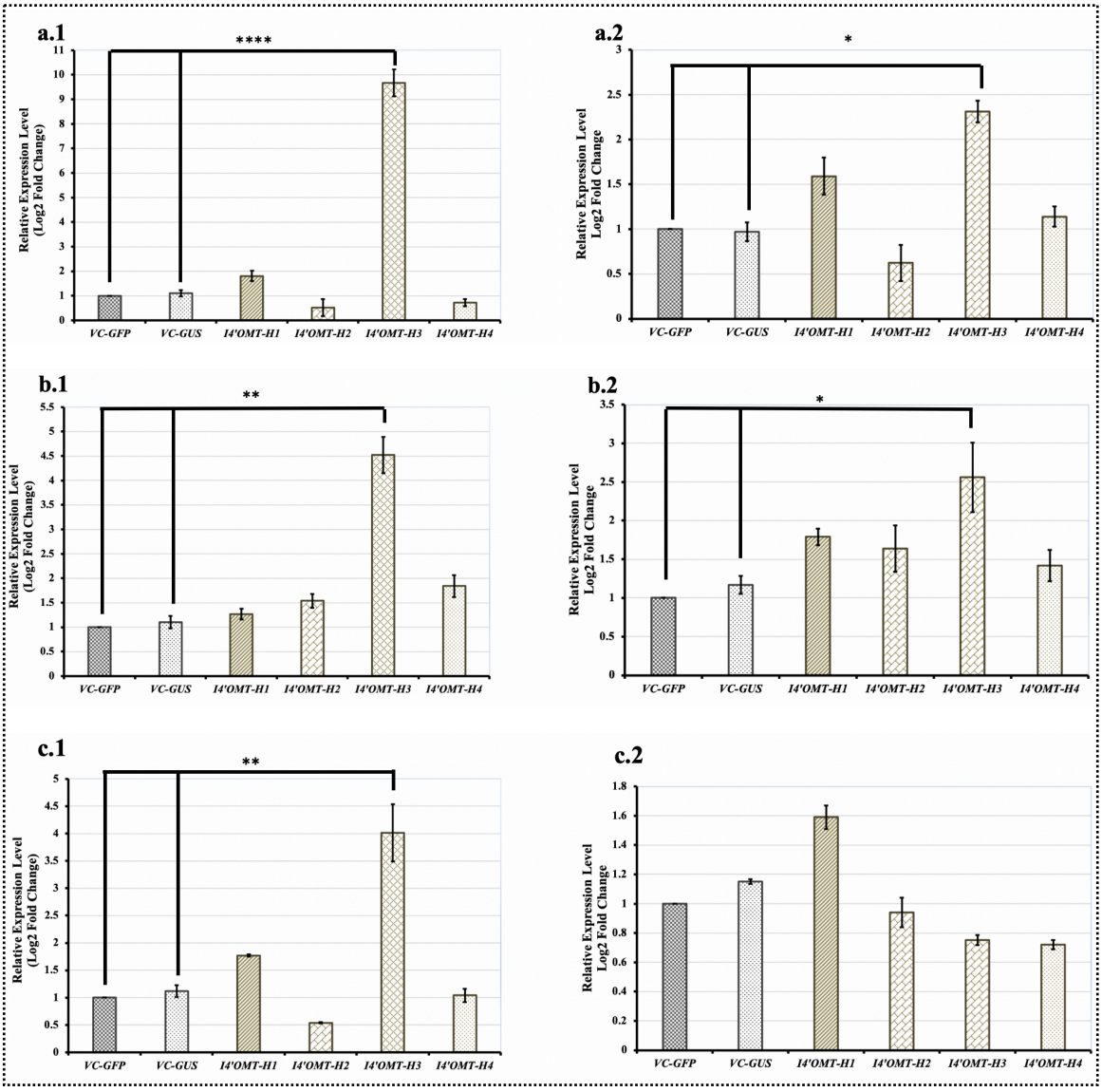
GFP gene only and *GUS* representing the transgenic hairy roots with *GUS* gene and *GFP* marker gene. qRT-PCR was performed with *HI4'OMT*/*I4'OMT* gene specific primers. Expression values were normalized to the expression levels of the soybean ubiquitin-1 gene (*GmUB1*) in respective samples. The level of *HI4'OMT*/*I4'OMT* expression in the '*VC*' (*GFP*) transgenic hairy roots of both races (SCN2 and SCN5) was arbitrarily set at 1.0. Each bar represents the mean relative expression level of five independent biological replicates with standard errors of mean. The bar with asterisks is significantly different (unpaired student's *t*-test $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$).

3.4.2 Expression of phenyl-propanoid and iso-flavonoid pathway genes in transgenic hairy roots overexpressing *HI4'OMT*/*I4'OMT-H1-H4* alleles

To understand the impact of *HI4'OMT*/*I4'OMT* alleles overexpression on other associated genes from phenyl-propanoid and iso-flavonoid pathway, we analyzed the expression of additional genes (*CHS7*, *CHS8*, *GmIFS1*, *GmIFS2* and *GmIFR*) in haplotype I to haplotype IV (*HI4'OMT*/*I4'OMT-H1-H4*) and compare its expression with *VCs* (*GFP* and *GUS*) containing transgenic hairy roots. Chalcone synthase (*CHS*) catalyzes the critical reaction for isoflavonoid and bioflavonoid synthesis [Figure 3.2.1; (Dhaubhadel *et al*., 2007)]. Also, there are two homologs of *CHS* family in soybean, *CHS7* (Glyma.01g228700) and *CHS8* (Glyma.11g110500) which have been previously identified for their significant role in defense by iso-flavonoid productions at biotic stress conditions (Yi *et al*., 2010 and Chennupati *et al*., 2012). When the roots with *HI4'OMT*/*I4'OMT-H1-H4* and with *GFP* (negative control) and *GUS* (positive control) genes were compared, both *CHS7* and *CHS8* showed significantly higher expression in *HI4'OMT*/*I4'OMT-H3* transgenic hairy roots than control and other alleles (*H1*, *H2* and *H4*) containing transgenic hairy roots among

both races. However, *CHS7* showed higher transcript abundance (9.67 times) as compare to the *CHS8* (4.52 times) in SCN2 root samples; while in SCN5, both *CHS7* and *CHS8* showed nearly similar expression [(*CHS7*, 2.31 times and *CHS8*, 2.56 times) (Figure 3.4.2.1a.1.2-b.1.2)].

2-hydroxyisoflavanone/*isoflavone synthase* (*IFS*) catalyze the key metabolic entry point for the formation of significant isoflavonoid(s) like daidzein (precursor of formononetin) and genistein [Figure 3.2.1; (Cheng *et al*., 2008)]. There are two homologs of *IFS* in soybean, *GmIFS1* (Glyma.07g202300) and *GmIFS2* (Glyma.13g173500). We quantified and compared the expression of both *IFS* genes in different transgenic roots. The expression of *GmIFS2* gene was induced significantly in *HI4'OMT*/*I4'OMT-H3* transgenic hairy roots than that the control and other alleles containing hairy roots among both races (SCN2, 4.51 times and SCN5, 2.15 times), whereas *GmIFS1* showed significant higher expression in the *HI4'OMT*/*I4'OMT-H3* containing root samples with SCN2 [(4.01 times) (Figure 3.4.2.1c.1.2-d.1.2)] only as compare to the SCN5 root samples. Isoflavone reductase (*IFR*) is a crucial enzyme involved in the biosynthesis of isoflavonoid phytoalexins [Figure 3.2.1; (Cheng *et al*., 2015)]. To determine whether overexpression of *HI4'OMT*/*I4'OMT-H3* has also affected the expression of phytoalexin synthesis genes, we also analyzed the expression of *isoflavone reductase*, *GmIFR* (Glyma.01g72600) gene in various transgenic hairy roots. Overexpression of *HI4'OMT*/*I4'OMT-H3* gene significantly enhances the transcript abundance of *GmIFR* gene in transgenic root with *H3* construct than that of control and other alleles in SCN2 (3.98 times) only. The expression of *GmIFR* was very less in *HI4'OMT*/*I4'OMT-H3* (0.723 times) in SCN5 infected root samples (Figure 3.4.2.1e.1.2).


a.1
****
Relative Expression Level (Log2 Fold Change)
VC-GFP
VC-GUS
I4'OMT-H1
I4'OMT-H2
I4'OMT-H3
I4'OMT-H4
a.2
*
Relative Expression Level Log2 Fold Change
b.1
**
b.2
*
c.1
**
c.2

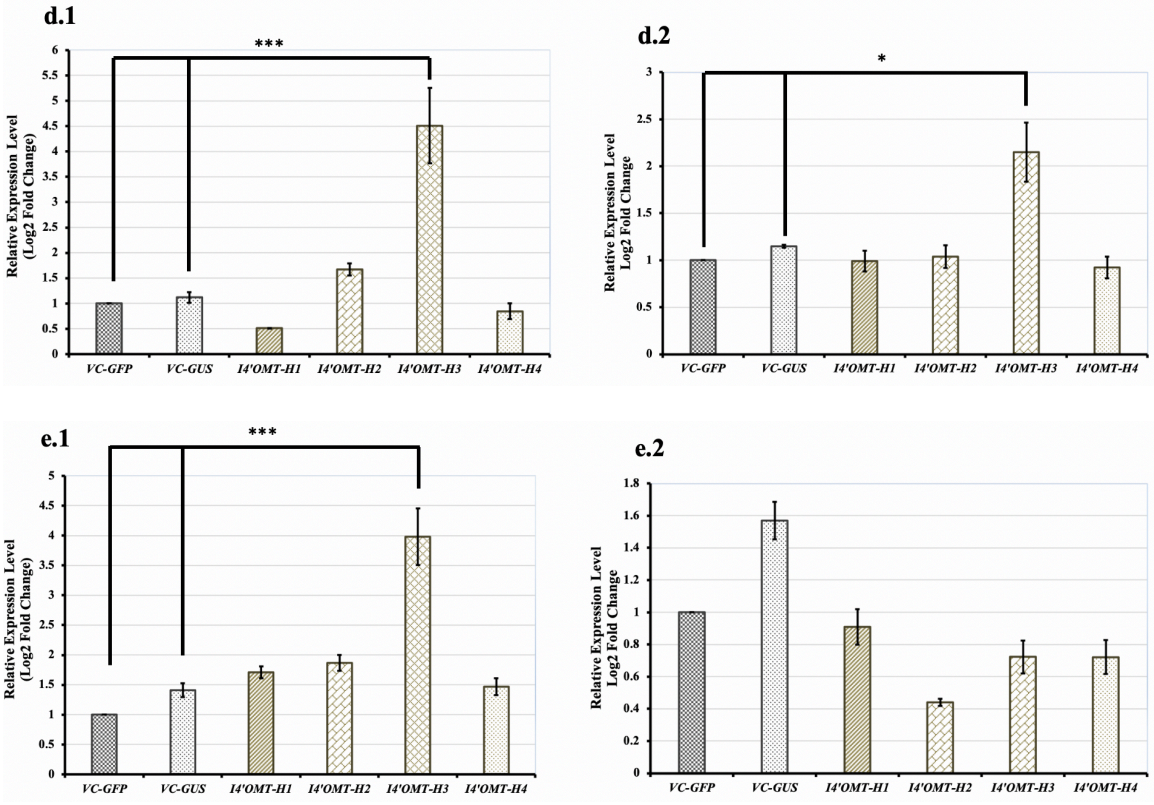


Figure 3.4.2.1: Relative transcript abundance of (a) *CHS7*, **(b)** *CHS8*, **(c)** *GmIFS1*, **(d)** *GmIFS2* and **(e)** *GmIFR* in transgenic hairy roots containing *HI4'OMT*/*I4'OMT-H1-H4* transgenes and green fluorescent protein reporter gene and *GUS* vector controls (*VCs*) produced in the 'Williams 82' background was analyzed. qRT-PCR was performed with *HI4'OMT*/*I4'OMT* specific primers. Expression values were normalized to the expression levels of the soybean ubiquitin-1 gene (*GmUB1*) in respective samples. The relative transcript abundance of *VC* (with GFP gene only) transgenic hairy roots without *HI4'OMT*/*I4'OMT-H1-H4* and *GUS* constructs was arbitrarily set at 1.0. Each bar represents the mean relative expression levels of five biological replicates with standard errors of mean. The bar with asterisks is significantly different (unpaired student's *t*-test $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$; $^{****}P \leq 0.0001$).

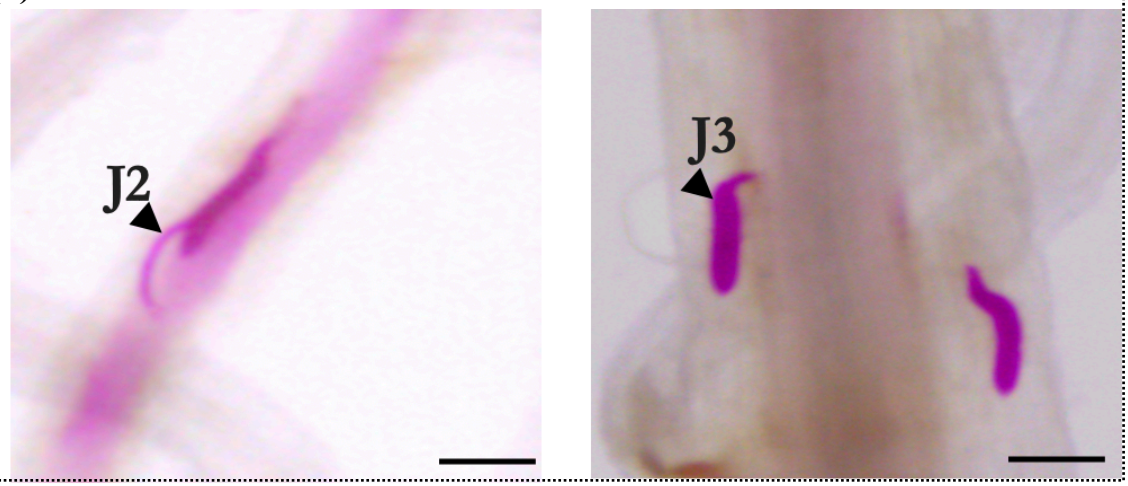
3.4.3 Evaluation of SCN development and female cysts on transgenic hairy roots overexpressing *H1-H4* candidate alleles from *HI4'OMT/I4'OMT* gene

We examined the effectiveness of *HI4'OMT/I4'OMT-H1-H4* alleles overexpression on broad spectrum SCN (SCN2 and SCN5) resistance using the transgenic hairy roots. Six types of transgenic hairy roots, which comprised overexpression of *HI4'OMT/I4'OMT* alleles (*H1* to *H4*) and vector controls (*VCs*) i.e. *GFP* (negative control) and *GUS* (positive control) gene in 'Williams 82', were used. Susceptible genotype 'Williams 82'- with vector controls (*GFP* and *GUS*), and vector with *HI4'OMT/I4'OMT-H1-H4* candidate genes, were subjected to infection with SCN race 2 and race 5. After 5 days of inoculation with second-stage juvenile (J2) nematodes, transgenic hairy roots were evaluated for SCN development following a recently established protocol by measuring the ratio of the nematodes developed beyond J2 stage to the total number of penetrating nematodes (Lin *et al*., 2013 and Melito *et al*., 2010). Nematodes either in J2 stage or beyond J2 (J3 stage) were observed for most sections of transgenic hairy roots, whereas J4 and adult stage was observed occasionally (Figure 3.4.3.1a).

The mean number of total nematodes in *HI4'OMT/I4'OMT-H1-H4* overexpressing and control transgenic hairy roots infected for SCN2 and SCN5 races was found in the range of 266 to 462 and 326.67 to 490.67 respectively. Similarly, the mean number of nematodes at J2 stage was ranged between the 138.33 to 191.33 for SCN2 and 134 to 212.33 for SCN5 and for beyond J2 stage (J3+J4+adult) ranged between 113.66 to 287.66 for SCN2 and SCN5, 130.33 to 278.33 respectively (Annexure II; Supplemental Table 3.4.3.1a-b). There was no significant difference found on the total number of nematodes ($P > 0.05$), and number of nematodes at J2 stage ($P > 0.05$). However, there was a

significant difference observed for the normalized ratio of the number of SCNs that had developed beyond J2 stage (J3+J4+adult) to the total number of nematodes among transgenic and control roots for both races ($P < 0.01$; $P < 0.001$).Therefore, the normalized ratio of the number of SCNs that had developed beyond J2 stage (J3+J4+adult) to the total number of nematodes per root was used as an index to indicate the resistance levels of the transgenic hairy roots harboring the *HI4'OMT*/*I4'OMT-H1-H4* and *VC* (*GFP* and *GUS*) constructs (Lin *et al*., 2013 and Melito *et al*., 2010). The resistance indices in transgenic hairy roots overexpressing *HI4'OMT*/*I4'OMT-H1-H4* were found in the range of 42.72% to 59.53% for SCN2 and 39.78% to 60.18% for SCN5 races relative to the *GFP* (SCN2, 62.56% and SCN5, 63.06%) and *GUS* (SCN2, 62.33% and SCN5, 61.54%) vector controls in Williams 82 (Figure 3.4.3.1b and c). However, in both races the lowest resistance index was observed in the transgenic hairy roots overexpressing the *HI4'OMT*/*I4'OMT-H3* (SCN2, 42.72% and SCN5, 39.78%) as compare to all other transgenic roots including *H1*, *H2*, *H4* alleles and both positive (*GUS*) and negative (*GFP*) controls (Annexure II; Supplemental Table 3.4.3.1a-b; Figure 3.4.3.1b and c).

(a)



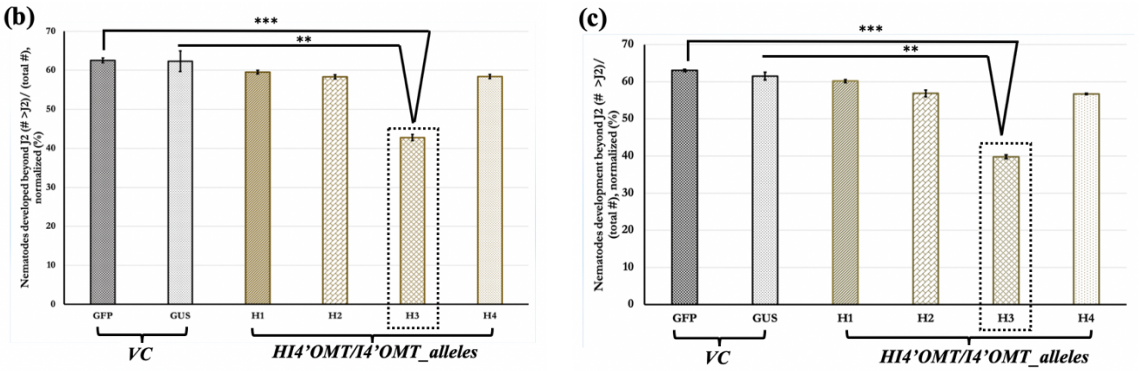
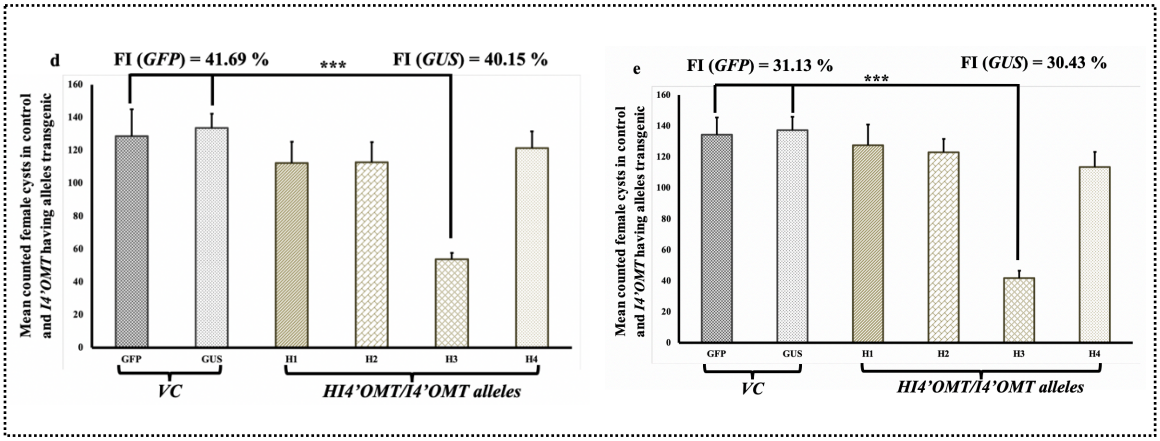


Figure 3.4.3.1: SCN performance in transgenic hairy roots. (a) Representative stages of SCN in the hairy roots. Bars = 250 μm. Six types of transgenic hairy roots from susceptible cultivar 'Williams 82' genetic background was compared and analyzed. '*HI4'OMT/I4'OMT-H1-H4*' represents transgenic hairy roots containing the *HI4'OMT/I4'OMT-H1-H4* candidate gene's alleles with '*GFP*' gene for coexpression. '*VC*' represents the transgenic hairy roots produced using a control vector containing '*GFP*' gene only and '*GUS*' with '*GFP*' for positive control. Each bar represents the **(b-c)** average of resistance index, normalized ratio of the nematodes developed beyond J2 (J3+J4+adult) stage to total number of nematodes per transgenic hairy root for SCN2 and SCN5 races respectively and, **(d-e)** the average of total number of cysts counted on independent transgenic hairy roots in SCN2 and SCN5 races with the standard error of means. Bars with asterisks are significantly different (unpaired student's *t*-test, $^{*}P \leq 0.05$ $^{**}P \leq 0.01$; $^{***}P \leq 0.001$).

SCN resistance between transgenic and control hairy roots was also analyzed by counting of total female cysts numbers. The numbers of cysts formed after ~25 days of second-stage juvenile (J2) inoculation for SCN2 and SCN5 races was counted in 10-12 independent hairy root lines including transgenic roots overexpressing the *HI4'OMT/I4'OMT-H1-H4*

(expression vector) and *GFP* and *GUS* (control vectors) genes. Total number of cysts counted within *HI4'OMT*/*I4'OMT-H1-H4* overexpressing and control transgenic hairy roots was found in the range of 644 to 1,604 and 502 to 1,650 for SCN2 and SCN5 races respectively. The lowest number of cysts were scored in *H3* allele containing transgenic roots 35 (H3-R-11) and 21 (H3-R-35) respectively for SCN2 and SCN5 races as compared to other transgenic roots including all other alleles (*H1*, *H2* and *H4*) and controls [(*GFP* and *GUS*) (Annexure II; Supplemental Table 3.4.3.1c-d)]. Mean numbers of cysts developed on transgenic roots overexpressing the *HI4'OMT*/*I4'OMT-H3* candidate gene was found statistically significantly different (SCN2, ∑ = 53.67; for *GFP*, *P* = 0.000143 and for GUS, *P* = 2.295E-08; SCN5, ∑ = 41.84; for GFP, *P* = 1.0609E-07 and for *GUS*, *P* = 1.45E-09) as compare to both the controls i.e. *GFP* (SCN2, ∑ = 128.75 and SCN5, ∑ = 134.42) and *GUS* (SCN2, ∑ = 133.67 and SCN5, ∑ = 137.50) among both races. Also, when mean female cysts number of *HI4'OMT*/*I4'OMT-H3* overexpressing transgenic roots was normalized with *GFP* and *GUS* control female cysts mean (female index, FI) it exhibited 41.69 % and 40.15 % resistance for race 2 (HG type 1.2.5.7) and 31.13 % and 30.43 % for race 5 (HG type 2.5.7) respectively (Figure3.4.3.1d-e).



CHAPTER 4: DISCUSSION

4.1 Untargeted Metabolomics for Broad-Spectrum SCN Resistance

4.1.1 Diverse defense mechanism(s) in response to various environmental stresses

Environmental stresses such as biotic and abiotic stresses are the major threats for any plant species in respect to their growth, development or to disturb their metabolic homeostasis, which also demand in plants the activation of multiple genes to produce different proteins, and primary and secondary metabolites (Shulaev *et al*., 2008). This further leads to onset of several metabolic response(s) mainly involved in signaling, physiological regulation, and defense reactions (Fraire-Velazquez and Balderas-Hernandez, 2013). Various metabolic mechanism(s) can be triggered in response to different stress stimuli, for instances, osmotic adjustment (via accumulation of amino acids, carbohydrates, polyols including others) for water deficit condition, regulatory frost condensation by increasing the concentration of soluble sugars, sugar alcohols and nitrogen containing compounds under cold stress to protect cell membranes (Rodziewicz *et al*., 2014), activation of ion-homeostasis mechanism in salinity stress, that includes transportation and storage of absorbed salt either in the vacuole or older tissues to neutralize the cell pH (Gupta and Huang, 2014) and development of systemic or induced acquired resistance via inter- or intra- priming and/or production of plant based toxins upon pathogen(s) or herbivores attack (Tenenboim and Brotman, 2016; Figure 3.1.1). However, understanding on which type of metabolite(s) [primary vs. secondary] and how they produced [constitutive (metabolites present all the time) vs. induced (accumulation of metabolites under specific stress conditions)] is also crucial to further explore the plant's defense strategies towards various stresses.

A number of abiotic stresses are found responsible for several changes in plant metabolism including restricted activities of metabolic enzymes, substrate shortage, excess requirement of specific compound(s), or/and combination of two or more factors (Obata and Fernie, 2012). A number of phenotype-metabolite interaction studies have been carried out in recent years to find the particular pattern(s) linked with various abiotic stresses and mainly highlighted the fine adjustment of amino acids, carbohydrates, and Krebs cycle intermediates [(primary metabolism) (Arbona *et al*., 2013 and Fraire-Velazquez and Balderas-Hernandez, 2013)]. One of the drought stress tolerance study showed the constitutive response by maintaining high relative water content (RWC), ascorbic acid content and membrane stability and decreasing the concentration of hydrogen peroxide (H_2O_2) with less lipid peroxidation in wheat cultivars (Sairam and Srivastava, 2001). Another drought study in wheat cultivars leaf tissues though demonstrated the accumulation of amino acids specifically proline, tryptophan, and the branched amino acids including leucine, isoleucine, and valine [(induced defense) (Bowne *et al*., 2012)]. Adaptation of alkali-salt tolerance among wild soybean and bluegrass (*Poa pratensis*) was mainly found associated with the accumulation of amino acids (proline, valine, glutamic acid, asparagine, glutamine, phenylalanine, lysine), sugars (glucose, sucrose, trehalose), organic acids (palmitic acid, lignoceric acid, α-ketoglutaric acid, gluconate, galactarate, glucarate) and those involved in the citric acid cycle [(citric acid, malic acid, isocitric acid, succinic acid, maleiate, aconitate) (Zhang *et al*., 2016 and Hu *et al*., 2015)]. These examples show the significant role of both induced and constitutive mechanisms with primary metabolism in various abiotic stresses. In one of our current study, we have also demonstrated the constitutive expression of organic acids, citric acid cycle intermediate

metabolites and sugars among glycophyte and halophyte accessions of wild sweet potato for salinity stress (unpublished research data).

On the other hand, plant biotic interactions are mainly accomplished by changes in secondary metabolism i.e. the production of wide range of secondary metabolites including flavonoids, iso-flavonoids, phenolic acids, terpenoids or alkaloids mainly by induced mechanism (Wittstock and Gershenzon, 2002). One of the studies by Soriano et al. (2004) demonstrated the induction of O-methyl-apigenin-C-deoxyhexoside-O-hexoside (flavonoid conjugate) in oats (*Avena sativa*) upon the nematode (*Heterodera avena*) attack. Jones et al. (2007) also showed the production and accumulation of flavonoids at developing nematode feeding sites in *Arabidopsis thaliana* upon root knot nematode attack (*Heterodera schachtii*) as part of defense mechanism. Chemical interaction between *Medicago tuncatula* and Asian soybean rust (*Phakopsora pachyrhizi*) showed the inducement of iso-flavonoids (formononetin-7-O-glucoside), flavonoid (narigenin-7-O-glucoside, liquirtin, liquiritigenin, including others), phtyoalexin (medicarpin) and triterpenoid saponin glycosides (Ishiga *et al*., 2015). Beside legumes various other plant species also showed similar defense mechanism, as significant induction of phenylpropanoid, flavonoids (secondary metabolites) have been identified in the grape berries attacked by fungal plant pathogen [(*Botrytis cinerea*) (Hong *et al*., 2012)]. Similarly, insecticide exposure in *Bacillus thuringensis* (*Bt*) transgenic rice (*Oryza sativa*) led to stronger expression and regulation of signaling molecules (salicylic acid; phenolic acid), shikimate mediated secondary metabolism and antioxidants [(α-tocopherol and dehydroascorbate/ascorbate) (Zhou *et al*., 2012)]. In another recent study accumulation of the defense related flavanone hesperidin (secondary metabolite) has demonstrated in leaves

and roots of citrus (*Citrus sinesis* grafted on to *C. limonia*) affected by bacterial infection [(*Xylella fastidiosa*) (Soares *et al*., 2015)]. Most of the plant-pathogen based metabolic studies demonstrate the strong role of induced defense mechanism under diverse biotic stresses. Outcome of the current study is also consistent with the previous researches and found the accumulation of secondary metabolites (mainly iso-flavonoids, phenolic acids and their conjugates) in the roots of resistant genotype (S54-T) against SCN2 and SCN5 infection. This indicates that the inducement of secondary metabolites as prevalent mechanism for broad-spectrum SCN resistance (Figure 3.1.4.1a and b; Figure 3.1.3.1.1; Group-4 and Figure 3.1.3.2.1; Group-2) as compare to the constitutive defense mechanism (Figure 3.1.3.1.1; Group-2 and Figure 3.1.3.2.1; Group-1).

4.1.2 Resistant metabolites in response to SCN infection

Hypersensitive response including the formation of necrotic lesions, induction of pathogenesis-related (*PR*) genes, and the biosynthesis of secondary metabolites and signaling molecules towards pathogen attack is a major defense mechanism in biotic resistant plants (Durner *et al.,* 1997). A significant phenolic phytohormone known as salicylic acid ((SA) or 2-hydroxybenzoic acid) is well known for their role in plant defense (Zhang *et al.,* 2013; Yang *et al.,* 2014 and Shine *et al.,* 2016). Elevation of salicylic acid is found linked accumulation of reactive oxygen species (H_2O_2, $O_2^{\cdot-}$) and induction of *PR* genes for local and systemic acquired resistance. It can also act as an antioxidant to kill the pathogen invading cells directly via autolysis or by inducing the formation of other defensive metabolites in neighboring cells (Durner *et al.,* 1997). However, SA is also found negative regulator of plant fitness. Higher accumulation of free SA can result in stunted growth, reduced photosynthesis, early leaf senescence, inhibited seed germination

including others (Rivas-San Vicente and Plasencia, 2011). This phytohormone is also demonstrated strong competitive inhibitor of *PBS3* gene, which encodes an acyl-adenylate/thioester forming (*GH3*) enzyme that catalyzes the amino acid conjugation of phenolic acid(s), metabolites which were previously found involved in bacterial and oomycete pathogens resistance (Okrent *et al*., 2009). Methylation, glycosylation or amino-acid conjugation of bioactive free SA to its derivatives/conjugates has been so far found the effective mechanisms to reduce the phytotoxicity and other adverse effects of SA. For instances, in one of the Arabidopsis study by Bartsch et al. (2010) showed the accumulation of xylose-conjugated form of SA (2, 3-DHBX; 2-hydroxy-3-β-*O*-D-xylopyranosyloxy benzoic acid) when stressed by *Pseudomonas syringae*. Few other studies indicated the isomerization as a prevalent mechanism to reduce the negative impacts of SA and for the potential usage of its positional isomer(s) for biotic resistance. Like, 4-hydorxybenzoic acid (4-HBA; one of SA positional isomer) accumulation has been detected in phloem fluids of cucumber leaves when inoculated with *Pseudomonas syringae* pv. *syringae* (Smith-Becker *et al*., 1998). Also, further hydroxylation (addition of –OH group) of SA isomer (e.g. 4-hydroxybenzoic acid; 4-HBA) at C-3 and C-5 (3^{rd} and 5^{th} carbon positions) can derivatized 4-HBA into 3, 4, 5-trihydroxybenzoic acid (3, 4, 5-THBA; gallic acid), which were purified from *Terminalia nigrovenulosa* bark to test their nematicidal activity against root knot nematode i.e. *Meloidogyne incognita* (Nguyen *et al*., 2013 and Nguyen *et al*., 2013). Experimental data showed that 3, 4, 5-THBA can significantly inhibit the juveniles hatching and can destruct the shape of eggs, when used at higher concentration (1.0 mg/ml). However, free forms of phenolic acids (SA (2-HBA), and 4-HBA) sometimes can be access by pathogens for their better survival in the host. For instances, a functional

4-hydroxybenzoic acid degradation pathway found in the phytopathogen *Xanthomonas campestris* a causal agent of crucifer's black rot. Bioinformatics, genetic and biochemical data demonstrated that *X. campestris* use 4-HBA as a substrate and convert it into 3, 4-DHBA after hydroxylation. Further they encode a new form of β-ketoadipate succinyl-coenzyme- A, a transferse enzyme to degrade the 3, 4-DHBA into simpler forms that can be use in TCA cycle as their food source and for their development, growth and to express the pathogenicity (Wang *et al*., 2015). Conversely, degradation of phenolic acid glucoside and methylated conjugates has been rarely seen by microbes for their growth and development. This indicates the more stability of phenolic acid conjugates and significance in biotic resistance then their free forms.

Therefore, detected reduced levels of salicylic acid (SA) and higher accumulation of its glycosylated (1-O-4-hydroxybenzoyl-β-D-glucose ester, 1-O-3, 4-dihydroxybenzoyl-β-D-glucose ester, 1-O-3, 4, 5-trihydroxybenzoyl-β-D-glucose ester), methylated (1-O- 4-hydroxy-3-methoxy-benzoyl-β-D-glucose ester) and aldehyde (4-hydroxybenzaldehyde) conjugated forms, from the present study can also be linked to the above explanation. These metabolites might help in controlling the phytotoxicity of free phenolic acids at one side but intensifying their effects against SCN pathogen by converting them into their conjugated forms on the other side (Figure 3.1.3.1.1; Group-4 and Figure 3.1.3.2.1; Group-2). This also suggests that glycosylation and methylation mechanisms are perhaps vital for the stabilization of phenolic acids in SCN infected resistant roots for defense (Figure 4.1.2.1).

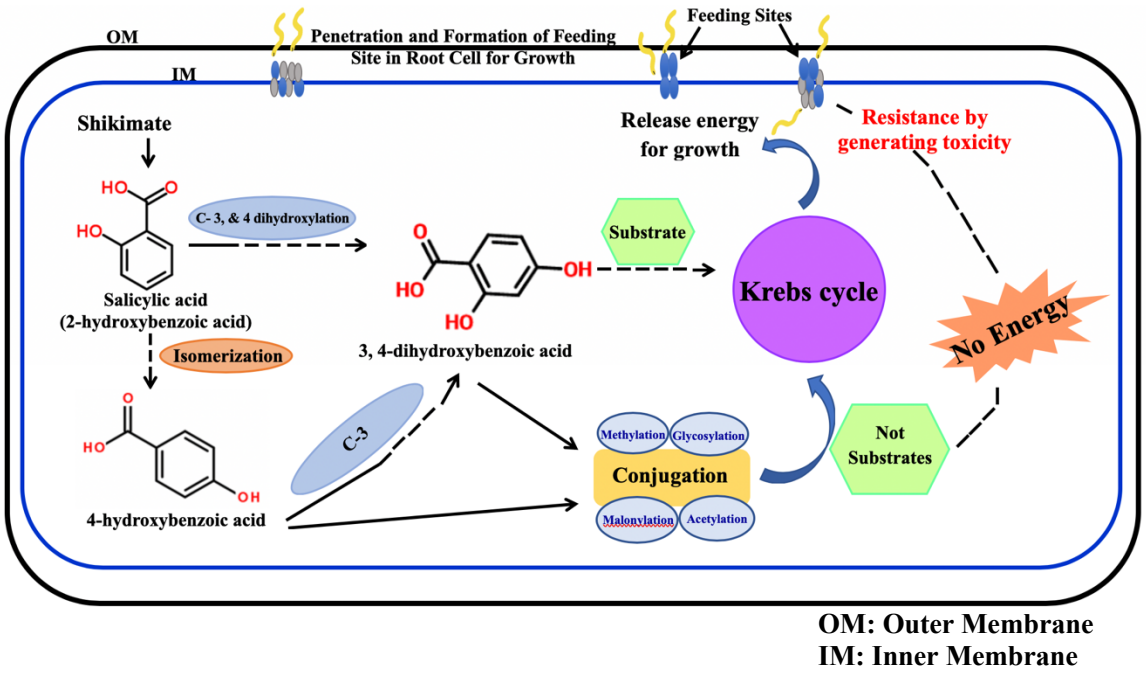


Figure 4.1.2.1: Proposed model for broad-spectrum resistance mechanism in resistant (S54) wild soybean genotype. During resistance response conjugates might directly release at feeding sites without being used as substrate in Kreb cycle for energy production. After their release at feeding site they might convert back into the free phenolic acid forms which can be toxic to the growing nematodes.

Another reason of higher accumulation of conjugates can be explained by the presence of more N-benozyl-L-glutamic acid (amino acid conjugate of phenolic acid, Figure 3.1.3.1.1; Group-4 and Figure 3.1.3.2.1; Group-2) in SCN resistant infected roots, as amino acid conjugation requires more expression of *PBS3* gene but less expression of its negative regulator (i.e. salicylic acid). Glycosylation and methylation are also required for effective cell permeability and volatility, subsequently conjugates formation can also enhance the attaining of systemic resistance via intra-priming (within plant between organs) and inter-priming [(between plants) (McQualter *et al*., 2005)]. Endogenous uridine diphosphate

(UDP)-glucosyltransferases actively converts the phenolic acids into their glucoside or glucose ester forms. Both conjugated forms can transport from the cytosol to vacuole within cell or between cells as a deactivated storage form and can get activated back during pathogen attack (McQualter *et al*., 2005). Therefore, high accumulation of glucose ester forms found in present research gives a hint that probably these compounds can also be linked with intra-priming. Methylation is also increasing the cell permeability and volatility, thus allowing more effective long-distance transport of defense signals to other remote plants and helping them to prepare for any future attacks (Zhang *et al*., 2013 and Tenenboim and Brotman, 2016). There have been reports that showed conversion of salicylic acid into methyl salicylate (MeSA) is critical to acquire the systemic resistance between plants (Park *et al*., 2007). Higher accumulation of methylated conjugated form of 3, 4-DHBA (1-O- 4-hydroxy-3-methoxy-benzoyl-β-D-glucose ester, Figure 3.1.3.1.1; Group-4 and Figure 3.1.3.2.1; Group-2) also indicates this compound can be associated with systemic resistance via inter-priming between infected and non-infected remote plants. Alongside species unspecific phenolic acid resistance molecules, legume specific iso-flavonoids were the other assorted group of plant natural metabolites, which were found critical for SCN resistance (Figure 3.1.3.1.1; Group-4 and Figure 3.1.3.2.1; Group-2 respectively; Figure 3.1.4.1a and b). These metabolites are not found involved in the plants primarily physiological mechanisms like photosynthesis, respiration, transpiration, differentiation and solute transport or translocation (Mazid *et al.,* 2011) but mainly require for the survival of an organism in a particular environment or stressful conditions. Induced concentration of iso-flavonoids such as the daidzein, genistein and their glucoside (daidzin), methylated (formononetin, iso-formononetin) and malonyl conjugates (malonyl-

daidzin) had been previously detected upon bacterial, fungal and herbivores attack (Morris *et al.,* 1999 and Murakami *et al.,* 2014).

Current research findings also showed induced level of iso-flavonoids (daidzein, daidzin, malonyl-daidzin, genistein, formononetin, iso-formononetin; Figure 3.1.4.1a and b) upon SCN2 and SCN5 infection (expression pattern differences in SCN-resistant and SCN-susceptible wild soybean genotypes; Figure 3.1.3.1.1; Group-4 and Figure 3.1.3.2.1; Group-2). Based on our so far knowledge this is the first established report that shows the role of phenolic acids and iso-flavonoids conjugates to SCN resistance in wild soybean. Conversely, these metabolites have been extensively studied for their health promoting effects and nutritive properties. For instances, daidzein and geinstein has been found linked to the suppression of alcohol consumption, and prevention or inhibition of mammary, prostate and stomach tumor growth respectively (Dixon and Steele, 1999). Therefore, the enhanced level of iso-flavonoids/phenolic acids and their accumulation in plants can be found useful for both pathogens mediated defense and to enhance nutritional values for human health.

4.1.3 Basal resistance versus pathogen specific resistance: adjusted incorporation as required

Plants resistance can either occur by using mutual metabolites or by means of producing pathogen-specific metabolites upon microbial attack. Metabolites, which express among resistant and non-resistant genotypes, can provide basal resistance for various ranges of pathogens however, phytochemicals which produce and get instigated upon intrusion of specific strain of microbe in resistant genotype essentially considered for pathogen specific defense (Morris *et al.,* 2006 and Kant *et al.,* 2015). Alternately, the distribution and

expenses of resources to synthesize metabolites for both basal and specific resistance can be expensive. Therefore, according to essentiality and demand, plants need to embrace one defense strategy and trade-off other in course of sparing energy, survival cost and to effectively diminish the pathogenicity. The basal resistance phytochemicals primarily known as "phytoanticipins" mainly act as a first layer of defense and release as an antimicrobial and antifungal compound upon pathogen attack (Jones *et al.,* 2007; Gonzalez-Lamothe *et al*., 2009 and Balmer *et al.,* 2013). A various group of phytoanticipins is made out of secondary metabolites including saponins and flavonoids or iso-flavonoids. As Hummelova et al. (2015) demonstrated the anti-bacterial activity of demethyltexasin iso-flavonoid against various strains of *Staphylococcus aureus*. Moreover, flavonoids-C-glucosides have been distinguished to have anti-nematode activity. Like because of *Heterodera avena* attack in oats (*Avena sativa*), induction of O-methyl-apigenin-C-deoxyhexoside-O-hexoside occur, eventually which turned to be an effective protectant against nematode infection (Soriano *et al.*, 2004). Additionally, apigenin-C-glucoside also represents one of the dominant flavonoids group for having anti-cancerous, anti-inflammatory, antioxidant and anti-microbial properties, and might include in therapeutics to treat various type of cancers, hepatitis, inflammation and several kinds of infections (Martens and Mithofer, 2005). Another flavonoid namely chrysin secreted by legumes (e.g. *Medicago sativa*), at root tips and emerging root hair zone act as a specific symbiotic signal for *Rhizobium* bacteria.

It also possesses antioxidant and anti-inflammatory biological activities (Ahad *et al.*, 2014 and Samarghandian *et al.*, 2015). Similarly, saponins are found to be involved in a variety of biological activities including allelopathy (secretion of chemical compounds

against competing species), beneficial anti-fungal properties (Huhman and Sumner, 2002), and deterrence to insect pest (Gholami *et al.*, 2014). Like in alfalfa (*Medicago sativa*) species soysaponin I was released as insecticidal upon attack of *Spodoptera littoralis* (Agrell *et al.*, 2003). Current study likewise demonstrates the presence of various flavonoids (sojagol) and iso-flavonoids (genistein and glycitein), and their conjugates like 8C-hexosylchrysin, apigeninidin 5-O-glucoside, 2'-hydroxypseudobatigenin, and soysaponins (II and V) in both genotypes (resistant and susceptible) infected roots with about same concentration or little less concentration at later (Figure 3.1.3.1.1; Group-2 and Figure 3.1.3.2.1; Group-1) days of SCN infections. Here, we can foresee that flavonoid derivatives (8C-hexosylchrysin, apigeninidin 5-O-glucoside), iso-flavonoids (2'-hydroxypseudobatigenin, glycietin, genistein) and soysaponins probably can provide basal resistance to both genotypes against SCN pathogen, however effective defense may need to reduce their production to save energy, survival cost and to triggered the production of SCN specific phytoalexins (like phenolic acid and iso-flavonoids, and their glycosylated and methylated conjugates in resistant genotype specifically at later days of infection (5 or 8 dpi; Figure 3.1.3.1.1; Group-2 and Figure 3.1.3.2.1; Group-1).

4.1.4 Metabolites diversity in response to various biological challenges

Defense mechanism is inevitable for plant existence but effective resistance capability against particular pathogen or pest can only resist at the sacrifice of several biological and ecological value having metabolites (Huot *et al.,* 2014). Despite the fact that an extensive variety of secondary metabolites were identified in the infected roots of both genotypes (Figure 3.1.3.1.1; Group-2 and Figure 3.1.3.2.1; Group-1) yet key compounds linked to SCN resistance were more accumulated only in resistant infected (S54-T) roots (Figure

3.1.3.1.1; Group-4 and Figure 3.1.3.2.1; Group-2), which demonstrates the requirement of specific metabolites to provide effective protection against SCN attack. However, susceptible (S67) genotype also accumulated some other biologically relevant metabolites (including chalcones, flavanones, flavonoids and iso-flavonoids; Figure 3.1.3.1.1; Group-1 and Group-3 and Figure 3.1.3.2.1; Group-3), which might require for their better survival and local environmental adaptation. Chalcones, flavanones and flavonoids are widely distributed group of secondary metabolites *in planta* having various biological and ecological functions including: flavoring agents, health promoting nutraceuticals, protection against abiotic and biotic stresses, endogenous compounds for growth and development, root nodulation, regulation of local auxin transport and providing coloration to flowers, seeds and fruits to attract pollinators (Dixon and Steele, 1999; Ferreyra *et al.*, 2012 and Su *et al.*, 2015).

Iso-flavonoids have previously been detected as important class of metabolites with high nutritional values and medicinal benefits to humans (e.g., having anti-cancerous and antioxidant properties). Also, these compounds provide resistance to insect pests, herbivores and against ultraviolet radiation, such as ozone stress tolerance in various plant species (Ferreyra *et al.*, 2012 and Lin *et al.,* 2014). Iso-flavonoid(s) like glycitin (glycitein-7-O-glucoside), glycitein 6”-O-acetylglucoside, genistin-6”-O-acetate, 2’-hydroxyformononetin and 2’-hydroxydaidzein cause estrogenic and anticoagulant effects, and found toxic for herbivores (Dixon and Paiva, 1995). Iso-flavonoid i.e. genistin (genistein-7-O-glucoside) act as a phytoalexin in legume (*Phaseolus vulgaris*) against fungal pathogen (*Fusarium solani* f. sp. *phaseoli*, and *Monilinia fructicola*) and copper chloride chemical compound (Reichling, 2010). Most of the above mentioned iso-

flavonoids were found highly accumulated in roots of susceptible (S67) genotype either at early days of SCN infection (3 dpi; Figure 3.1.3.2.1; Group-3), or at all time points (Figure 3.1.3.2.1; Group-3) when infected with SCN2 and SCN5 respectively. This indicates that probably these metabolites were induced by susceptible genotype in response to SCN attack however, might be these compounds wouldn't be able to provide them effective resistance due to their diverse functionality in tolerance for other stresses or having significant role in other physiological mechanisms. Therefore, the higher levels of identified chalcones, flavonoids and iso-flavonoids in infected roots of susceptible (S67) genotype from the current study indicates their involvement either in the enhancement of nutritive values for the genotype or providing them resistance from other environmental stresses.

4.2 Phenyl-propanoid or Iso-flavonoid Pathway's Genes Involvement in Broad-Spectrum SCN Resistance

The aim of the quantitative estimation of phenyl-propanoid or iso-flavonoid pathways' genes was to provide insights on how SCN infection can affect their expression in wild soybean roots. Though induction of phenyl-propanoid genes such as *PAL*, *CHS*, *C4H*, *IFS* and *IFR* has been previously detected in soybeans and other plant systems upon bacterial and fungal infections (Zhang *et al*., 2017; Cheng *et al*., 2015; Jadhav *et al*., 2013 and Chennupati *et al*., 2012), however very little has been reported on gene activities of enzymes in the pathway leading to the synthesis of phenolics and isoflavones in response to *H. glycines*. By keeping this in view, we have characterized the expression of genes encoding enzymes at key flux control points (i.e. *PAL* and *CHS*) and at a branch point in

isoflavones biosynthesis (i.e. *IFS*, *IFR*, *CYP93A1* and *HI4'OMT*/*I4'OMT*) in wild soybeans following *H. glycine* (SCN2 and SCN5) challenge.

The findings of current study represent the first report that genes encoding key control and branch point enzymes in the phenyl-propanoid pathway leading to isoflavones and phenolic acid biosynthesis are affected by *H. glycine* infection. The differential expression of *PAL* and *HI4'OMT*/*I4'OMT* genes among SCN resistant (S54) and susceptible (S67) wild soybean genotypes has significantly shed light on our understanding of the complex mechanisms of SCN-plant interactions. Our metabolomics data revealed the involvement of phenolic acid and isoflavone derivatives in the broad-spectrum SCN resistance and *phenylalanine ammonia lyase* (*PAL*) is a key checkpoint enzyme in the production of the substrates required for derivatization of phenolic acid and isoflavones. Although no or very less transient change in *PAL1.1* and *PAL1.2* expression was determined at 8 dpi between S54 and S67, however their strong induction in S54 roots during the early stages of infection (3 and 5 dpi) suggests that, S54 may have a more efficient metabolic pathway than S67 for high production of isoflavone and phenolic acid for immediate response to SCN, and their availability for conversion into methylated, glycosylated, hydroxylated forms at the point of SCN intensification (8 dpi) for robust nematicidal activity (Spoel and Dong, 2012 and Fu and Dong, 2013). Our results are also in agreement of a recent study which shows as compare to ICSs (*isochorismate synthase*); *PALs* have important functions in the metabolism of phenolic acids in soybean (Shine *et al*., 2016).

The upregulation of the enzymes catalysing metabolites modification may lead to the accumulation of compound derivatives/conjugates, which can act as phytoalexins

(Jadhav *et al*., 2013). Significantly higher accumulation of formononetin (methylated derivative of daidzein isoflavone) at later days (8 dpi) of SCN infection has been detected in our metabolomics study. These results are consistent with the present analysis, as that, the production of formononetin is mainly accomplished by the methylation of 2, 7, 4'-trihydroxy-isoflavanone or daidzein substrates. These reaction are largely catalysed by either h*ydroxyisoflavanone-4'-methyl-transferse* (*HI4'OMT*) or *isoflavone-4'-methyl-transferase* (*I4'OMT*) enzyme, whose corresponding gene's (*HI4'OMT*/*I4'OMT*; Glyma.13g173300) higher expression has been determined in the roots of S54 as compare to S67 at 8 dpi (Figure 3.2.2.1c). The distinct expression pattern of isoflavone-methylation gene (*HI4'OMT*/*I4'OMT*) between S54 and S67 suggest that this gene discriminatorily activated more synthesis of related enzyme, thus leading to the higher accumulation of formononetin in resistant (S54) genotype. Formononetin has been found to act as anti-microbial compound (Vishnuvathan *et al*., 2016) and nematicides against soil-borne pathogens and root-feeding-insects (Chin *et al*., 2018). Therefore, its abundance can increase the degree of SCN resistance in roots of S54 specifically at prolonged duration of infection. This is consistent with the results of our previous study, which shows that, the SCN development in S54 roots at 8 dpi was much more inhibited than that at 3 and 5 dpi (Zhang *et al*., 2017).

Higher expression of *PALs* and *HI4'OMT*/*I4'OMT* genes against *H. glycines* opens avenues to use them as relevant candidates for soybean transformation in SCN susceptible genotypes. The successful gene transformation of *PALs* and *HI4'OMT*/*I4'OMT* can produce natural phytoalexins to mitigate the impact of nematode damage on soybean. For the present research we choose *HI4'OMT*/*I4'OMT* as a potential candidate to understand

their role in SCN resistance. The selection of gene has been accomplished based on hypothesis that the up regulation of *HI4'OMT*/*I4'OMT* gene can be important in broad-spectrum SCN resistance by enhancing the biosynthesis of significant metabolites such as formononetin and other isoflavone derivatives. Molecular cloning (genomic and cDNA cloning) and hairy root transformation of *HI4'OMT*/*I4'OMT* gene has been further carried to understand their role in SCN defense. The detailed description of molecular cloning and hairy root transformation of *HI4'OMT*/*I4'OMT* gene and their impact on SCN resistance has been provided in the chapters#2 and 3 respectively.

4.3 Role of *Hydroxyisoflavanone*/*Isoflavone-4'-methyl-transferase* (*HI4'OMT*/*I4'OMT*) Gene in Broad-Spectrum SCN Resistance

4.3.1 Selection of *H3* candidate allele from various alleles of *HI4'OMT*/*I4'OMT* gene

Broad spectrum resistance and durability from various plant diseases can be accomplished efficiently through non-race specific resistance as compare to the race specific resistance (Li *et al*., 2017). Through metabolomics and genomics analysis, we have identified the wild soybean specific (S54 specific, resistant to SCN2 and SCN5 races) haplotype (haplotype III; *H3*) of *hydroxyisoflavanone*/*isoflavone-4'-methyl-transferase* (*HI4'OMT*/*I4'OMT*) gene. A comparison of amino acid sequences of *HI4'OMT*/*I4'OMT* gene among currently cloned and prior available genotypes of soybean/wild soybean also reveals that the haplotype III (*H3*) allele exists rarely in wild soybeans (*G.soja*; S54-H3, cloned from the current study and *G. soja*; E2-H3, NCBI accession no. AIE16653; Cheng *et al*., 2008, Annexure II; Supplemental Figure 4.3.1). Haplotype III (*HI4'OMT*/*I4'OMT*-*H3*) carries three single nucleotide polymorphisms including one for amino acid variation

[(isoleucine to valine, 41st amino acid; 121 bp, A/G) (Figure 3.3.2.1)]. Despite the fact that isoleucine and valine both contains non-polar side chains (R-functional group) and belong to the class of branched chain amino acids, nonetheless, their substitution with one another can impact the protein functionality. One recent example of single amino acid (isoleucine to valine) swapping affecting function of PYL proteins for substrate (pyrabactin, an ABA agonist) binding has been reported in *Arabidopsis thaliana* (Yuan *et al*., 2010). PYL proteins act as abscisic acid (ABA) receptors in *A. thaliana*, which binds to the ABA agonist pyrabactin. This protein contains 14 members, with high degree of sequence similarity in the species. However, two members PYL1 and PYL2 showed single amino acid differences at the positions 137 and 114 respectively, that leads to their diverse binding capability to pyrabactin. In PYL1, 137th amino acid has been substituted by valine from isoleucine, and in PYL2 it was vice versa at 114th amino acid position. Due to change of a single amino acid in PYL1 (I137V), it was predominantly selected by pyrabactin, which further leads to highly effective ABA signaling as compare to PYL2 (Yuan *et al*., 2010). There has been another study that showed how one or two amino acid variations can change the functionality of proteins or enzymes. Two selectively favored amino acid changes in the glucosinolate-biosynthetic locus (*BCMA*) cause a gain of novel enzyme function to control the damage by insect herbivores and higher reproduction in the wild relative of *Arabidopsis thaliana* (i.e. *Boechera stricta*). Amino acid swapping at 134 (G134L; glycine134leucine) and 536 (P536K; proline536lysine) position leads to generate two new copies (*BCMA1* and *BCMA3*) of the ancestral *BCMA* (*BCMA2*) copy. The new copies evolved with novel catalytic activity for glucosinolate producing enzyme and showed greater resistance from the damage by a diverse community of herbivores (Prasad *et al*.,

2012). Similarly, a few rice varieties got broad spectrum resistance from various strains of *Magnaporthe oryzae* pathogen due to single alterations. Li et al. (2017) have found SNP variation (33bp, A/G) at the promoter of *bsr-d1* gene. Resistant varieties were having (SNP33-G), while susceptible ones were carrying SNP33-A. Mutated allele with SNP33-G enhances the binding of MYBS1 transcription factor, which suppresses the gene expression. Less gene expression leads to lesser production of peroxidase enzyme and more availability of H_2O_2 (hydrogen peroxide) compound, which was found highly toxic to the strains of *M. oryzae*.

The discussed examples have shown how single alterations in coding or regulatory region can affects protein or gene functionality in quantitative manner and may develop favorable traits for adaptive evolution. Although different haplotypes of *HI4'OMT*/*I4'OMT* gene were found highly conservative based on sequence alignment and amino acid comparisons from the current study, however, subtle variations can play a dominant role in determining their functions. The experimental results suggest here might be function of *HI4'OMT*/*I4'OMT-H3* haplotype can differ from others due to single alteration in its amino acid sequence (isoleucine to valine, 41st amino acid). Thus, *HI4'OMT*/*I4'OMT-H3* was identified to be the most promising candidate gene, which might confer broad spectrum resistance, and was selected for further experimentations to validate their role in SCN defense for both SCN races (SCN2 and SCN5) in susceptible soybean cultivar 'Williams 82'. We had also compared the role of other alleles in the SCN resistance by checking their function through hairy root transformation.

4.3.2 Role of *HI4'OMT*/*I4'OMT-H1-H4* alleles in broad-spectrum SCN resistance

Adaptation of resistance by soybean cyst nematodes from previously identified SCN resistant genes is inevitable as nematodes evolve faster and possess more genetic diversity then host soybeans. In an effort to enhance soybean resistance against SCN; researchers have been elucidating the novel defense mechanism(s) using overexpression and gene silencing approaches (Mitchum, 2016). Though the identified and tested resistant genes have shown some potential, however no new SCN resistant trait is released in the market so far. Most of the resistant or susceptible genes that have been tested are mainly identified from microarray gene expression profiling. This technique led to the identification of a large number of genes based on the expression differences among resistant and susceptible genotypes at both global (Khan *et al*., 2004; Klink *et al*., 2007; Mazarei *et al*., 2011) and cellular levels (Kandoth *et al*., 2011; Klink *et al*., 2010, 2011 and Matsye *et al*., 2011). In addition, gene selection for function testing solely based on microarray expression analysis is not advantageous due to the complex interaction of genes with soybean and SCN. In one of the recent studies by Matthews et al. (2013), overexpression analysis of several genes has been performed on susceptible soybean roots and assessed for impacts on SCN development. The tested 100 genes have been selected exclusively from microarray gene expression analysis and from them only about 10% showed moderate SCN resistance (50% reduction of SCN). On the other hand, about 17% of the genes increased the number of SCN by 150% instead of reduction and remaining genes showed no significant effect on SCN development. These studies increase the requirement of functional validation of more putative soybean defense gene(s) critically selected *via* multiple experimental approaches and should show role in broad-spectrum SCN resistance.

Though initial identification of *HI4'OMT*/*I4'OMT* as a resistant candidate gene in the present study for both SCN races was also based on relative gene and associated metabolite (formononetin) expression differences within SCN-resistant (S54) and SCN-susceptible (S67) wild soybean genotypes (Figure 3.3.1.1). However, resistance can not only be explained by the genetic or metabolomic expression differences, because allele(s) can be monomorphic and might show similar effect on pathogens. Due to this, finalization of candidate gene for functional validation was carried out based on the allelic or haplotype variation *via* molecular cloning. There were four alleles or haplotypes of *HI4'OMT*/*I4'OMT* gene were detected from the cDNA cloning of S54, and S67. The amino acid difference was only found in two of the haplotypes i.e. haplotype II (*H2*; G202D) and haplotype III (*H3*; I41V). Haplotype II was found in both resistant and susceptible genotypes, but haplotype III (*H3*) was found specific to S54 (Figure 3.3.2.1) and chosen as a potential allele to validate their role in reduction of SCN susceptibility among soybean by hairy root transformation including other alleles to compare their roles in the defense as well.

The gene encoding *isoflavone-4'-methyl-transferase* (*I4'OMT*) or *hydroxyisoflavanone-4'-methyl-transferase* (*HI4'OMT*), are the key enzymes in the biosynthesis of methylated conjugates of isoflavonoids in legumes and act as effective phytoalexins. These genes have been cloned before in several legumes including *Glycine max* (accession, Williams 82; https://phytozome.jgi.doe.gov/pz/portal.html#!gene?search=1&detail=1&method=4433&searchText=transcriptid:30501362), Glycyrrhiza *echinata* (Akashi *et al*., 2003), *Lotus japonicus* (Asamizu *et al*., 2000), *Medicago truncatula* (accession no. XP_013441548.2),

and *Pueraria lobata* (Li *et al*., 2016) including others. However, there is limited knowledge available about their role in biotic stresses. Here, we report for the first time that *HI4'OMT*/*I4'OMT-H3* transgenic soybean plants inoculated with *H. glycine* display significantly altered responses to soybean cyst nematode (SCN2 and SCN5) pathogen infections. The normalized ratio of the number of SCNs that had developed beyond J2 stage to the total number of nematodes per plant was significantly lower for *HI4'OMT*/*I4'OMT-H3* hairy roots than the controls and other alleles of *HI4'OMT*/*I4'OMT* candidate gene when tested in the susceptible soybean cultivar ('Williams 82'). When the susceptibility levels of the control transgenic hairy roots were arbitrarily set at 100%, the transgenic soybean hairy roots overexpressing *HI4'OMT*/*I4'OMT-H3* attained 57.23% and 60.22% reduction in susceptibility for SCN2 and SCN5 races, respectively.

Similarly, the mean number of developed female cysts was found significantly lower in the *HI4'OMT*/*I4'OMT-H3* overexpressing hairy roots than the controls and other alleles containing roots among both SCN races. When the female index of the control transgenic hairy roots was arbitrarily set at 100%, the transgenic soybean hairy roots overexpressing *HI4'OMT*/*I4'OMT-H3* were found 58.31% compared with *GFP*; 59.85% compared with *GUS* and 68.87% compared with *GFP*; 69.57% compared with *GUS*, reduction in susceptibility for SCN2 and SCN5 races, respectively. These results suggest that Haplotype III (*H3* allele) of *HI4'OMT*/*I4'OMT* gene can be an effective allele for developing resistant soybean varieties for different races of SCN. For the two SCN races (SCN2 and SCN5) in the background of 'Williams 82' despite the significant developmental differences beyond J2 stage and female cysts, there were no differences observed in total SCN numbers between transgenics hairy roots having *H1-H4* constructs

and controls. This suggests that *HI4'OMT*/*I4'OMT-H3* might only involve in the hindrance of nematode growth i.e. the development beyond J2 stage and in maturity progression (female cysts) against attacking SCNs in the roots of soybeans not in the penetration of nematodes.

Although, the mechanism behind *HI4'OMT*/*I4'OMT-H3* gene in the enhancement of SCN resistance does not get crystal clear from the present study, however, the differential expression of *HI4'OMT*/*I4'OMT-H3* candidate gene with other tested genes from phenylpropanoid pathway in overexpressed and control transgenic root hints at a possible mechanism. The *HI4'OMT*/*I4'OMT-H3* overexpressed transgenic roots showed significantly induced level of transformed gene (*HI4'OMT*/*I4'OMT-H3*) with additional genes from phenylpropanoid pathway (*GmCHS7*, *GmCHS8*, *GmIFS1* and *GmIFS2*) as compare to the control roots and other alleles. To recap, *HI4'OMT*/*I4'OMT* functions as an *S*-adenosyl-L-methionine-dependent daidzein methyl transferase. It catalyzes the formation of formononetin (4'-methylated-daidzein) using daidzein as a substrate. Higher accumulation of *HI4'OMT*/*I4'OMT-H3* gene indicates the availability of more *isoflavone-4'-methyl-transferase* enzyme which possibly can induce the fast and higher conversion of daidzein into formononetin. *CHS* and *IFS* are the other key enzymes of the phenylpropanoid pathway, which play an important role in the biosynthesis of daidzein (Figure 2; Cheng *et al*., 2015). The upregulation of their gene members *GmCHS7*, *GmCHS8*, *GmIFS1* and *GmIFS2* in overexpressed transgenic roots also indicate their cooperative role in the higher accumulation of formononetin. We predicted that the four genes might enhance the availability of daidzein substrate by controlling their steps in phenylpropanoid pathway at a specific manner. This might lead to the faster conversion of

daidzein into its methylated form (formononetin) in *HI4'OMT*/*I4'OMT-H3* overexpressing transgenic roots than the control roots or the roots having *HI4'OMT*/*I4'OMT-H1*, *H2* or *H4* haplotypes constructs.

Here, the discussed data of gene and metabolomic comparisons was processed and collected from the overexpressed and control transgenic roots prior to any nematode inoculation in both races. This suggests that the higher level of resistance in *HI4'OMT*/*I4'OMT-H3*-overexpressing hairy roots is because of higher production of formononetin due to constitutive promoter, which may have primed the plants for defense responses through the modulation of genes and metabolites expression in phenylpropanoid pathway before even SCN infection. The overall results of present study advise that the addition of *HI4'OMT*/*I4'OMT-H3* candidate gene in soybean might be impactful for broad-spectrum SCN resistance.

REFERENCES

Abedinia, M., Henry, R.J., Blakeney, A.B., and Lewin, L.G. (2000). Accessing genes in the tertiary gene pool of rice by direct introduction of total DNA from *Zizania palustris* (Wild rice). *Plant Molecular Biology Reporter*, **18:** 133–138.

Aflitos, S. et al. (2014). Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *The PlantJournal*, **80:** 136–148.

Aflitos, S. et al. (2014). Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. *The PlantJournal*, **80:** 136–148.

Agrell, J., Oleszek, W., Stochmal, A., Olsen, M. and Anderson, P. (2003). Herbivore-induced responses in alfalfa (*Medicago sativa*). *Journal of Chemical Ecology*, **29(2):** 303-320.

Ahad, A., Ganai, A. A., Mujeeb, M. and Siddiqui, W. A. (2014). Chrysin, an anti-inflammatory molecule, abrogates renal dysfunction in type 2 diabetic rats. *Toxicology and Applied Pharmacology*, **279:** 1-7.

Akashi, T., Sawada, Y., Shimada, N., Sakurai, N., Aoki, T. and Ayabe, S-I. (2003). cDNA cloning and biochemical characterization of *S*-adenosyl-$_{L}$-methionine: 2,7,4'-trihydroxyisoflavanone methyltransferase, a critical enzyme of the legume isoflavonoid phytoalexin pathway. *Plant and Cell Physiology*, **44:** 103-112.

Akhunov, E., Nicolet, C., and Dvorak, J. (2009). Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theoretical and Applied Genetics*, **119:** 507–517.

Akhunov, E., Nicolet, C., and Dvorak, J. (2009). Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theoretical and Applied Genetics*, **119:** 507–517.

Alonso A., Marsal S. and Julia, A. (2015). Analytical methods in untargeted metabolomics: state of the art in 2015. *Frontiers in Bioengineering and Biotechnol*ogy, **3**: 23. doi: 10.3389/fbioe.2015.00023.

An, C., and Mou, Z. (2014). "Salicylic acid and defense responses in plants," in Phytohormones: A Window to Metabolism, Signaling and Biotechnological Applications, eds L.-S. P. Tran and S. Pal (New York: Springer Science and Business Media).

Arbelaez, J. D. et al. (2015). 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. *MolecularBreeding*, **35:** 81.

Arbona, V., Manzi, M., de Ollas, C. and Gomez-Cadenas, A. (2013). Metabolomics as a tool to investigate abiotic stress tolerance in plants. *International Journal of Molecular Sciences*, **14:** 4885-4911. doi: 10.3390/ijms 14034885.

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S. (2000). Generation of 7137 non-redundant expressed sequence tags from a legume, *Lotus japonicus*. *DNA Research*, **7:** 127-130.

Ballini, E., Berruyer, R., Morel, J. B., Lebrun, M. H., Notteghem, J. L., &Tharreau, D. (2007). Modern elite rice varieties of the 'Green Revolution' have retained a large introgression from wild rice around the *Pi33* rice blast resistance locus. *New Phytologist*, **175:** 340–350.

Balmer D., Flors V., Glauser, G. and Mauch-Mani, B. (2013). Metabolomics of cereals under biotic stress: current knowledge and techniques. *Frontiers in Plant Science*, **4:** 82. doi: 10.3389/fpls.2013.0082.

Baranski, R., Klocke, E. and Schumann, G. (2006). Green fluorescent protein as an efficient selection marker for *Agrobacterium rhizogenes* mediated carrot transformation. *Plant Cell Reports*, **25(3):** 190-197.

Bartsch, M., Bednarek, P., Vivancos, P. D., Schneider, B., von Roepenack-Lahaye, E., Foyer, C. H., Kombrink, E., Scheel, D. and Parker, J. E. (2010). Accumulation of isochorismate derived 2, 3-dihydroxybenzoic 2-*O*-*β*-D-xylosidein Arabidopsis resistance to pathogens and ageing of leaves. *Journal of Biological Chemistry*, **285(33):** 25654-25665.

Bektas, Y. and Eulgem, T. (2015). Synthetic plant defense elicitors. *Frontiers in Plant Science*, **5:** 804. doi: 10.3389/fpls.2014.00804.

Bleeker, P. M. et al. (2012). Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. *Proceedings of the National Academy of Sciences of the UnitedStates of America*, **109:** 20124–20129.

Bleeker, P. M. et al. (2012). Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. *Proceedings of the National Academy of Sciences of the UnitedStates of America*, **109:** 20124–20129.

Bleeker, P.M. et al. (2011). Tomato-produced 7-epizingiberene and R-curcumene act as repellents to whiteflies. *Phytochemistry*, **72:** 68–73.

Bleeker, P.M. et al. (2011). Tomato-produced 7-epizingiberene and R-curcumene act as repellents to whiteflies. *Phytochemistry*, **72:** 68–73.

Bodirsky, B.L., Rolinski, S., Biewald, A., Weindl, I., Popp, A. and Lotze-Campen, H. (2015). Global food demand scenarios for the 21st century. *PLoS One*; doi: 10.1371/journal.pone.0139201.

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, **33:** 41–52.

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Advances*, **33:** 41–52.

Bowne, J. B., Erwin, T. A., Juttner, J., Schnurbusch, T., Langridge, P., Bacic, A. and Roessner, U. (2012). Drought responses of leaf tissues from wheat cultivars of differing drought tolerance at the metabolite level. *Molecular Plant*, **5(2):** 418-429.

Brown, A.H.D. (1989). Core collections –A practical approach to genetic-resources management. *Genome*, **31:** 818–824.

Brown, A.H.D. (1989). Core collections –A practical approach to genetic-resources management. *Genome*, **31:** 818–824.

Brozynska, M., Furtado, A., and Henry, R.J. (2015). Genomics of crop wild relatives: Expanding the gene pool for crop improvement. *PlantBiotechnology Journal*, **14:**1070–1085.

Brozynska, M., Furtado, A., and Henry, R.J. (2015). Genomics of crop wild relatives: Expanding the gene pool for crop improvement. *PlantBiotechnology Journal*, **14:**1070–1085.

Buckler, E.S., Thornsberry, J.M. and Kresovich, S. (2001). Molecular diversity, structure and domestication of grasses. *Genetical Research*, **77:** 213–218.

Bybd D. W., Kirkpatrick T. and Barker K. R. (1983). An improved technique for clearing and staining plant tissues for detection of nematodes. *Journal of Nematol*ogy, **15**: 142-143.

Byrd, D., Kirkpatrick, T. and Barker, K. (1983). An improved technique for clearing and staining plant tissue for detection of nematodes. *Journal of Nematology*, **15:** 142–143.

Castaneda-Alvarez, N.P. et al. (2016). Global conservation priorities for crop wild relatives. *Nature Plants*, **2:** 16022.

Catford, J. G., Staehelin, C., Larose, G., Piche, Y. and Vierheilig, H. (2006). Systemically suppressed isoflavonoids and their stimulating effects on nodulation and mycorrhization in alfalfa split-root systems. *Plant Soil*, **285:** 257-266.

Chang, R.Z., Chen, Y.W., Shao, G.H., and Wan, C.W. (1994). Effect of salt stress on agronomic characters and chemical quality of seeds in soybean. *Soybean Science*, **13:** 101–105.

Chen, H., Seguin, P. and Jabaji, S. H. (2009). Differential expression of genes encoding the phenylpropanoid pathway upon infection of soybean seedlings by *Rhizoctonia solani*. *Canadian Journal of Plant Pathology*, **31(3):** 356-367.

Cheng, H., Yu, O. and Yu, D. (2008). Polymorphism of *IFS1* and *IFS2* gene are associated with isoflavone concentrations in soybean seeds. *Plant Science*, **175:** 505-512.

Cheng, H., Yu, O. and Yu, D. (2008). Polymorphisms of IFS1 and IFS2 gene are associated with isoflavone concentrations in soybean seeds. *Plant Science*, **175:** 505-512.

Cheng, Q., Li, N., Dong, L., Zhang, D., Fan, S., Jiang, L., Wang, X., Xu, P. and Zhang, S. (2015). Overexpression of soybean isoflavone reductase (*GmIFR*) enhances resistance to *Phytophthora sojae* in soybean. *Frontiers in Plant Science*, **23(6):** 1024. doi: 10.3389/fpls.2015.01024.

Cheng, Q., Li, N., Dong, L., Zhang, D., Fan, S., Jiang, L., Wang, X., Xu, P. and Zhang, S. (2015). Overexpression of soybean isoflavone reductase (*GmIFR*) enhances resistance to *Phytophthora sojae* in soybean. *Frontiers in Plant Science*, **6:** 1024. doi: 10.3389/fpls.2015.01024.

Cheng, Q., Li, N., Dong, L., Zhang, D., Fan, S., Jiang, L., Wang, X., Xu, P. and Zhang, S. (2015). Overexpression of soybean isoflavone reductase (*GmIFR*) enhances resistance to *phytophthora sojae* in soybean. Frontiers in Plant Science, **6:** 1024. doi: 10.3389/fpls.2015.01024.

Chennupati, P., Seguin, P., Chamoun, R. and Jabaji, S. (2012). Effects of high-temperature stress on soybean isoflavone concentration and expression of key genes involved in isoflavone synthesis. *Journal of Agricultural and Food Chemistry*, **60:** 12421-12427.

Chennupati, P., Seguin, P., Chamoun, R., and Jabaji, S. (2012). Effects of high-temperature stress on soybean isoflavone concentration and expression of key genes involved in isoflavone synthesis. *Journal of Agricultural and Food Chemistry*, **60:** 12421-12427.

Chin, S., Behm, C.A., and Mathesius, U. (2018). Functions of flavonoids in plant-nematode interactions. *Plants*, **7:** 85. doi: 10.3390/plants7040085.

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. and Chan, A.P. (2012). Predicting the functional of effect amino acid substitutions and indels. *PLoS One*, **7:** e46688.

Chu, S., Wang, J., Cheng, H., Yang, Q. and Yu, D. (2014). Evolutionary study of the isoflavonoid pathway based on multiple copies analysis in soybean. *BMC Genetics*, **15:** 76. doi: 10.1186/1471-2156-15-76.

Chu, S., Wang, J., Cheng, H., Yang, Q. and Yu, D. (2014). Evolutionary study of the isoflavonoid pathway based on multiple copies analysis in soybean. *BMC Genetics*, **15:** 76. http://www.biomedcentral.com/1471-2156/15/76.

Chunwongse, J., Chunwongse, C., Black, L., and Hanson, P. (2002). Molecular mapping of the *Ph-3* gene for late blight resistance in tomato. *TheJournal of Horticultural Science & Biotechnology*, **77:** 281–286.

Clark, R. M., Linton, E., Messing, J., &Doebley, J. F. (2004). Pattern of diversity in the genomic region near the maize domestication gene *tb1.Proceedings of the National Academy of Sciences of the United States ofAmerica*, **101:** 700–707.

Clough, J.C., Tuteja, J.H., Laura, M.L., Marek, F., Shoemaker, R.C. and Vodkin, L.O. (2004). Features of a 103-kb gene-rich region in soybean include an inverted perfect repeat cluster of CHS genes comprising the *I* locus. *Genome*, **47:** 819–831.

Colmer, T.D., Flowers, T.J., and Munns, R. (2006). Use of wild relatives to improve salt tolerance in wheat. *Journal of Experimental Botany*, **57:** 1059–1078.

Colmer, T.D., Flowers, T.J., and Munns, R. (2006). Use of wild relatives to improve salt tolerance in wheat. *Journal of Experimental Botany*, **57:** 1059–1078.

Cook, D.E., Lee, T.G., Guo, X., Melito, S., Wang, K., Bayless, A.M., Wang, J., Hughes, T.J., Willis, D.K., Clmente, T.E., Diers, B.W., Jiang, J., Hudson, M.E. and Bent,

A.F. (2012). Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science*, **338:** 1206-1209. doi:10.1126/science.1228746.

Dhaubhadel, S., Gijzen, M., Moy, P. and Farhangkhoee, M. (2007). Transcriptome analysis reveals a critical role of *CHS7* and *CHS8* genes for isoflavonoid synthesis in soybean seeds. *Plant Physiology*, **143:** 326-338.

Dhillon, T. and Stockinger, E.J. (2013). *Cbf14* copy number variation in the A, B and D genomes of diploid and polyploid wheat. *Theoretical and Applied Genetics*, **126:** 2777-2789. doi:10.1007/s00122-013-2171-0.

Diab, A.A., Teulat-Merah, B., This, D., Ozturk, N.Z., Benscher, D., and Sorrells, M.E. (2004). Identification of drought-inducible genes and differentially expressed sequence tags in barley. *Theoretical and Applied Genetics*, **109:** 1417–1425.

Dinkova-Kostova, A.T., and Kostov, R.V. (2012). Glucosinolates and iso- thiocyanates in health and disease. *Trends in Molecular Medicine*, **18:** 337–347.

Dixon, R. A. and Paiva, N. L. (1995). Stress-induced phenylpropanoid metabolism. *The Plant Cell*, **7:** 1085-1097.

Dixon, R.A. and Steele, C.L. (1999). Flavonoids and isoflavonoids- a gold mine for metabolic engineering. *Trends in Plant Science*, **4(10):** 1360-1385.

Du, B. et al. (2009). Identification and characterization of *Bph14*, a gene conferring resistance to brown planthopper in rice. *Proceedings of the National Academy of Sciences of the United States of America*, **106:** 22163–22168.

Durner, J., Shah, J. and Klessig D.F. (1997). Salicylic acid and disease resistance in plants. *Trends in Plant Science*, **2(7):** 266-274.

Durner, J., Shah, J. and Klessig D.F. (1997). Salicylic acid and disease resistance in plants. *Trends in Plant Science*, **2(7):** 266-274.

Eshed, Y., and Zamir, D. (1994). Introgressions from *Lycopersiconpennellii*can improve the soluble solids yield of tomato hybrids. *Theoretical and Applied Genetics*, **88:** 891–897.

Fernie, A. R., & Schauer, N. (2009). Metabolomics-assisted breeding: A viable option for crop improvement? *Trends in Genetics*, **25:** 39–48.

Fernie, A. R., & Schauer, N. (2009). Metabolomics-assisted breeding: A viable option for crop improvement? *Trends in Genetics*, **25:** 39–48.

Fernie, A.R., and Schauer, N. (2009). Metabolomics-assisted breeding: A viable option for crop improvement? *Trends in Genetics*, **25:** 39–48.

Fernie, A.R., and Schauer, N. (2009). Metabolomics-assisted breeding: A viable option for crop improvement? *Trends in Genetics*, **25:** 39–48.

Ferreyra, M. L. F., Rius, S. P. and Casati, P. (2012). Flavanoids: biosynthesis, biological functions, and biotechnological applications. *Frontiers in Plant Science*, **3:** 222. doi: 10. 3389.fpls.2012.00222.

Fetch, T., Johnston, P.A., and Pickering, R. (2009). Chromosomal location and inheritance of stem rust resistance transferred from *Hordeumbulbosum* into cultivated barley (*H. vulgare*). *Phytopathology*, **99:** 339–343.

Fischer, I., Steige, K.A., Stephan, W., and Mboup, M. (2013). Sequence evolution and expression regulation of stress-responsive genes in natural populations of wild tomato. *PLoS ONE*, **8:** e78182.

Foolad, M.R., and Panthee, D.R. (2012). Marker-assisted selection in tomato breeding. *Critical Reviews in Plant Sciences*, **31:** 93–123.

Fraire-Velazquez, S. and Balderas-Hernandez, V. E. (2013). Abiotic stress in plants and metabolic responses. Chapter: 2. doi: 10.5772/54859.

Frelichowski, J.E., and Juvik, J.A. (2001). Sesquiterpene carboxylic acids from a wild tomato species affect larval feeding behavior and survival of *Helicoverpazea*and *Spodoptera exigua* (Lepidoptera: Noctuidae). *Journalof Economic Entomology*, **94:** 1249–1259.

Frelichowski, J.E., and Juvik, J.A. (2001). Sesquiterpene carboxylic acids from a wild tomato species affect larval feeding behavior and survival of *Helicoverpazea*and *Spodoptera exigua* (Lepidoptera: Noctuidae). *Journalof Economic Entomology*, **94:** 1249–1259.

Fu, Z.Q., and Dong, X.N. (2013). Systemic acquired resistance: Turning local infection into global defense. *Annual Review of Plant Biology*, **64:** 839-863.

Fujii, M., Yokosho, K., Yamaji, N., Saisho, D., Yamane, M., Takahashi, H., Sato, K., Nakazono, M. and Ma, J.F. (2012). Acquistion of aluminium tolerance by modification of a single gene in barley. *Nature Communications*, **3:** 713.

Gholami, A., Geyter, N. De., Pollier, J., Goormachtig, S. and Goossens, A. (2014). Natural product biosynthesis in *Medicago* species. *Natural Products Reports,***31:** 356-380. doi: 10.1039/c3np70104b.

Gilbert, N. (2014). Cross-bred crops get fit faster. *Nature* **513:** 292292.

Ginwal, H.S., and Mittal, N. (2010). An efficient genomic DNA isolation protocol for RAPD and SSR analysis in *Acorus calamus* L. *Indian Journal of Biotechnology*, **9:** 213-216.

Gonzalez-Lamothe, R., Mitchell, G., Gattuso, M., Diarra, M. S., Malouin, F. and Bouarab, K. (2009). Plant antimicrobial agents and their effects on plant and human pathogens. *International Journal of Molecular Sciences,***10:** 3400-3419.

Guan, R.X., et al. (2014). Salinity tolerance in soybean is modulated by natural variation in *GmSALT3*. *The Plant Journal*, **80:** 937–950.

Guiterrez-Gonzalez, J.J., Guttikonda, S.K., Tran, L.S., Aldrich, D.L., Zhong, R., Yu, O., Nguyen, H.T. and Sleper, D.A. (2010). Differential expression of isoflavone biosynthetic genes in soybean during water deficits. *Plant Cell physiology*, **51:** 936-948.

Gupta, B. and Huang, B. (2014). Mechanism of salinity tolerance in plants: physiological, biochemical and molecular characterization. *International Journal of Genomics*. doi: 10.1155/2014/801596.

Hajjar, R., and Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica*, **156:** 1–13.

Hanikenne, M., Talke, I.N., Haydon, M.J., Lanz, C., Nolte, A., Motte, P., Kroymann, J., Weigel, D. and Kramer, U. (2008). Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of *HMA4*. *Nature*, **453:** 391-395. doi:10.1038/nature06877.

Harlan, J.R., and de Wet, J.M.J. (1971). Toward a rational classification of cultivated plants. *Taxon*, **20:** 509–517.

Haverkort, A.J., Struik, P.C., Visser, R.G.F., and Jacobsen, E. (2009). Applied biotechnology to combat late blight in potato caused by *Phytophthorainfestans*. *Potato Research*, **52:** 249–264.

Hayashi, K. and Yoshida, H. (2009). Refunctionalization of the ancient rice blast disease resistance gene Pit by the recruitment of a retrotransposon as a promoter. *The Plant Journal*, **57:** 413-425.

He, X-Z and Dixon, R.A. (2000). Genetic manipulation of *isoflavone 7-O-methyl transferase* enhances biosynthesis of 4'-O-methylated isoflavonoid phytoalexins and disease resistance in Alfalfa. *The Plant Cell,* **12:** 1689-1702.

He, X. Z. and Dixon, R. A. (2000). Genetic manipulation of isoflavone 7-O-methyltransferase enhances biosynthesis of 4'-O-methylated isoflavonoid phytoalexins and disease resistance in alfalfa. *Plant Cell*, **12**: 1689-1702.

Höfgen, R. and Willmitzer, L. (1988). Storage of competent cells for *Agrobacterium* transformation. *Nucleic Acids Research*, **16:** 9877.

Hofmann, J., El Ashry Abd El, N., Anwar, S., Erban, A., Kopka, J. and Grundler, F. (2010). Metabolic profiling reveals local and systemic responses of host plants to nematode parasitism. *The Plant Journal*, **62:** 1058-1071.

Hong, Y-S, Martinez, A., Liger-Belair, G., Jeandet, P., Nuzillard, J-M and Cilindre, C. (2012). Metabolomics reveals simultaneous influences of plant defense system and fungal growth in *Botrytis cinerea* infected *Vitis vinifera* cv. Chardonnay berries. *Journal of Experimental Botany*, **63(16):** 5773-5785.

Honsdorf, N., March, T.J., Berger, B., Tester, M., and Pillen, K. (2014). High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. *PLoS ONE*, **9:** e97047.

Hu, L., Zhang, P., Jiang, Y. and Fu, J. (2015). Metabolomic analysis revealed differential adaptation to salinity and alkalinity stress in Kentucky Bluegrass (*Poa pratensis*). *Plant Molecular Biology Reporter*, **33:** 56-68.

Hua, L., Wang, D. R., Tan, L., Fu, Y., Liu, F., Xiao, L., Zhu, Z., Fu, Q., Sun, X., Gu, P., Cai, H., McCouch, S.R. and Sun, C. (2015). *LABA1*, a domestication gene associated with long, barbed awns in wild rice. *The Plant Cell*, **27:** 1875–1888.

Huang J. S. and Barker, K. R. (1991). Glyceollin-I in soybean-cyst nematode interactions - spatial and temporal distribution in roots of resistant and susceptible soybeans. *Plant Physiology*, **96:** 1302-1307.

Huhman, D. V. and Sumner, L. W. (2002). Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. *Phytochemistry*, **59:** 347-360.

Hummelova, J., Rondevaldova, J., Balastikova, A., Lapcik, O. and Kokoska, L. (2015). The relationship between structure and in vitro antibacterial activity of selected isoflavones and their metabolites with special focus on antistaphylococcal effect of demethyltexasin. *Letters in Applied Microbiology*, **60(3):** 242-246. doi: 10.1111/lam.12361.

Huot, B., Yao, J., Montgomery, B. L. and He, S. Y. (2014). Growth-defense tradeoffs in plants: A balancing act to optimize fitness. *Molecular Plant*, **7:** 1267-1287.

Hyma, K.E. et al. (2015). Heterozygous mapping strategy (HetMappS) for high resolution genotyping-by-sequencing markers: A case study in grapevine. *PLoS ONE*, **10:** e0134880.

Hyten, D. L. et al. (2010). High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMCGenomics*, **11:** 38.

Hyten, D.L., Cannon, S.B., Song, Q., Weeks, N., Fickus, E.W., Shoemaker, R.C., Specht, J.E., Farmer, A.D., May, G.D. and Cregan, P.B. (2010). High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics*, **11:** 38. doi.org/10.1186/1471-2164-11-38.

Ishiga, Y., Uppalapati, S. R., Gill, U. S., Huhman, D., Tang, Y. and Mysore, K. S. (2015). Transcriptomic and metabolomic analyses identify a role for chlorophyll catabolism and phytoalexin during *Medicago* nonhost resistance against Asian soybean rust. *Scientific Reports*, **5:** 13061. doi:10.1038/srep13061.

Jadhav, P.R., Mahatma, M.K., Mahatma, L., Jha, S., Parekh, V.B., and Khandelwal, V. (2013). Expression analysis of key genes of phenylpropanoid pathway and phenol profiling during *Ricinus communis-Fusarium oxysporum* f. sp. *ricini* interaction. *Industrial Crops and Products*, **50:** 456-461.

James, R.A., Davenport, R.J., and Munns, R. (2006). Physiological characterization of two genes for Na+ exclusion in durum wheat, *Nax1* and *Nax2*. *Plant Physiology*, **142:** 1537–1547.

Jena, K.K. (2010). The species of the genus *Oryza* and transfer of useful genes from wild species into cultivated rice, *O. sativa*. *Breeding Science*, **60:** 518–523.

Ji, Y., Scott, J.W., Hanson, P., Graham, E., and Maxwell, D.P. (2007). Sources of resistance, inheritance, and location of genetic loci conferring resistance to members of the tomato-infecting begomoviruses. In H. Czosnek (Ed.), *Tomato yellow leaf curl virus disease: Management, molecularbiology, breeding for resistance* (pp. 343–362). Dordrecht, The Netherlands: Springer.

Jiao Y., Vuong T.D., Liu Y., Meinhardt C., Liu Y., Joshi T., Cregan P.B., Xu D., Shannon G., and Nguyen H.T. (2015). Identification and evaluation of quantitative trait loci underlying resistance to multiple HG types of soybean cyst nematode in soybean PI 437655. *Theortical and Applied Genetics*, **128:** 15-23.

Jones, J. T., Furlanetto, C. and Phillips, M. S. (2007). The role of flavonoids produced in response to cyst nematode infection of *Arabidopsis thaliana*. *Nematology,***9(5):** 671-677.

Jorge, T. F., Rodrigues, J. A., Caldana, C., Schmidt, R., van Dongen, J. T., Thomas-Oates, J. and Antonio, C. (2016). Mass spectrometry-based plant metabolomics: metabolite responses to abiotic stress. *Mass Spectrometry Reviews*, **35:** 620-649.

Kadam, S. et al. (2016). Genomic-assisted phylogenetic analysis and marker development for next generation soybean cyst nematode resistance breeding. *Plant Science*, **242:** 342–350.

Kandoth, P.K., Ithal, N., Recknor, j., Maier, T., Nettleton, D., Baum, T.J. and Mitchum, M.G. (2011). The soybean Rhg1 locus for resistance to the soybean cyst nematode *Heteroderaglycines* regulates the expression of a large number of stress-and defense-related genes in degenerating feeding cells. *Plant Physiology*, **155:** 1960-1975.

Kanehisa M., Sato Y., Kawashima, M., Furumichi M. and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research,***44**: D457-D462.

Kant M. R, Jonckheere, W., Knegt, B., Lemos, F., Liu, J., Schimmel, B. C. J., Villarroel, C. A., Ataide, L. M. S., Dermauw, W., Glas, J. J., Egas, M., Janssen, A., Van Leeuwen, T., Schuurink, R. C., Sabelis, M. W. and Alba, J. M. (2015). Mechanisms and ecological consequences of plant defence induction and suppression in herbivore communities. *Annals of Botany*, **115:** 1015-1051. doi: 10.1093/aob/mcv054.

Kawase, M., Fukunaga, K. and Kato, K. (2005). Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable insertions. *Molecular Genetics and Genomics*, **274:** 131-140.

Kelly, D., Vatsa, A., Mayham, W., Ngô, L., Thompson, A., and Kazic, T. (2016). Extracting complex lesion phenotypes in *Zea mays*. *Machine vision andApplications*, **27:** 681. doi:10.1007/s00138-015-0728-4.

Kereszt, A., Li, D., Indrasumunar, A., Nguyen, C. D., Nontachaiyapoom, S., Kinkema, M. and Gresshoff, P.M. (2007). *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nature Protocols*, **2(4):** 948-952.

Khan, R., Alkharouf, N., Beard, H., MacDonald, M., Chouikha, I., Meyer, S., Grefenstette, J., Knap, H. and Matthews, B. (2004). Microarray analysis of gene expression in soybean roots susceptible to the soybean cyst nematode two days post invasion. *Journal of Nematology*, **36:** 241-248.

Kinzler, A.J., Prokopiak, Z.A., Vaughan, M.M., Erhardt, P.W., Sarver, J.G., Trendel, J.A., Zhang, Z., and Dafoe, N.J. (2016). Cytochrome P_{450}, *CYP93A1*, as defense marker in soybean. *Biologia Plantarum*, **60(4):** 724-730.

Klink V. P., Overall C. C., Alkharouf N. W., MacDonald M. H. and Matthews, B. F. (2007). Laser capture microdissection (LCM) and comparative microarray expression analysis of syncytial cells isolated from incompatible and compatible soybean (*Glycine max*) roots infected by the soybean cyst nematode (*Heterodera glycines*). *Planta*, **226:** 1389-1409.

Klink, V.P., Hosseini, P., Matsye, P.D., Alkharouf, N.W. and Matthews, B.F. (2011). Differences in gene expression amplitude overlie a conserved transcriptomic program occurring between the rapid and potent localized resistant reaction at the syncytium of the *Glycine max* genotype Peking (PI548402) as compared to the prolonged and potent resistant reaction of PI88788. *Plant Molecular Biology*, **75:** 141-165.

Klink, V.P., Overall, C.C., Alkharouf, N.W., MacDonald, M.H. and Matthews, B.F. (2007). A time-course comparative microarray analysis of an incompatible and compatible response by *Glycine max* (soybean) to *Heterodera glycines* (soybean cyst nematode) infection. *Planta*, **226(6):** 1423-1447.

Klink, V.P., Overall, C.C., Alkhourf, N.W., MacDonald, M.H., and Matthews, B.F. (2010). Microarray detection call methodology as a means to identify and compare transcripts expressed within syncytial cells from soybean (*Glycine max*) roots undergoing resistant and susceptible reactions to the soybean cyst nematode (*Heterodera glycines*). *Journal of Biomedical and Biotechnology*, **2010:** 491217.

Koenning S.R. and Wrather J.A. (2010). Suppression of soybean yield potential in the continental United States from plant diseases estimated from 2006 to 2009. *Plant Health Program*; http://dx.doi.org/10.1094/PHP-2010-1122-01-RS.

Kofsky, J., Zhang, H., and Song, B-H. (2018). The Untapped Genetic Reservoir: The Past, Current, and Future Applications of the Wild Soybean (*Glycine soja*). *Frontiers in PlantScience*, **9:** 949. doi: 10.3389/fpls.2018.00949.

Konishi, S., Lin, S.Y., Ebana, K., Fukuta, Y., Izawa, T., Sasaki, T., and Yano, M. (2006). A SNP caused the loss of seed shattering during rice domestication. *Science*, **312:** 1392–1396.

Krens, F.A., et al. (2015). Cisgenic apple trees; development, characterization, and performance. *Frontiers inPlant Science*, **6:** 286.

Kumar, A., Tiwari, K.L., Datta, D., and Singh, M. (2014). Marker assisted gene pyramiding for enhanced Tomato leaf curl virus disease resistance in tomato cultivars. *Biologia Plantarum*, **58:** 792–797.

Kumar, M., Rawat, P., Kureel, J., Singh, A.K., Singh, D., and Maurya, R. (2011). One step synthesis of 2-hydroxymethylisoflavone and their osteogenic activity. *Bioorganic and Medicinal Chemistry Letter*, **21:** 1706–1709.

Kushalappa, A. C. and Gunnaiah, R. (2013). Metabolo-proteomics to discover plant biotic stress resistance genes. *Trends in Plant Science*, **18:** 522-531.

Lanfermeijer, F.C., Warmink, J., and Hille, J. (2005). The products of the broken *Tm-2* and the durable *Tm-2*2 resistance genes from tomato differ in four amino acids. *Journal of Experimental Botany*, **56:** 2925–2933.

Langridge, P., and Fleury, D. (2011). Making the most of '*omics*' for crop breeding. *Trends in Biotechnology*, **29:** 33–40.

Laothawornkitkul, J., Jansen, R. M. C., Smid, H. M., Bouwmeester, H. J., Muller, J. and van Bruggen, A. H. C. (2010). Volatile organic compounds as a diagnostic marker of late blight infected potato plants: A pilot study. *Crop Protection Journal*, **29:** 872-878.

Larson, G. et al. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy ofSciences of the United States of America*, **111:** 6139–6146.

Li, D., Pfeiffer, T.W., and Cornelius, P.L. (2007). Soybean QTL for yield and yield components associated with *Glycine soja* alleles. *Crop Science*, **48:** 571–581.

Li, J., Li, C., Gou, J., Wang, X., Fan, R. and Zhang, Y. (2016). An alternative pathway for formononetin biosynthesis in *Pueraria lobata*. *Frontiers in Plant Science*, **7:** 861. doi: 10.3389/fpls.2016.00861.

Li, W., Zhu, Z. et al. (2017). A natural allele of a transcription factor in rice confers broad spectrum blast resistance. *Cell*, **170:** 114-126.

Li, Y.H. et al. (2014). *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*, **32:** 1045–1052.

Li, Y.H. et al. (2014). *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*, **32:** 1045–1052.

Lin, H., Rao, J., Shi, J., Hu, C., Cheng, F., Wilson, Z. A., Zhang, D. and Quan, S. (2014). Seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars. *Journal of Integrative Plant Biology*, **56(9):** 826-836.

Lin, J., Mazarei, M., Zhao, N., Hatcher, C.N., Wuddineh, W.A., Rudis, M., Tschaplinski, T.J., Pantalone, V.R., Arelli, P.R., Hewezi, T., Chen, F. and Stewart, C.N. Jr. (2016). Transgenic soybean overexpressing *GmSAMT1* exhibits resistance to

multiple-HG types of soybean cyst nematode *Heteroderaglycines*. *Plant Biotechnology Journal*, **pp.1-10.**doi: 10.1111/pbi.12566.

Lin, J., Mazarei, M., Zhao, N., Zhu, J.J., Zhuang, X., Liu, W., Pantalone, V.R., Prakash, R.A., Stewart Jr, C. N. and Chen, F. (2013). Overexpression of a soybean salicylic acid methyltransferase gene confers resistance to soybean cyst nematode. *Plant Biotechnology Journal*, **11:** 1135-1145.

Liu S. et al. (2012). A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature*, **492:** 13. doi: 10.1038/nature11651.

Liu, C-J., Deavours, B.E., Richard, S.B., Ferrer, J-L., Blount, J.W., Huhman, D., Dixon, R.A. and Noel, J.P. (2006). Structural basis for dual functionality of iso-flavonoid *O*-methyltransferases in the evolution of plant defense response. *The Plant Cell*, **18:** 3656-3669.

Liu, S. et al. (2017). The soybean *GmSNAP18* gene underlies two types of resistance to soybean cyst nematode. *Nature Communications*, **8:** 14822. doi: 10.1038/ncomms14822.

Liu, S., Kandoth, P.K., Lakhssassi, N., Kang, J., Colantonio, V., Heinz, R., Yeckel, G., Zhou, Z., Bekal, S., Dapprich, J., Rotter, B., Cianzio, S., Mitchum, M.G. and Meksem, K. (2017). The soybean *GmSNAP18* gene underlies two types of resistance to soybean cyst nematode. *Nature Communications*, **8:** 14822. doi: 10.1038/ncomms14822.

Livak, K. J. and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods*, **25:** 402-408.

Luo J (2015). Metabolite-based genome-wide association studies in plants. *Current Opinion in Plant Biology*, **24:** 31-38. doi: 10.1016/j.pbi.2015.01.006.

Luo, J. (2015). Metabolite-based genome-wide association studies in plants. *Current Opinion in Plant Biology*, **24:** 31–38.

Lusser, M., Parisi, C., Plan, D., and Rodriguez-Cerezo, E. (2012). Deployment of new biotechnologies in plant breeding. *Nature Biotechnology*, **30:** 231–239.

Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., Li, J., Phan Tran, L-S., and Qin, F. (2015). A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. Nature Communications, 6: 8326. doi: 10.1038/ncomms93261.

Marais, G.F., Pretorius, Z.A., Marais, A.S., and Wellings, C.R. (2003). Transfer of rust resistance genes from *Triticum* species to common wheat. *SouthAfrican Journal of Plant and Soil*, **20:** 193–198.

Maron, L.G., Guimaraes, C.T., Kirst, M. et al. (2013). Aluminium tolerance in maize is associated with higher *MATE1* gene copy number. *PNAS*, **110:** 5241-5246.

Marroni, F., Pinosio, S., and Morgante, M. (2014). Structural variation and genome complexity: is dispensable really dispensable? *Current Opinion in Plant Biology*, **18:** 31-36.

Martens, S. and Mithofer, A. (2005). Flavones and flavone synthases. *Phytochemistry*, **66:** 2399-2407.

Matthews B.F., MacDonald M.H., Thai V.K. and Tucker M.L. (2003). Molecular characterization of arginine kinases in the soybean cyst nematode (*Heteroderaglycines*). *Journal of Nematol*ogy, **35**: 252-258.

Matthews, B.F., Beard, H., MacDonald, M.H., Kabir, S., Youssef, R.M., Hosseini, P. and Brewer, E. (2013). Engineered resistance and hypersusceptibility through functional metabolic studies of 100 genes in soybean to its major pathogen, the soybean cyst nematode. *Planta*, **237:** 1337-1357.

Maxted, N., and Kell, S.P. (2009). *Establishment of a global network for the in-situ conservation of crop wild relatives: Status and needs*. Rome, Italy: Food and Agriculture Organization of the United Nations Commissionon Genetic Resources for Food and Agriculture.

Maxted, N., Ford-Lloyd, B.V., Jury, S., Kell, S., and Scholten, M. (2006). Towards a definition of a crop wild relative. *Biodiversity and Conservation*, **15:** 2673–2685.

Mazarei, M., Liu, W., Al-Ahmad, H., Arelli, P., Pantalone, V.R. and Stewart, C.N. Jr. (2011). Gene expression profiling of resistant and susceptible soybean lines infected with soybean cyst nematode. *Theoretical and Applied Genetics*, **123:** 1193-1206.

Mazid, M., Khan, T. A. and Mohammad, F. (2011). Role of secondary metabolites in defense mechanisms of plants. *Biology and Medicine*, **3(2):** 232-249.

McQualter, R. B., Chong, B. F., Meyer, K., Van Dyk, D. E., O'Shea, M. G., Walton, N. J., Viitanen, P. V. and Brumbley, S. M. (2005). Initial evaluation of sugarcane as a production platform for *p*-hydroxybenzoic acid. *Plant Biotechnol Journal*, **3:** 29-41.

Melito, S., Heuberger, A.L., Cook, D., Diers, B.W., MacGuidwin, A.E. and Bent, A.F. (2010). A nematode demographics assay in transgenic roots reveals no significant impacts of the Rhg1 locus LRR-Kinase on soybean cyst nematode resistance. *BMC Plant Biology*, **9(10):** 104. doi: 10.1186/1471-2229-10-104.

Menda, N. et al. (2014). Analysis of wild-species introgressions in tomato inbreds uncovers ancestral origins. *BMC Plant Biology*, **14:** 287.

Meyer, R.S., DuVal, A.E. and Jensen, H.R. (2012). Patterns and processes in crop domestication: An historical review and quantitative analysis of 203 global food crops. *New Phytologist*, **196:** 29–48.

Miller, A. J., & Gross, B. L. (2011). From forest to field: Perennial fruit crop domestication. *American Journal of Botany*, **98:** 1389–1414.

Mitchum, M. G., Wrather, J. A., Heinz, R. D., Shannon, J. G. and Danekas, G. (2007). Variability in distribution and virulence phenotypes of *Heteroderaglycines* in Missouri during 2005. *Plant Disease*, **91:** 1473-1476.

Mitchum, M.G. (2016). Soybean resistance to the soybean cyst nematode *Heteroderaglycines*: An update. *Phytopathology*, **106(12):** 1444-1450.

Morris, P. F., Savard, M. E. and Ward, E. W. B. (1999). Identification and accumulation of isoflavonoids and isoflavone glucosides in soybean leaves hypocotyls in resistance responses to *Phytophthora megaspermaf*.sp. *glycinea*. *Physiological and Molecular Plant Pathology,***39:** 229–244.

Morris, W. F., Brain Traw, M. and Bergelson, J. (2006). On testing for a tradeoff between constitutive and induced resistance. *Oikos*, **112:** 102-110.

Moy, P., Qutob, D., Chapman, B.P., Atkinson, I., and Gijzen, M. (2004). Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. *Molecular Plant-Microbe Interactions*, **17:** 1051-1062.

Mujeeb-Kazi, A., and De Leon, J.L.D. (2002). Conventional and alien genetic diversity for salt tolerant wheats: Focus on current status and new germplasm development. *Prospects for Saline Agriculture*, **37:** 69–82.

Munns, R. et al. (2012). Wheat grain yield on saline soils is improved by an ancestral Na+ transporter gene. *Nature Biotechnology*, **30:** 360–364.

Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annual Review of Plant Biology*, **59:** 651–681.

Munns, R., Hare, R.A., James, R.A., and Rebetzke, G.J. (2000). Genetic variation for improving the salt tolerance of durum wheat. *Australian Journalof Agricultural Research*, **51:** 69–74.

Murakami, S., Nakata, R., Aboshi, T., Yoshinaga, n., Masayoshi, T., Okumoto, Y., Ishihara, A., Morisaka, H., Huffaker, A., Schmelz, E.A. and Mori, N. (2014).

Insect-induced daidzein, formononetin, and their conjugates in soybean leaves. *Metabolites*, **4:** 532-546. doi: 10.3390/metabo4030532.

Narayanan, R.A., Atz, R, Denny, R., Young, N.D. and Somers, D.A. (1999). Expression of soybean cyst nematode resistance in transgenic hairy roots of soybean. *Crop Science*, **39:** 1680-1686.

Naz, A.A., Arifuzzaman, M., Muzammil, S., Pillen, K., and Leon, J. (2014). Wild barley introgression lines revealed novel QTL alleles for root and related shoot traits in the cultivated barley (*Hordeum vulgare* L.). *BMCGenetics*, **15:** 107.

Nevo, E., and Chen, G.X. (2010). Drought and salt tolerances in wild relatives for wheat and barley improvement. *Plant Cell and Environment*, **33:** 670–685.

Nguyen D. M. C., et al. (2013). Nematicidal activity of 3,4-dihydroxybenzoic acid purified from *Terminalia nigrovenulosa* bark against *Meloidogyne incognita*. *Microbal Pathogenesis*, **59-60:** 52-59.

Nguyen, D-M-C., Seo, D-J., Nguyen, V-N., Kim, K-Y., Park, R-D. and Jung, W-J. (2013). Nematicidal activity of gallic acid purified from *Terminalia nigrovenulosa* bark against the root knot nematode *Meloidogyne incognita*. *Nematology*, **15:** 507-518.

Niblack, T. L., Wrather, J. A., Heinz, J. A. and Donald, P. A. (2003). Distribution and virulence phenotypes of *Heteroderaglycines* in Missouri. *Plant Disease*, **87:** 929-932.

Nürnberger, T., and Lipka, V. (2005). Non-host resistance in plants: new insights into an old phenomenon. *Molecular Plant Pathology*, **6:** 335–345. doi: 10.1111/j.1364-3703.2005.00279.

Obata, T. and Fernie, A. R. (2012). The use of metabolomics to dissect plant responses to abiotic stresses. *Cellular and Molecular Life Sciences*, **69:** 3225-3243. doi: 10.1007/s00018-012-1091-5.

Okrent, R. A., Brooks, M. D. and Wildermuth, M. C. (2009). Arabidopsis GH3.12 (PBS3) conjugates amino acids to 4-substituted benzoates and is inhibited by salicylate. *Journal of Biological Chemistry*, **284(15):** 9742-9745.

Palaisa, K., Morgante, M., Tingey, S., and Rafalski, A. (2004). Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. *Proceedings ofthe National Academy of Sciences of the United States of America*, **101:** 9885–9890.

Pandey, M.K., Pandey, R., Singh, V.P., Pandey, V.B., and Singh, U.P. (2002). Antifungal activity of 4′, 7-dimethoxyisoflavone against some fungi. *Mycobiology*, **30:** 55–56.

Park, S. W., Kaimoyo, E., Kumar, D., Mosher, S. and Klessig, D. F. (2007). Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. *Science*, **5(318):** 113-116.

Parniske, M., Wulff, B.B.H., Bonnema, G., Thomas, C.M., Jones, D.A., and Jones, J.D.G. (1999). Homologues of the *Cf-9* disease resistance gene (*Hcr9s*) are present at

multiple loci on the short arm of tomato chromosome1. *Molecular Plant-MicrobeInteractions*, **12:** 93–102.

Pieterse, C.M., Van Der Does, D., Zamioudis, C., Leon-Reyes, A., and Van wees, S.C. (2012). Hormonal modulation of plant immunity. *Annual Review of Cell and Development Biology,* **28:** 489-521. doi: 10.1146/annurev-cellbio-092910-154055.

Placido, D.F., Campbell, M.T., Folsom, J. J., Cui, X.P., Kruger, G.R., Baenziger, P.S., and Walia, H. (2013). Introgression of novel traits from a wild wheat relative improves drought adaptation in wheat. *PlantPhysiology*, **161:** 1806–1819.

Prasad, V.S.K. Kasavajhala, Song, B-H., Olson-Manning, C., Anderson, J.T., Lee, C-R., Eric Schranz, M., et al. (2012). A gain of function polymorphism controlling complex traits and fitness in nature. *Science*, **31:** 337(6098). doi: 10.1126/science.1221636.

Prasanna, H.C. et al. (2015). Pyramiding *Ty-2* and *Ty-3* genes for resistance to monopartite and bipartite tomato leaf curl viruses of India. *Plant Pathology*, **64:** 256–264.

Prasanna, H.C., Kashyap, S., Krishna, R., Sinha, D.P., Reddy, S., and Malathi, V.G. (2015). Marker assisted selection of *Ty-2* and *Ty-3* carrying tomato lines and their implications in breeding tomato leaf curl disease resistant hybrids. *Euphytica*, **204:** 407–418.

Prischmann, D.A., Dashiell, K. E., Schneider, D.J., and Eubanks, M.W. (2009). Evaluating *Tripsacum*-introgressed maize germplasm after infestation with western corn

rootworms (Coleoptera: Chrysomelidae). *Journal ofApplied Entomology*, **133:** 10–20.

Qi, X. et al. (2014). Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. *Nature Communications*, **5:** 4340.

Rahaman, M.M., Chen, D.J., Gillani, Z., Klukas, C., and Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Frontiers in Plant Science*, **6:** 619.

Reichling, J. (2010). Plant-microbe interactions and secondary metabolites with antibacterial, antifungal and anti-viral properties. Annual Plant Reviews, Functions and Biotechnology of Secondary Metabolites (Second Edition), volume 39, chapter-4, page no. 224. edited by Michael Wink, publication by Wiley-Blackwell (A John Wiley & Sons, Ltd., Publication).

Rivas-San Vicente, M. and Plasencia, J. (2011). Salicylic acid beyond defence: its role in plant growth and development. *Journal of Experimental Botany*, **62(10):** 3321-3328. doi: 10.1093/jxb/err031.

Rodziewicz, P., Swarcewicz, B., Chmielewska, K., Wojakowska, A. and Stobiecki, M. (2014). Influence of abiotic stresses on plant proteome and metabolome changes. *ActaPhysiologia Plantarum*, **36(1):** 1-19.

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods of Molecular Biology*, **132:** 365–386.

Ryan, P.R., Raman, H., Gupta, S., Sasaki, T., Yamamoto, Y. and Delhaize, E. (2010). The multiple origins of aluminium resistance in hexaploid wheat include *Aegilops tauschii* and more recent cis mutations to *TaALMT1*. *The Plant Journal*, **64:** 446-455.

Saintenac, C., Zhang, W.J., Salcedo, A., Rouse, M.N., Trick, H.N., Akhunov, E., and Dubcovsky, J. (2013). Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. *Science*, **341:** 783–786.

Sairam, R. K. and Srivastava, G. C. (2001). Water stress tolerance of wheat (*Triticum aestivum* L.): Variations in hydrogen peroxide accumulation and antioxidant activity in tolerant and susceptible genotypes. *Journal of Agronomy and Crop Science*, **186:** 63-70.

Samarghandian, S., Azimi-Nezhad, M., Samini, F. and Farkhondeh, T. (2015). Chrysin treatment improves diabetes and its complications in liver, brain, and pancreas in streptozotocin-induced diabetic rats. *Canadian Journal of Physiology and Pharmacology*, **94(4):** 388-393.

Sarikamis, G., Marquez, J., MacCormack, R., Bennett, R.N., Roberts, J., and Mithen, R. (2006). High glucosinolate broccoli: A delivery system for sulforaphane. *Molecular Breeding*, **18:** 219–228.

Saxena, R.K., Edwards, D., and Varshney, R.K. (2014). Structural variations in plant genomes. *Briefing in Functional Genomics*, **13:** 296-307. doi: 10.1093/bfgp/elu016.

Schaart, J.G., van de Wiel, C.C.M., Lotz, L.A.P., and Smulders, M.J.M. (2016). Opportunities for products of new plant breeding techniques. *Trends in Plant Science*, **21:** 438–449.

Schaart, J.G., van de Wiel, C.C.M., Lotz, L.A.P., and Smulders, M.J.M. (2016). Opportunities for products of new plant breeding techniques. *Trends in Plant Science*, **21:** 438–449.

Schmalenbach, I., Korber, N., and Pillen, K. (2008). Selecting a set of wild barley introgression lines and verification of QTL effects for resistance to powdery mildew and leaf rust. *Theoretical and Applied Genetics*, **117:** 1093–1106.

Schmollinger, S., Muhlhaus, T., Boyle, N.R., Blaby, I.K., Casero, D., Mettler, T., and Merchant, S.S. (2014). Nitrogen-sparing mechanisms in *Chlamydomonas* affect the transcriptome, the proteome, and photosynthetic metabolism. *Plant Cell*, **26:** 1410–1435.

Seah, S., Yaghoobi, J., Rossi, M., Gleason, C.A., and Williamson, V.M. (2004). The nematode-resistance gene, *Mi-1*, is associated with an inverted chromosomal segment in susceptible compared to resistant tomato. *Theoretical and Applied Genetics*, **108:** 1635–1642.

Sepiol, C.J., Yu, J., and Dhaubhadel, S. (2017). Genome-wide identification of *chalcone reducatse* gene family in soybean: Insight into root-specific *GmCHRs* and *Phytophthora sojae* resistance. *Frontiers in Plant Science*, **8:** 2073. doi: 10.3389/fpls.2017.02073.

Shine, M. B., Yang, J-W., El-Habbak, M., Nagyabhyru, P., Fu, D-Q., Navarre, D., Ghabrial, S., Kachroo, P. and Kachroo, A. (2016). Cooperative functioning between phenylalanine ammonia lyase and isochorismate synthase activities contributes to salicylic acid biosynthesis in soybean. *New Phytologist*, **212:** 627-636.

Shulaev, V., Cortes, D., Miller, G. and Mittler, R. (2008). Metabolomics for plant stress response. *Physiologia Plantarum*, **132:** 199-208.

Smith-Becker, J., Marois, E., Huguet, E. J., Midland, S. L., Sims, J. J. and Keen, N. T. (1998). Accumulation of salicylic acid and 4-hydroxybenzoic acid in phloem fluids of cucumber during systemic acquired resistance is preceded by a transient increase in phenylalanine ammonia-lyase activity in petioles and stems. *Plant Physiology*, **116:** 231-238.

Soares, M. S., da Silva, D. F., Forim, M. R., da Silva, M. F., Fernandes, J. B., Vieira, P. C., Silva, D. B., Lopes, N. P., de Carvalho, S. A., de Souza, A. A. and Machado M. A. (2015). Quantification and localization of hesperidin and rutin in *Citrus sinesis* grafted on *C. limonia* after *Xylella fastidiosa* infection by HPLC-UV and MALDI imaging mass spectrometry. *Phytochemistry*, **115:** 161-170.

Song, Q.J., Hyten, D.L., Jia, G.F., Quigley, C.V., Fickus, E.W., Nelson, R.L., and Cregan, P.B. (2015). Fingerprinting soybean germplasm and its utility in genomic research. *G3: Genes Genomes Genetics*, **5:** 1999–2006.

Song, W.Y. et al. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science*, **270:** 1804–1806.

Soria-Guerra, R., Rosales-Mendoza, S., Chang, S., Haudenshield, J.S., Padmanaban, A., and Rodriguez-Zas, S. (2010). Transcriptome analysis of resistant and susceptible genotypes of *Glycine tomentella* during *Phaskopsorapachyrhizi* infection reveals novel rust resistance genes. *Theoretical and Applied Genetics*, **120:** 1315-1333.

Spoel, S.H., and Dong, X.N. (2012) How do plants achieve immunity? Defence without specialized immune cells. *Nature Reviews Immunology*, **12:** 89-100.

Strauch R. C., Svedin E., Dilkes B., Chapple C. and Li, X. (2015). Discovery of a novel amino acid racemase through exploration of natural variation in *Arabidopsis thaliana*. *Proceedings of the Natural Academy of Sciences USA*, **112**: 11726-11731.

Su, H., Jiang, H. and Li, Y. (2015). Effects of *PAL* and *ICS* on the production of total flavonoids, daidzein and puerarin in *Pueraria thomsonii*Benth. suspension cultures under low light stress. *Journal of Plant Biochemistry and Biotechnology,***24(1):** 34-41. doi: 10.1007/s13562-013-0233-7.

Subramanian, S., Hu, X., Lu, G., Odelland, J. T. and Yu, O. (2004). The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. *Plant Molecular Biology*, **54:** 623–639.

Suprunova, T., Krugman, T., Distelfeld, A., Fahima, T., Nevo, E., and Korol, A. (2007). Identification of a novel gene (*Hsdr4*) involved in water-stress tolerance in wild barley. *Plant Molecular Biology*, **64:** 17–34.

Suprunova, T., Krugman, T., Fahima, T., Chen, G., Shams, I., Korol, A., and Nevo, E. (2004). Differential expression of dehydrin genes in wild barley,

Hordeumspontaneum, associated with resistance to water deficit. *Plant Cell and Environment*, **27:** 1297–1308.

Sutton, T., Baumann, U., Hayes, J., Collins, N.C., Shi, B.J., Schnurbusch, T., Hay, A., Mayo, G., Pallotta, M., Tester, M. and Langridge, P. (2007). Boron-toxicity tolerance in barley arising from efflux transporter amplification. *Science*, **318:** 1446-1449. doi:10.1126/science.1146853.

Swamy, B.P.M., and Sarla, N. (2008). Yield-enhancing quantitative trait loci (QTLs) from wild species. *Biotechnology Advances*, **26:** 106–120.

Sweeney, M., and Mccouch, S. (2007). The complex history of the domestication of rice. *Annals of Botany*, **100:** 951–957.

Tabashnik, B.E. (2010). Communal benefits of transgenic corn. *Science*, **330:** 189–190.

Tanksley, S. D. et al. (1996). Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L-pimpinellifolium*. *Theoretical and Applied Genetics*, **92:** 213–224.

Tanksley, S.D., and McCouch, S.R. (1997). Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science*, **277:** 1063–1066.

Tenenboim H and Brotman Y (2016). Omic relief for the biotically stressed: metabolomics of plant biotic interactions. *Trends in Plant Science*, **21(9):** 781-791.

Tian, Z., Liu, S.B., Wang, Y.C., Li, X.Q., Zheng, L.H. and Zhao, M.G. (2013). Neuroprotective effects of formononetin against NMDA-induced apoptosis in cortical neurons. *Phytotherapy Research*, **27:** 1770-1775.

Tovkach, A., Ryan, P.R., Richardson, A.E., Lewis, D.C., Rathjen, T.M., Ramesh, S., Tyerman, S.D. and Delhaize, E. (2013). Transposon-mediated alteration of *TaMATE1B* expression in wheat confers constitutive citrate efflux from root apices. *Plant Physiology*, **161:** 880-892.

Tsao, R., Papadopoulos, Y., Yang, R., Young, J. C. and McRae, K. (2006). Isoflavone profiles of red clovers and their distribution in different parts harvested at different growing stages. *Journal of Agricultural and Food Chemistry*, **54**: 5797-5805.

Tylka, G.L. and Marett, C.C. (2017). Known distribution of the soybean cyst nematode, *Heteroderaglycines*, in the United States and Canada, 1954 to 2017. *Plant Health Progress*, 18: 167-168.

Uitdewilligen, J.G.A.M.L., Wolters, A.M.A., D'hoop, B.B., Borm, T.J.A., Visser, R.G.F., and van Eck, H.J. (2013). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous auto-tetraploid potato. *PLoS ONE*, **8:** e62355.

Vanblaere, T., Szankowski, I., Schaart, J., Schouten, H., Flachowsky, H., Broggini, G.A.L., and Gessler, C. (2011). The development of a cisgenic apple plant. *Journal of Biotechnology*, **154:** 304–311.

Varshney, R.K., Graner, A., and Sorrells, M.E. (2005). Genomics-assisted breeding for crop improvement. *Trends in Plant Science*, **10:** 621–630.

Vidavski, F., Czosnek, H., Gazit, S., Levy, D., and Lapidot, M. (2008). Pyramiding of genes conferring resistance to *Tomato yellow leaf curl virus* from different wild tomato species. *Plant Breeding*, **127:** 625–631.

Vishnuvathan, V.J., Lakshmi, K.S., and Srividya, A.R. (2016). Medicinal uses of formononetin- A review. *The Journal of Ethnobiology and Traditional Medicine*: *Photon*, **126:** 1197-1209.

von Korff, M., Wang, H., Leon, J., &Pillen, K. (2004). Development of candidate introgression lines using an exotic barley accession (*Hordeumvulgare* ssp. spontaneum) as donor. *Theoretical and Applied Genetics*, **109:** 1736–1745.

Wan J., et al. (2015). Whole-genome gene expression profiling revealed genes and pathways potentially involved in regulating interactions of soybean with cyst nematode (*Heteroderaglycines* Ichinohe). *BMC Genomics*, **16:** 148.

Wang, J-Y., Zhou, L., Chen, B., Sun, S., Zhang, W., Li, M., Tang, H., Jiang, B-L., Tang, J-L. and He, Y-W. (2015). A functional 4-hydroxybenzoate degradation pathway in the phytopathogen *Xanthomonas campestris* is required for full pathogenicity. *Scientific Reports*, **5:** 18456 (doi: 10.1038/srep18456).

War A. R. et al. (2012). Mechanisms of plant defense against insect herbivores. *Plant Signaling and Behaviour*, **7**: 1306–1320.

Wittstock, U. and Gershenzon, J. (2002). Constitutive plant toxins and their role in defense against herbivores and pathogens. *Current Opinion in Plant Biology*, **5:**doi: 10.1016/S1369-5266.

Wright, S. I., Bi, I. V., Schroeder, S. G., Yamasaki, M., Doebley, J. F., McMullen, M. D., &Gaut, B. S. (2005). The effects of artificial selection of the maize genome. *Science*, **308:** 1310–1314.

Xavier, A., Muir, W.M., and Rainey, K.M. (2016). Assessing predictive properties of genome-wide selection in soybeans. *G3 (Bethesda)*, **6:** 2611–2616.

Xia J. G., Sinelnikov I. V., Han B. and Wishart, D. S. (2015). MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Research,***43**: W251-W257.

Xie, X. B. et al. (2008). Fine mapping of a yield-enhancing QTL cluster associ- ated with transgressive variation in an *Oryza sativa* × *O-rufipogon*cross. *Theoretical and Applied Genetics*, **116:** 613–622.

Xu, K., Xu, X., Fukao, T., Canlas, P., Maghirang-Rodriguez, R., Heuer, S., Ismail, A.M., Bailey-Serres, J., Ronald, P.C. and Mackill, D.J. (2006). *Sub1A* is an ethylene-response-factor- like-gene that confers sub-mergence tolerance to rice. *Nature*, **442:** 705-708. doi:10.1038/nature04920.

Xu, X-H et al. (2015). Friend or foe: differential responses of rice to invasion by mutualistic or pathogenic fungi revealed by RNA-seq and metabolite profiling. *Scientific Reports*, **5:** 13624. doi: 10.1038/srep13624.

Xu, X. et al. (2012). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology*, 30: 105–111.

Yan, K., Liu, P., Wu, C-A., Yang, G-D., Xu, R., Guo, Q-H., Huang, J-G., and Zheng, C-C. (2012). Stress induced alternative splicing provides a mechanism for the regulation of microRNA processing in Arabidopsis thaliana. *Molecular Cell*, **48(4):** 521-531.

Yang, L., Li, B., Zheng, X-Y., Li, J., Yang, M., Dong, X., He, G., An, C. and Deng, X.W. (2014). Salicylic acid biosynthesis is enhanced and contributes to increased biotrophic pathogen resistance in Arabidopsis hybrids. *Nature Communications,* **6:** 7309. doi: 10.1038/ncomms8309.

Yi, J., Derynck, M.R., Chen, L., and Dhaubhadel, S. (2010). Differential expression of *CHS7* and *CHS8* genes in soybean. *Planta*, **231:** 741-753.

Yuan, X., Yin, P., Hao, Q., Yan, C., Wang, J. and Yan, N. (2010). Single amino acid alteration between valine and isoleucine determines the distinct pyrabactin selectivity by PYL1 and PYL2. The *Journal of Biological Chemistry*, **285(37):** 28953-28958.

Zabala, G., Zou, J., Tuteja, J., Gonzalez, D., Clough, S.J., and Vodkin, L.O. (2006). Transcriptome changes in the phenylpropanoid pathway of *Glycine max* in response to *Pseudomonas syringae* infection. *BMC Plant Biology*, **6:** 26.

Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nature Reviews Genetics*, **2:** 983–989.

Zamir, D. et al. (1994). Mapping and introgression of a *Tomato yellow leaf curl virus* tolerance gene, *Ty-1*. *Theoretical and Applied Genetics*, **88:**141–146.

Zhang H, Li C, Davis EL, Wang J, Griffin JD, Kofsky J and Song B-H (2016). Genome-wide association study of resistance to soybean cyst nematode (*Heteroderaglycines*) HG type 2.5.7 in wild soybean (*Glycine soja*). *Frontiers in Plant Science*, **7:** 1214. doi:10.3389/fpls.2016.01214.

Zhang H. and Song B. H. (2017). RNA-seq data comparisons of wild soybean genotypes in response to soybean cyst nematode (*Heteroderaglycines*). *Genomics data*, **14**: 36-39.

Zhang H. Y., Kjemtrup-Lovelace S., Li C. B., Luo Y., Chen L. P. and Song B. H. (2017). Comparative RNA-seq analysis uncovers a complex regulatory network for soybean cyst nematode resistance in wild soybean (*Glycine soja*). *Scientific Reports*, **7**. Article Number: 9699.

Zhang, C., Wang, X., Zhang, F., Dong, L., Wu, J., Cheng, Q., Qi, D., Yan, X., Jiang, L., Fan, S., Li, N., Li, D., Xu, P., and Zhang, S. (2017). *Phenylalanine ammonia-lyase 2.1* contributes to the soybean response towards *Phytophthora sojae* infection. *Scientific Reports*, **7:** 7242. doi: 10.1038/s41598-017-07832-2.

Zhang, H., Mittal, N., Leamy, L.J., Barazani, Oz. and Song, B-H. (2017). Back into the wild- Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications*, **10:** 5-24.

Zhang, J., Yang, D., Li, M. and Shi, L. (2016). Metabolic profiles reveal changes in wild and cultivated soybean seedlings leaves under salt stress. *PLoS ONE*, **11(7):** e0159622. doi: 10.1371/journal.pone.0159622.

Zhang, K., Halitschke, R., Yin, C., Liu, C-J. and Gan, S-S. (2013). Salicylic acid 3-hydroxylase regulates Arabidopsis leaf longevity by mediating salicylic acid catabolism. *Proceedings of the National Academy of Sciences,***110(36):** 14807-14812.

Zhao, N., Wang, G., Norris, A., Chen, X., and Chen, F. (2013). Studying plant secondary metabolism in the age of genomics. *Critical Reviews in Plant Sciences*, **32:** 369-382.

Zhou, J., Zhang, L., Chang, Y., Lu, X., Zhu, Z. and Xu, G. (2012). Alteration of leaf metabolism in *Bt*-transgenic rice (*Oryza sativa* L.) and its wild type under insecticide stress. *Journal of Proteome Research*, **11:** 4351-4360.

Zhou, Y. L. et al. (2009). Pyramiding *Xa23* and *Rxo1* for resistance to two bacterial diseases into an elite indica rice variety using molecular approaches. *MolecularBreeding*, **23:** 279–287.

Zhou, Y.L. et al. (2011). Improvement of bacterial blight resistance of hybrid rice in China using the *Xa23* gene derived from wild rice (*Oryza rufipogon*). *Crop Protection*, **30:** 637–644.

Zhou, Z.K. et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *NatureBiotechnology*, **33:** 408–414.

Zhu, J. and Park, K-C. (2005). Methyl salicylate, a soybean aphid-induced plant volatile attractive to the predator *Coccinellaseptempunctata*. *Journal of Chemical Ecology*, **31(8):** 1733-1746.

Annexure I: Composition of Different Stock Solutions

Table 2.1.3.1 Composition of pure metabolites standards

S. No.	Components	Composition
1	Metabolites standards (as per M.W.)	0.000122-0.000154 gm
2	Dimethylsulfoxide (DMSO)	1 mL
3	Isopropyl alcohol	9 mL
4	Total volume (volume made upto)	10 mL
Dissolve weighed standard metabolites in DMSO and isopropylalcohol.		

Table 2.2.3.1 Composition of 10X TAE buffer

S. No.	Components	Composition
1	89 mM Tris base	54 gm
2	89 mM Acetic acid	27.50 mL
3	2.24 mM Na_2.EDTA	4.15 gm
4	Sterile H_2O	to 500 mL
5	Total volume (volume made upto)	1000 mL

Table 2.3.2.1.1 Composition of CTAB extraction buffer

S. No.	Components	Composition
1	2% CTAB	2.000 gm
2	1.42 M NaCl	8.298 gm
3	20 mM EDTA	0.744 gm
4	100 mM Tris HCl (pH 8.0)	1.211 gm
5	4% PVP 40 (w/v)	4.000 gm
6	5 mM Ascorbic acid	0.088 gm
7	Sterile H_2O	to 70 mL
8	Total volume (volume made upto)	100 mL

Table 2.3.5.1a Composition of LB (Luria-Bertani) liquid medium

S. No.	Components	Composition
1	Tryptone	10 gm
2	NaCl	5.0 gm
3	Yeast extract	5.0 gm
4	Sterile H_2O	800 mL
5	Total volume (volume made upto)	1000 mL
Combine the reagents and shake until the solutes have dissolved. Sterilize by autoclaving for 45 minutes at 121 °C under 20 psi on liquid cycle		

Table 2.3.5.1b Composition of LB (Luria-Bertani) solid medium

S. No.	Components	Composition
1	Tryptone	10 gm
2	NaCl	5.0 gm
3	Yeast extract	5.0 gm
4	Agar	15 gm
5	Sterile H_2O	700 mL
6	Total volume (volume made upto)	1000 mL
Combine the reagents and shake until the solutes have dissolved. Sterilize by autoclaving for 45 minutes at 121 °C under 20 psi on liquid cycle		

Table 2.3.5.1c Composition of 50mg/ml ampicillin and kanamycin solutions

S. No.	Components	Composition
1	Ampicillin/Kanamycin	0.5 gm
2	Sterile H_2O	10 mL
Dissolve Ampicillin or Kanamycin completely in the sterile water and sterilize through 0.22 μm syringe filter and store 1 ml aliquots at -20 °C for up to 1 year.		

Table 2.3.5.1d Composition of 100mM (100X) IPTG Solution

S. No.	Components	Composition
1	IPTG	0.238 gm
2	Sterile H_2O	10 mL
Dissolve IPTG completely in the sterile water and sterilize through 0.22 μm syringe filter and store 1 ml aliquots at -20 °C for up to 1 year.		

Table 2.3.5.1e Composition of 20mg/ml X-Gal Solution

S. No.	Components	Composition
1	X-Gal	0.200 gm
2	Dimethylformamide (DMF)	10 mL
Dissolve X-Gal completely in DMF by vortex and sterilize by autoclaving for 45 minutes at 121 °C under 20 psi on liquid cycle. Wrap the tube in aluminum foil and store at -20 °C.		

Annexure II: Supplemental Tables and Figures

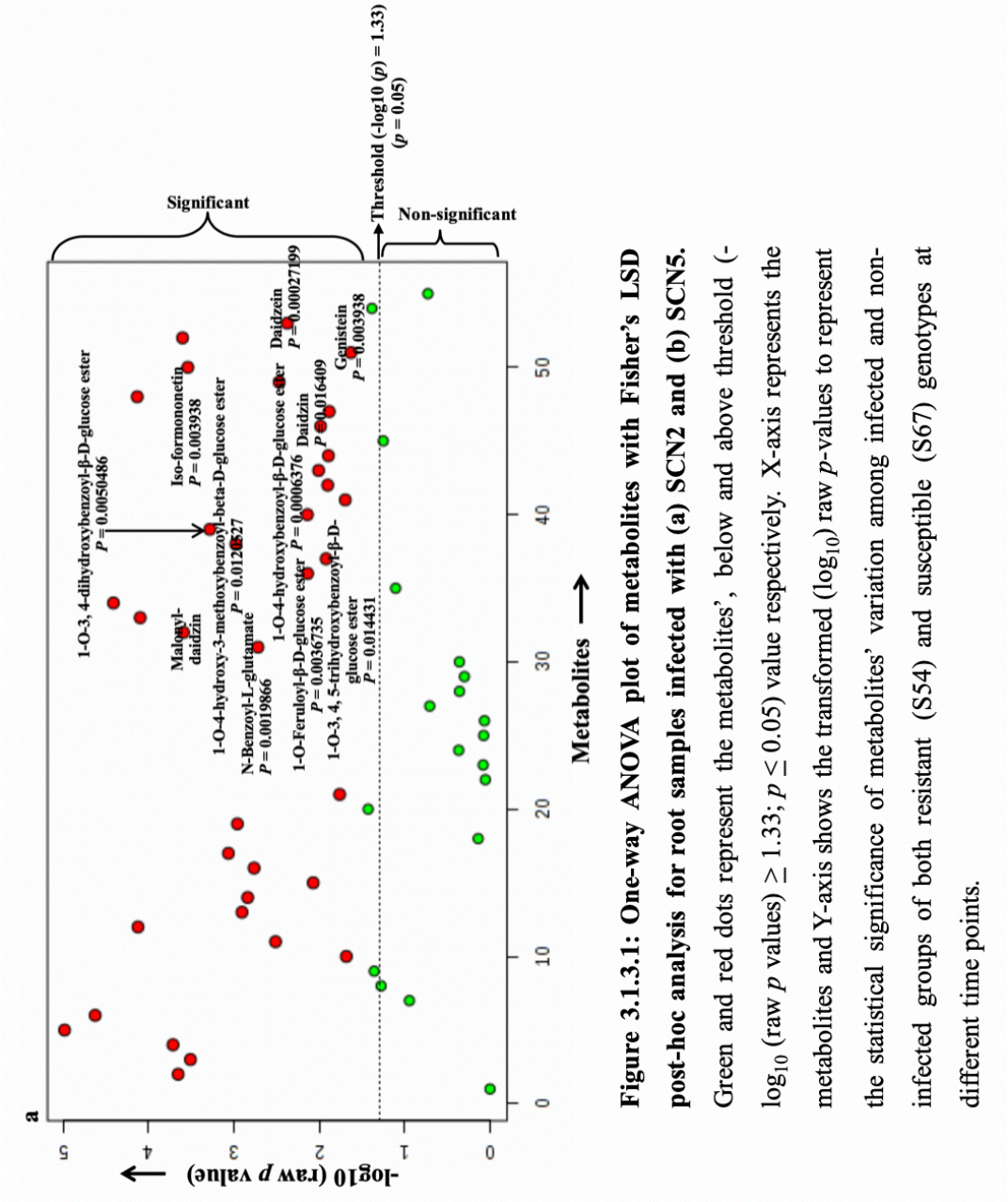
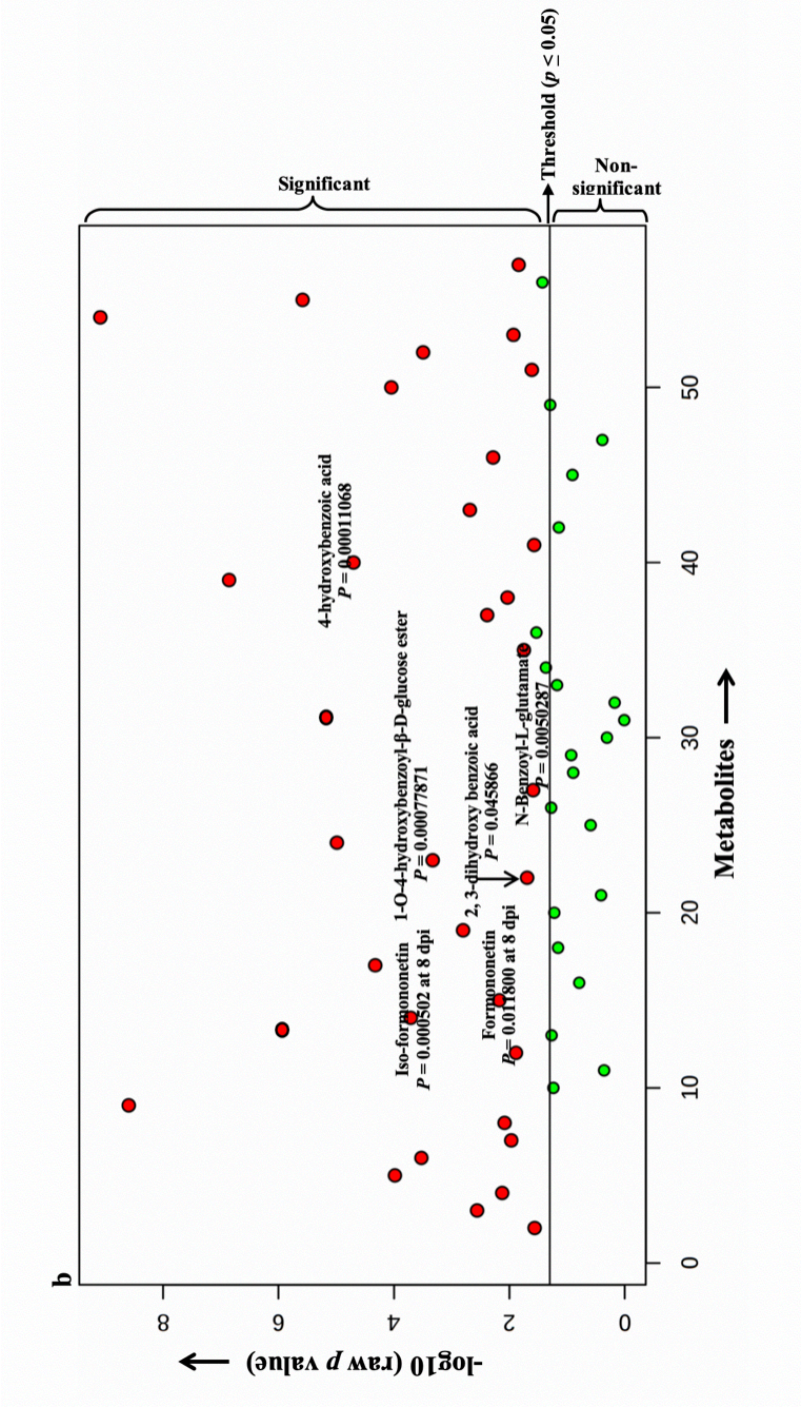


Figure 3.1.3.1: One-way ANOVA plot of metabolites with Fisher's LSD post-hoc analysis for root samples infected with (a) SCN2 and (b) SCN5. Green and red dots represent the metabolites', below and above threshold (-$\log_{10}$ (raw p values) ≥ 1.33; $p \leq 0.05$) value respectively. X-axis represents the metabolites and Y-axis shows the transformed ($\log_{10}$) raw p-values to represent the statistical significance of metabolites' variation among infected and non-infected groups of both resistant (S54) and susceptible (S67) genotypes at different time points.


b
Significant
Non-
significant
Threshold ($p \leq 0.05$)
4-hydroxybenzoic acid
$P = 0.00011068$
1-O-4-hydroxybenzoyl-β-D-glucose ester
$P = 0.00077871$
Iso-formononetin
$P = 0.000502$ at 8 dpi
2, 3-dihydroxy benzoic acid
$P = 0.045866$
Formononetin
$P = 0.011800$ at 8 dpi
N-Benzoyl-L-glutamate
$P = 0.0050287$
-log10 (raw p value)
Metabolites
0
2
4
6
8
0
10
20
30
40
50

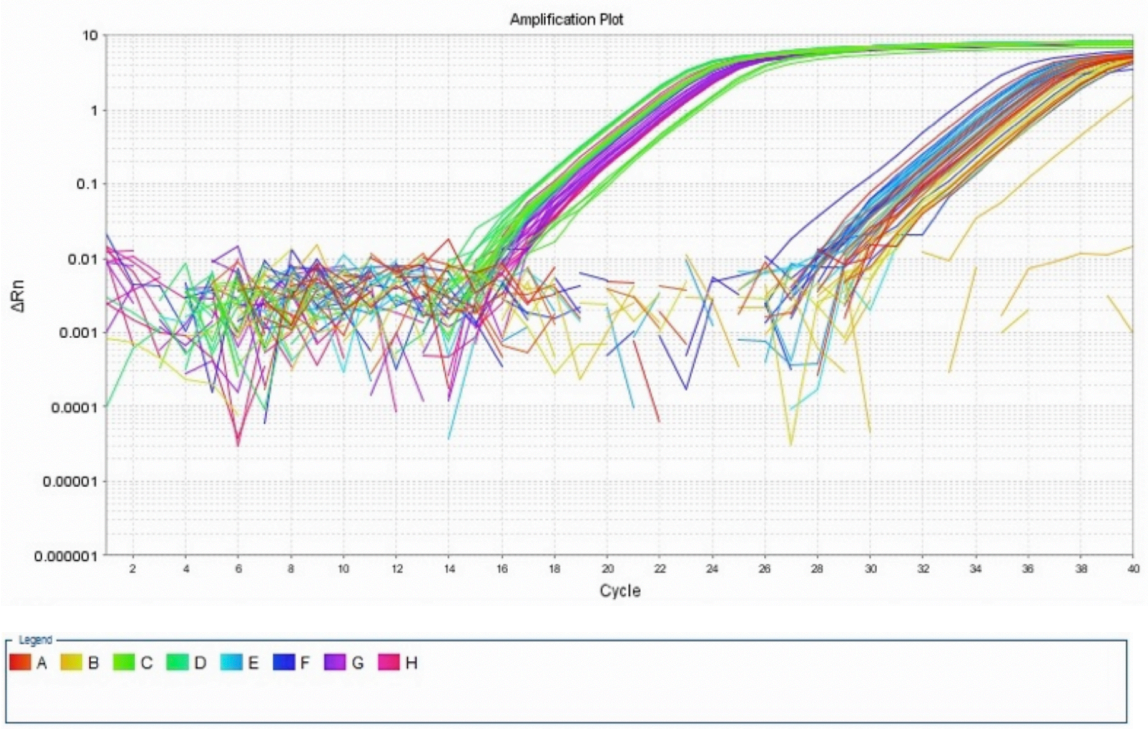


Figure 2.2.6.1: qRT-PCR amplification plot of phenyl-propanoid/iso-flavonoid pathway gene(s)

Table 3.2.1: RNA-seq Log_2 fold change gene expression data values from phenylpropanoid pathway among SCN-resistant (S54) and SCN-susceptible (S67) genotypes

S. No.	Enzyme Name	Enzyme I.D.	Gene I.D.	RNA-Sequencing: Log_2 Fold Change Expression			
				SCN2 (HG type 1.2.5.7)		SCN5 (HG type 2.5.7)	
				S54	S67	S54	S67
1	Phenylalanine ammonia lyase (*PAL*)	EC 4.3.1.24/ 4.3.1.25	Glyma.02g309300	1.286270535	0.816903353	0.151622513	-0.082018832
			Glyma.03g181600	NA	NA	NA	NA
			Glyma.03g181700	NA	NA	NA	NA
			Glyma.10g058200	NA	NA	NA	NA
			Glyma.10g209800	NA	NA	NA	NA
			Glyma.13g145000	0.643399609	0.394800721	-0.090633642	-0.150418924
			Glyma.19g182300	NA	NA	NA	NA
			Glyma.20g180800	NA	NA	NA	NA
2	Cinnamic acid hydroxylase (*C4H*)	EC 1.14.13.11	Glyma.02g236500	-0.08068741	0.224284493	-0.456242535	-0.485641235
			Glyma.10g275600	0	0	0	0
			Glyma.14g205200	0.251303555	0.371890745	-0.187573153	-0.078292197
			Glyma.20g114200	2.089688199	1.458955721	2.872298474	2.70603191
3	4-coumarate-CoA ligase (*4CL*)	EC 6.2.1.12	Glyma.01g232400	1.387128496	0.746470634	0.623290189	0.050269327
			Glyma.07g112700	0	0	0	0
			Glyma.11g010500	1.22881133	0.632160308	0.425659854	-0.236776361
			Glyma.11g091600	2.218881279	0	2.936367237	1.11872456
			Glyma.13g095600	0.086899954	0.215137532	-0.290826249	0.100772727
			Glyma.13g152900	1.72472175	0.225012808	1.502629516	-0.317907612
			Glyma.13g323000	0.86646062	0.225012808	-0.080705312	-0.265230229
			Glyma.13g372000	0.424884264	0.528805695	0.056391826	0.439904513
			Glyma.15g001700	-0.087718621	0.040128344	-0.026568348	0.110743382
			Glyma.17g064400	0.049240845	-0.200021074	-0.379818176	0.675124409
			Glyma.17g064500	0	0	0	0
			Glyma.17g064600	0.898826469	0.763175396	-0.159879568	0.135143442
			Glyma.18g044900	0.387824472	0.147981491	0.836209674	-0.067989546
			Glyma.18g136100	0	0	0	0
4	Chalcone synthase (*CHS*)	EC 2.3.1.74	Glyma.01g091400	0.237747459	2.508188553	0.274568351	-0.213278527
			Glyma.01g228700	0.790895131	0.480628318	0.104022693	-0.359572914
			Glyma.02g130400	1.201651566	0.430701673	0.614928371	0.177972432
			Glyma.05g153100	0	0	0	0
			Glyma.05g153200	0.986815512	0.921762682	1.203301701	0.344964857
			Glyma.06g118500	0	0	0	0.223247019
			Glyma.08g109200	3.590381921	1.130243292	1.58579943	0.154230686
			Glyma.08g109300	-1.575038576	2.040858486	0.887038595	0.542070101
			Glyma.08g109400	2.263557906	1.940118295	2.110658476	0.995461743
			Glyma.08g109500	1.145775159	0.725973263	1.799330743	0.40419698
			Glyma.08g110300	0	0	0	0
			Glyma.08g110400	-1.192693168	0	0	0
			Glyma.08g110500	0	0	0.35247839	0.086949838
			Glyma.08g110700	0	0	0	0
			Glyma.08g110900	0.657033842	1.614988051	1.567532794	1.541314072
			Glyma.09g074900	0	0	0	0
			Glyma.09g075200	1.265783404	1.884169872	0.559655361	-1.025373687
			Glyma.11g011500	0.605409881	0.610234992	-0.172093931	-0.441203922
			Glyma.13g034300	0	0	0	0
			Glyma.19g105100	1.381875814	1.690197788	-1.667355407	-0.364368735
5	Chalcone isomerase (*CHI*)	EC 5.5.1.6	Glyma.03g154600	0.311154421	-0.066738986	0.276105603	0.548393535
			Glyma.04g222400	0.054449392	0.27756474	-0.024045348	-0.033095517
			Glyma.06g143000	0.550022178	0.337750779	0.139461672	-0.120641897
			Glyma.10g292200	0.357676274	0.348953558	-0.289616983	-0.182056566
			Glyma.13g262500	-0.377825806	-0.282562926	-0.396179763	0.22316974
			Glyma.14g098100	0.61151808	0.319721461	0.325137896	0.141732521
			Glyma.15g242900	-0.441231989	-0.274402193	-0.083977509	0.581363443
			Glyma.17g226600	0.512760952	0.067624137	-0.245108767	-0.043904224
			Glyma.19g156900	0.167690666	0.195043293	-0.247836957	0.474091621
			Glyma.20g241500	0.171701615	0.173287941	-0.021638531	-0.146865432
			Glyma.20g241600	NA	NA	NA	NA
			Glyma.20g241700	NA	NA	NA	NA
6	2-hydroxyisolflavanone synthase (*IFS*)	EC 1.14.13.136	**Glyma.07g202300**	0.394926496	0.357576265	-0.213737722	-0.125261004

7	**2-hydroxyisoflavanone dehydratase (*HIDH*)**	**EC 4.2.1.105**	Glyma.01g239600	0.45498434	0.441863994	0.363985041	0.189537347
			Glyma.01g239300	-0.587795088	0.028255871	-0.495117711	0.299336568
			Glyma.01g239500	0.04362814	0.261281707	0.256266427	-0.043654924
			Glyma.02g134100	0.212187206	0.296153829	0.04249178	0.050824962
			Glyma.07g138600	0	0	0	0
			Glyma.10g250200	0.828485439	0.401219189	0.337363135	0.099898647
			Glyma.10g250300	1.654715254	0.513915639	1.788811416	0.265318982
			Glyma.11g004200	0.515533137	0.455335813	-0.288035169	-0.049947963
			Glyma.20g143400	-0.226579787	0.096032774	-0.0466707	-0.308290866
8	**2,7,4'-trihydroxyisoflavanone 4'-O methyltransferase (*HI4'OMT*) / isoflavone 4'-O-methyltransferase (*I4'OMT*)**	**EC 2.1.1.212**	**Glyma.13g173300**	3.639307457	2.790258977	4.725183617	1.485716045
			Glyma.13g173600	0	0	0	0
			Glyma.15g220700	0	-0.296488086	0	0
9	**Isoflavone-7-O-methyltransferase (*I7OMT*)**	**EC 2.1.1.150**	Glyma08g246900	0	0	0	0
			Glyma08g247000	0	0	0	0
			Glyma08g248000	0	0	0	0
			Glyma09g094400	2.510216153	3.283777264	1.434824647	0.145452921
			Glyma09g094600				
			Glyma10g176500	1.491005773	1.077376343	0.260983069	-0.371883152
			Glyma10g176600	1.262680835	4.406816585	-0.818762523	-1.969865562
			Glyma10g176700	0	0	0	0
			Glyma18g267500	-3.280734433	0	-1.701672432	-2.691520919
			Glyma18g267900	0	0	0	0
			Glyma18g269600	0.579311238	-0.002844237	0.178079815	-0.545726683
			Glyma20g213500	0.617631218	0.452551825	2.034387638	-0.497178677
			Glyma20g213600	0.59520022	0	0.261343825	0.519876644
			Glyma20g213700	0.559772071	0.696849053	0.05200705	-0.767080324
			Glyma20g213800	2.214621171	0	0.589736034	-0.546440414
10	**Isoflavone-7-O-glucosyltransferase (*IF7GT*)**	**EC 2.4.1.170**	Glyma.09g127200	2.380861333	2.30030051	2.66445123	-1.993555755
			Glyma.09g127300	-1.346789231	1.888599703	0	0
			Glyma.09g127700	-0.495261611	0.053638474	-0.668991606	-0.417294499
			Glyma.09g128200	0	0	0	-1.60879107
			Glyma.09g128300	1.126985801	-0.2978967	0.351663915	-0.78412246
			Glyma16g175200	NA	NA	NA	NA
			Glyma16g175300	NA	NA	NA	NA
			Glyma16g175400	NA	NA	NA	NA
			Glyma16g175500	NA	NA	NA	NA
			Glyma16g175600	-0.671227373	0.189138497	-1.514778209	-0.043764847
			Glyma16g175900	0.556770537	0	-0.960405965	-0.029586678
11	**Isoflavone 7-O-glucoside 6"-O-malonyl transferase (*IF7MAT*)**	**EC 23.1.115**	Glyma.08g247100	0	0.634451134	0	-0.379143178
			Glyma.08g247200	0	0.934890472	0	0.44426185
			Glyma.13g054000	0.817942474	-0.087111456	-1.145330723	-0.01524228
			Glyma13g056100	0.377946687	0.065409409	0.411492225	-0.029312916
			Glyma.18g258000	0.101739484	-0.156384203	-0.12798092	-0.075119615
			Glyma.18g268100	1.199388359	1.576879084	-0.167868736	-0.758364196
			Glyma.18g268200	0.950523365	0.430245752	0.063461619	-0.541198652
			Glyma.18g268300	0.486252999	0.104015893	-0.136137871	-0.496481122
			Glyma.18g268400	0.16297138	0.323482731	-0.160157179	-0.173612241
			Glyma.18g268500	0.227448859	0.425774538	-0.207162371	-0.579055061
			Glyma.18g268600	-0.126916256	0.123033037	-0.705600481	-0.398215441
			Glyma19g030500	0.121597825	-0.107871146	0.810175811	-0.215812186
			Glyma19g030700	NA	NA	NA	NA
			Glyma19g030800	NA	NA	NA	NA
12	**Isoflavone 2'-hydroxylase (*I2'H*)/Isoflavone-2-monooxygenase**	**EC 1.14.13.89**	Glyma.01g190400	0	0	0	0
			Glyma.02g067900	-2.112120972	-2.570141172	-2.59777092	0
			Glyma.08g089400	0	-0.562799503	1.767771429	0.103467212
			Glyma.08g089500	0	0	0	0
			Glyma.09g049200	1.401124572	1.344030375	0.815410619	0.577357367
			Glyma.09g049600	0	0	0	0
			Glyma.11g093100	0	0.654483964	-0.031303803	0.917368963
			Glyma.15g156100	1.103542316	1.996449527	1.669378088	1.596477492
			Glyma.16g149300	0.056669469	0.209527571	-0.196791292	-0.184479436

13	**2'-hydroxyisoflavone reductase (*IFR*)**	**EC 1.3.1.45**	Glyma.01g101300	0	0	0	0
			Glyma.01g172600	0.802452896	0.74905464	0.476472187	-0.036447849
			Glyma.01g172900	0.354457243	-0.468019197	0.619859482	0.134170805
			Glyma.01g211800	1.439472004	0.976746404	0.539430501	-0.181452898
			Glyma.11g070200	0.639973513	0.397901128	-0.195458212	-0.256910329
			Glyma.11g070500	0.994405171	0.531598483	0.095532456	-0.179751713
			Glyma.11g070600	1.142301481	0.697425028	0.470752862	-0.182522972
			Glyma.11g095400	0.942922465	0.397808669	-0.848469137	-0.645092066
			Glyma.11g116900	-1.86249011	1.39866987	0.273910748	0.987075867
			Glyma.12g042700	0	0	0	0
			Glyma.20g027800	-0.063944132	-0.426157305	0.157962041	0.338347364
14	**3,9-dihydroxypterocarpan 6a-monooxygenase (*D6aH*)**	**EC 1.14.14.93**	Glyma01g109000	0	0	0	0
			Glyma02g176200	-1.858786827	-1.974746839	-2.622124433	-2.011104331
			Glyma03g142000	-0.987214313	1.79022871	-3.547259878	-2.13125837
			Glyma03g142100	0.648042473	1.277374567	1.783547816	0.22774307
			Glyma03g143700	1.592045001	1.333506217	2.085594425	1.16200504
			Glyma10g092500	-0.125287605	0.477436499	-0.366073794	-0.783410111
			Glyma10g092600	-4.798836587	-2.969277604	-1.293714451	-1.575286962
			Glyma.12g173900	0	0	0	0
			Glyma.12g239100	0.360089349	-0.804701128	-0.724179529	0.115208461
			Glyma.17g162300	0	0	0	0
			Glyma.19g144600	-0.391163792	0.042286186	-0.991786801	-0.865871046
			Glyma.19g144700	0.529986361	0.6977946219	0.312457468	-0.448961097
			Glyma19g146800	1.269807653	0.299823764	0.157361241	0.20594812
15	**Trihydroxy pterocarpan dimethyl allyl transferase**	**EC 2.5.1.36**	Glyma12g239100	0.920369396	0.560798373	-0.724179529	0.115208461
			Glyma17g162300	-0.038109085	0	0	0
			Glyma19g144600	0	0	-0.991786801	-0.865871046
			Glyma19g144700	0	2.299186563	0.312457468	-0.448961097
			Glyma19g146800	3.464864003	0.972378068	0.157361241	0.20594812
			Glyma10g070200	0.888041888	0.453034596	0.77872955	0.383004387
			Glyma10g070300	1.443857877	0.511215219	-0.923056164	-1.498787646
			Glyma10g295300	2.099200904	1.054681467	1.636987728	1.011144498
			Glyma11g210300	1.003978201	0.64598842	0.391131216	-0.648910154
			Glyma11g210400	2.173354256	-0.310670945	0	0
			Glyma11g210500	-0.072362525	1.416836051	-0.596476195	-0.423879707
			Glyma20g245100	1.189259912	0.898930026	0.997576416	0.76942809
			Glyma20g245200	1.412888885	2.555429484	3.487700391	2.548883084
16	**Glyceollin synthase**	**EC 1.14.13.85**	Glyma.01g061100	-1.807271066	-0.40687036	-1.388207632	-0.677558655
			Glyma.02g119600	-4.837385828	-1.191088433	-1.441097802	-0.382584856
			Glyma.05g021800	-0.597887631	-0.530993274	-0.40865174	-0.287751515
			Glyma.05g021900	0.264445936	1.648784877	-2.133854028	0.363278597
			Glyma.05g022000	0	0	0	0
			Glyma.05g022100	1.283704731	1.71321262	1.483524845	0.409508004
			Glyma.05g147100	-1.154413129	-0.247985608	-0.901182388	-0.996315008
			Glyma.06g202300	1.894383317	2.206224945	1.310173868	-0.036897992
			Glyma.06g202400	0	0	0	0
			Glyma.07g039800	0.727815276	0.25197494	1.122931273	-0.077693757
			Glyma.07g052300	-2.032622604	-0.983738288	-1.375530204	-1.973672058
			Glyma.08g104100	-1.048925089	-0.258349194	-1.109614295	-1.482082128
			Glyma.09g186100	0	0	0	0
			Glyma.16g008600	1.349804	0.601254228	2.031587159	0.538821398
			Glyma.16g021200	-1.202110395	-0.40198467	-0.376434269	0.379707194
			Glyma.17g038700	-0.273452505	-0.208963566	0.228800098	0.302256271
			Glyma.17g077600	0	0	0	0
			Glyma.17g077700	-1.313514759	-0.387494596	-1.298338197	-0.71209945
			Glyma.19g240800	-1.121997335	-1.338127443	-0.45546341	-0.93258946
			Glyma.19g258700	-0.304769199	0.255936522	-0.434217473	0.521513141

*The bold highlighted ones were selected for the quantitative real-time PCR analysis to check their transcript abundance among S54 and S67 genotypes at 3, 5, and 8 dpi after SCN inoculation

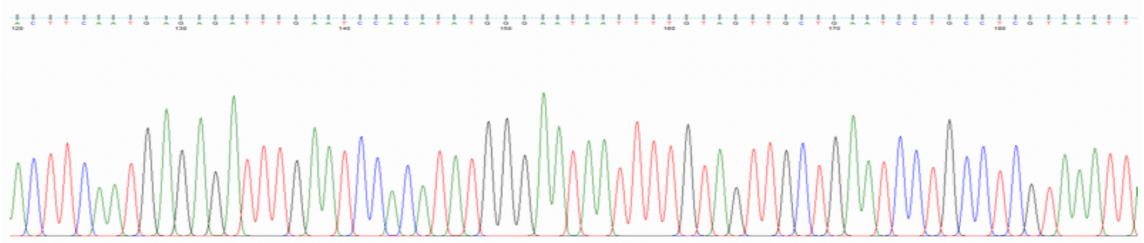


Figure 2.3.7.1a: Peak qualities of sequenced nucleotides

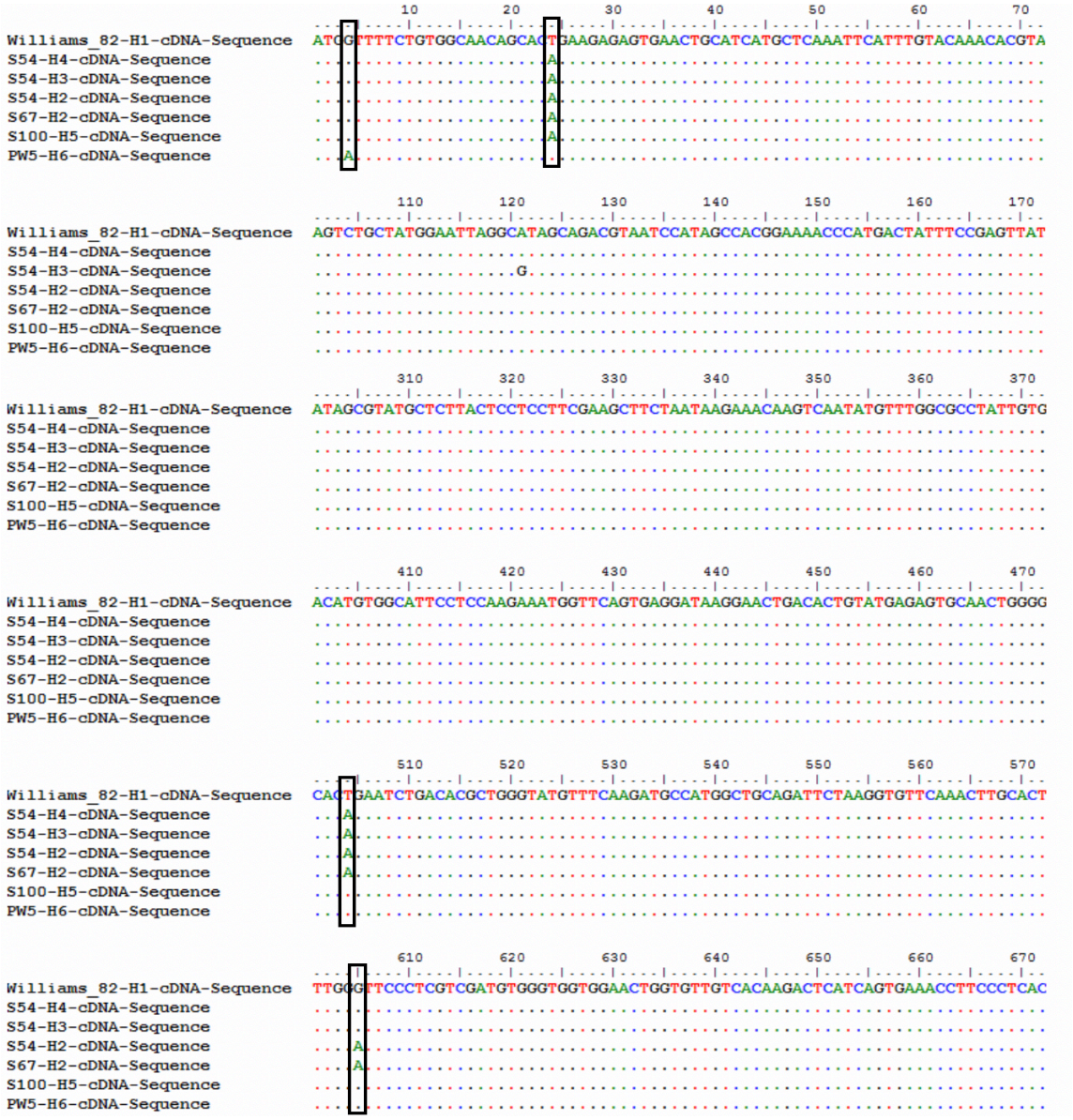


Figure 2.3.7.1b: cDNA sequences alignment and comparison among different wild soybean and soybean genotype

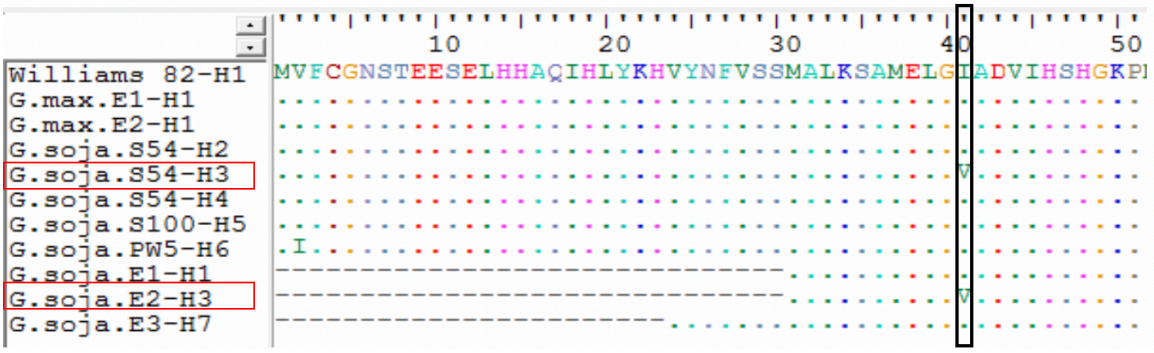


Figure 4.3.1: Amino acid sequence alignment and comparison of *HI4'OMT*/*I4'OMT* gene among different wild soybean and soybean genotype cloned in the current study and downloaded from NCBI database

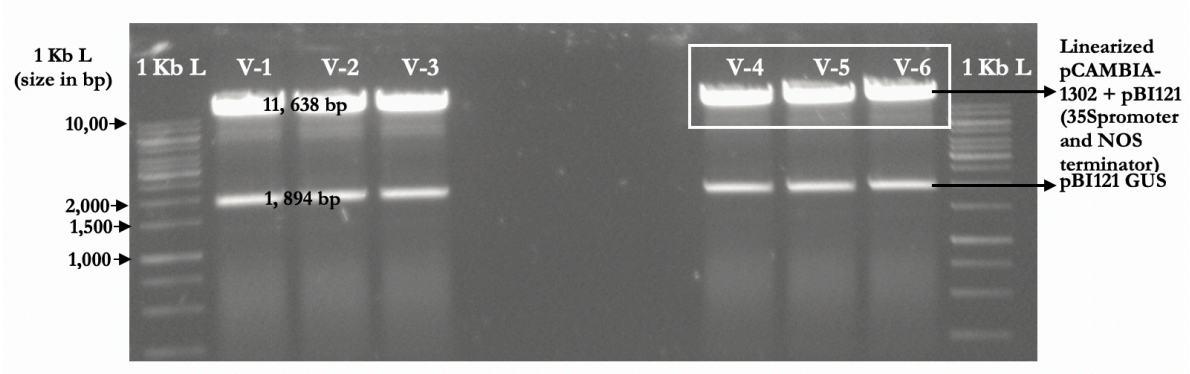
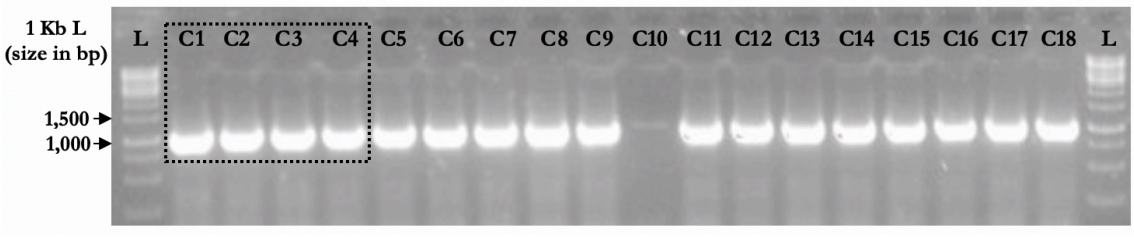
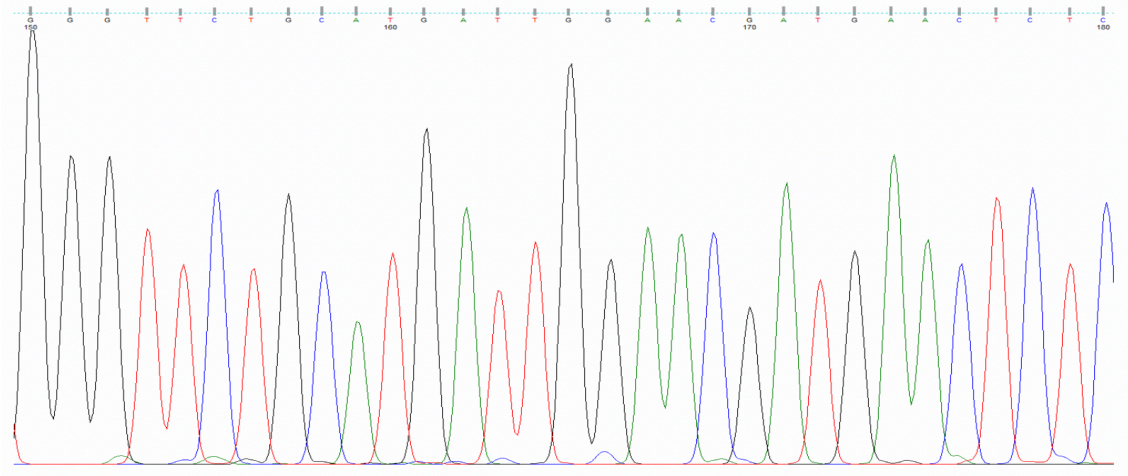


Figure 2.4.4.1: (a) Validation and purification of linearized binary vector without *GUS* and validation of Haplotype-3 (*H3*) allele inserted into the place of *GUS* within pCAMBIA-1302-*GFP*-35S::*GUS*/NOS binary vector by **(b)** PCR and **(c-d)** by sequencing

a) The expected size of linearized pCAMBIA+pBI121 (including 35Spromoter and NOS terminator) was 11,638 bp and the pBI121-*GUS* was 1,894 bp. The highlighted linearized pCAMBIA-1302+pBI121 were gel extracted purified for *H1* to *H4* gene ligation and expression vector construction



b) The expected size of amplified PCR product was 1,098 bp and the highlighted positive clones (C1-C4) were used for plasmid extraction and sequencing (figure 2.4.4.1c-d)



c) Peaks of the sequenced clone(s)

CCGATCTAGTAACATAGATGACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATATTTTGTTTTCTATCGCGTATT
AAATGTATAATTGCGGGACTCTAATCATAAAAACCCATCTCATAAATAACGTCATGCATTACATGTTAATTATTACATG
CTTAACGTAATTCAACAGAAATTATATGATAATCATCGCAAGACCGGCAACAGGATTCAATCTTAAGAAACTTTATTGC
CAAAGTTTGAAC**GATCGGGGAAATTCGAGCTC (*SacI* restriction site)**
TTAAGGATAAACTTCAATGAGAGATTTGAATCCACATATGGGAATAATTTTGTAGTTGCTGAATCCTGCCTCGTAAATT
AGTTTCTCCCACTCTTTCTTTTCTCTTTCTTTTCCATTAAACATAGTCAACATCACTAAATCATAGTCTAGCTTCAATTCA
GTCATTTCGCGATCATCACCTACTTCATCAATTGCTATGTCTATAATTATAACCTTTCCTTCTTTCCCTTTCCCTGAAATA
GCTTCTTTGCAGTTCTTCAATATCTTCACAGAGAGTTCATCGTTCCAATCATGCAGAACCCACTTGAGTAAAACTGCATC
GGCAGAGGGGATGGACTTGAACATATCTCCACCAACAAAGTTCAAATTTTCGTTTCCAGTTAAATTAGCCACAACCTGT
Reverse primer
GGTTGGTCAAACA**CAGTGCATTTCAAGTGAGGGA**AGGTTTCACTGATGAGTCTTGTGACAACACCAGTTCCACCACC
Forward primer (reverse complementary)
CACATCGACGAGGGAACCCAA**TCCCTCGAACACATGCTTGC**ATTCCTCGAGTGCAAGTTTGAACACCTTAGAATCTG
CAGCCATGGCATCTTGAAACATACCCAGCGTGTCAGATTCTGTGGTTTTGTTGAGAAAGTCCCAGAAACTCTCCCCAGT
TGCACTCTCATACAGTGTCAGTTCCTTATCCTCACTGAACCATTTCTTGGAGGAATGCCACATGTCAAGGGAGCTCGAA
TGAAGAGCTCCCTTCACAATAGGCGCCAAACATATTGACTTGTTTCTTATTAGAAGCTTCGAAGGAGGAGTAAGAGCAT
ACGCTATTTCTTCTCCTCCTTCTACACCGTTCTTTGACGGCAATATTGTTTTTGCAAAGAACCCATTGTGCGTTAAAAGG
CGGAGAAAGCGTTGGAGGACACTCACTTTTGAAGGGTGCAATTTGAGAGCTGAGGATAACTCGGAAATAGTCATGGGT
TTTCCGTGGCTATGGATTACGTCTGCTACGCCTAATTCCATAGCAGACTTAAGAGCCATGGAACTCACAAAGTTGTATA
CGTGTTTGTACAAATGAATTTGAGCATGATGCAGTTCACTCTCTTCTGTGCTGTTGCCACAGAAAACCAT
AAGGGACTGACCACCCGG**GGATCCTCTAGAGTCC (*BamHI* restriction site)**
CCCGTGTTCTCTCCAAATGAAATGAACTTCCTTATATAGAGGAAGGGTCTTGCGAAGGATAGTGGGATTGTGCGTCATC
CCTTACGTCAGTGGAGATATCACATCAATCCACT

d) Checked sequences indicating forward and reverse primers, with underlined *H3* gene sequences

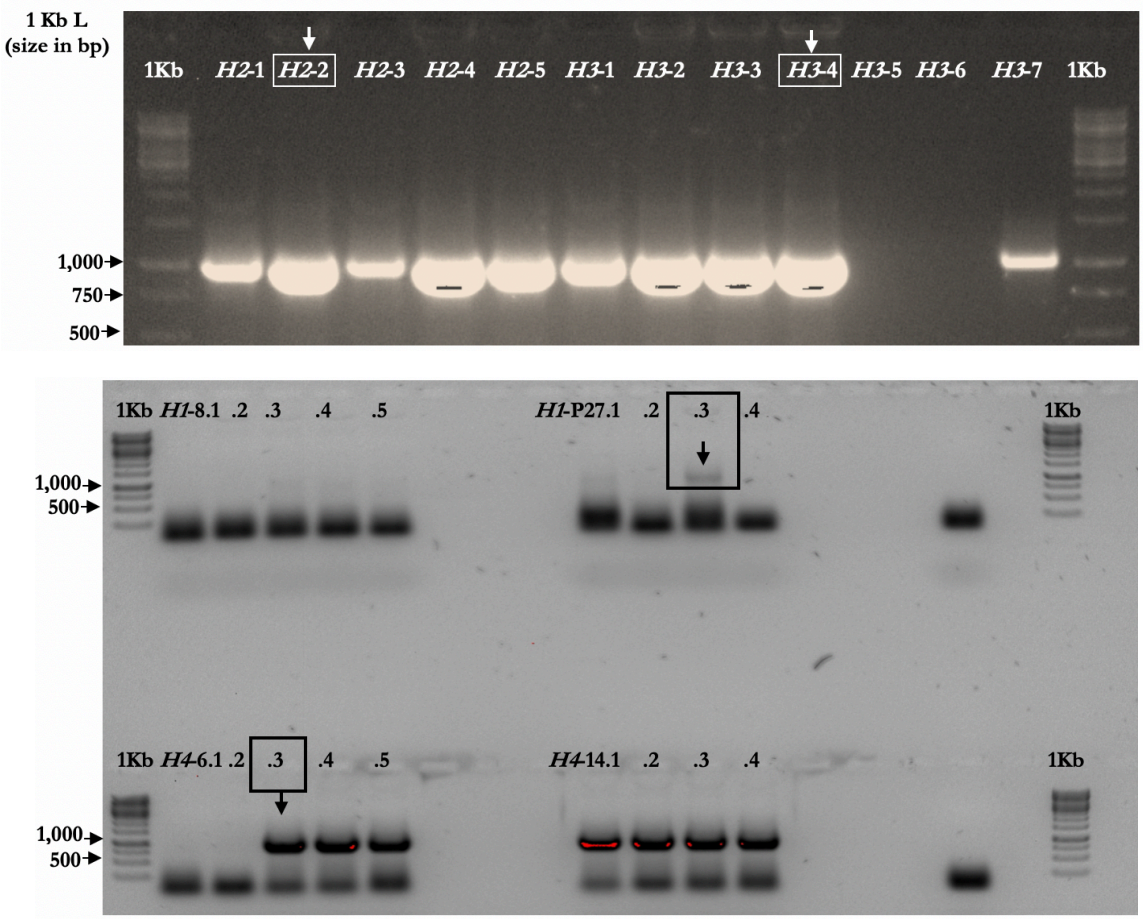


Figure 2.4.5.1: Validation of pCAMBIA-1302-GFP-35S::*H1*-*H4*/NOS expression vector after transforming in to the *Agrobacterium rhizogenes* K599 competent cells

The expected size of amplified PCR product was 1,098 bp and the highlighted positive clone (*H1*-P27.3, *H2*-2, *H3*-4 and *H4*-6.3) were used for sub-culturing and bacterial culture preparation

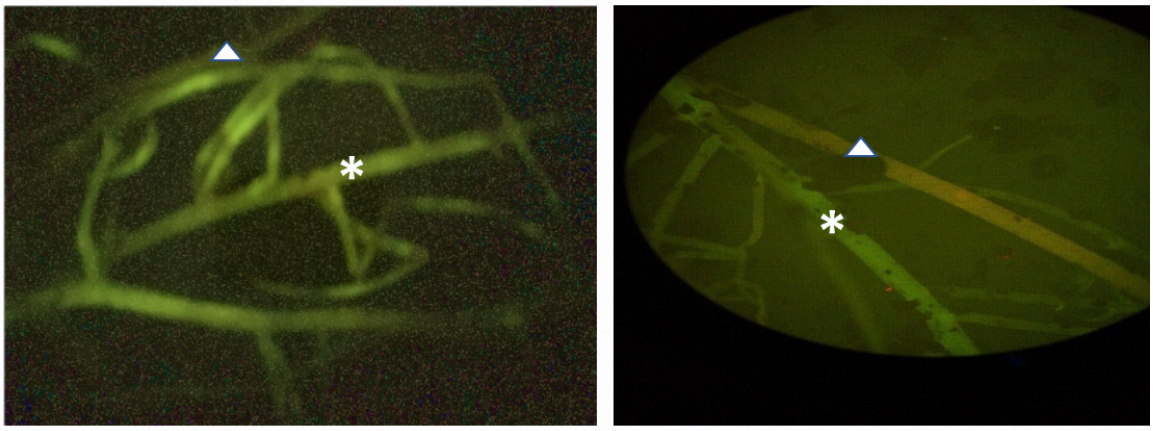
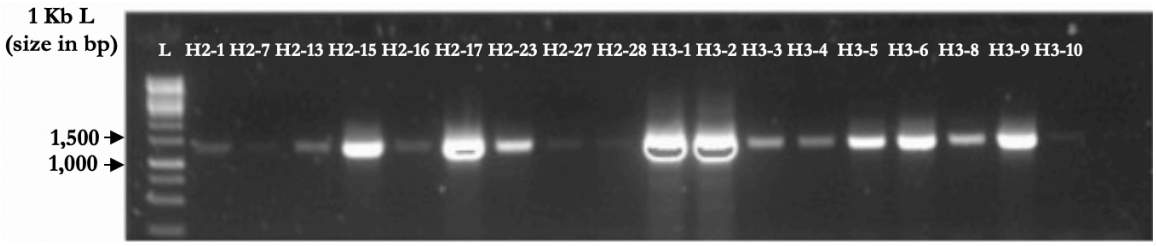
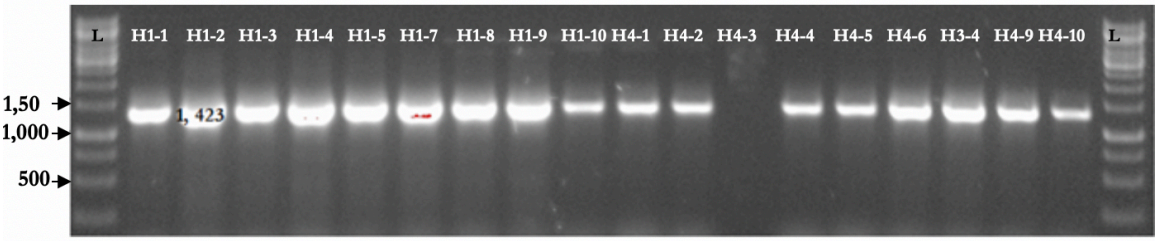


Figure 2.4.6.1: Validation of pCAMBIA 1302-*GFP*, *GUS* control vector and Haplotype-1 to Haplotype-4 (*H1-H4*) allele pCAMBIA-1302-GFP-35S::*H1-H4*/NOS expression vector containing transgenic roots by **(a)** *GFP* visualization and **(b-d)** by PCR

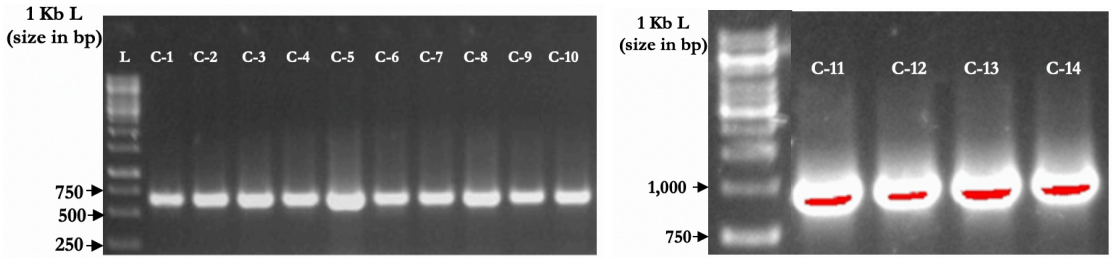
a) *GFP* positive transgenic hairy root (labeled with an asterisk) were separated with *GFP* negative transgenic hairy root (labeled with a triangle) under GFP filter



b) PCR amplification of pCAMBIA-1302-GFP-35S::*H2* &*H3*/NOS transgenic roots and the expected size of amplified PCR product was 1,423 bp (325 bp from vector + 1,098 bp gene size)



c) PCR amplification of pCAMBIA-1302-GFP-35S::*H1* & *H4*/NOS transgenic roots and the expected size of amplified PCR product was 1,423 bp (325 bp from vector + 1,098 bp gene size)



d) PCR amplification of pCAMBIA-1302-*GFP* and pCAMBIA-1302-*GFP*-35S::*GUS*/NOS transgenic roots and the expected size of amplified PCR product was 695 bp (325 bp from vector + 370 bp gene size) for *GUS* control (C-1 to C-10) and 1,036 bp (325 bp from vector + 711 bp gene size) for *GFP* control (C-11 to C-14)

Table 3.4.3.1: SCN performance in transgenic hairy roots. Counted total number of nematodes and nematodes at J2, beyond J2 (J3 + J4 + adult) stages with resistance index in individual transgenic hairy roots with '*VC* (*GFP* and *GUS*)' and '*HI4'OMT/I4'OMT-H1-H4'* gene constructs; transgenic hairy roots were infected with second stage juvenile (J2) nematodes, **(a) SCN2 and, (b) SCN5 races.**

S. No.	Transgenic hairy root	Total nematodes at beyond J2 (J3+J4+adult) stage	Total nematodes at J2 stage	Total nematodes	*Resistance Index
a.1. SCN2 infected transgenic hairy roots with *GFP* gene (negative vector control (*VC*))					
1.	CR-*GFP*-1	267	153	420	0.6358
2.	CR-*GFP*-3	178	105	283	0.6290
3.	CR-*GFP*-4	418	265	683	0.6120
Average:		**287.67**	**174.34**	**462**	**0.62560**
SEM (±)		**57.195**	**38.704**	**95.828**	**0.00578**
a.2. SCN2 infected transgenic hairy roots with *GUS* gene (positive vector control (*VC*))					
1.	CR-*GUS*-2	211	131	342	0.617
2.	CR-*GUS*-3	148	69	217	0.682
3.	CR-*GUS*-4	286	215	501	0.571
Average:		**215**	**138.34**	**353.34**	**0.62333**
SEM (±)		**32.567**	**34.542**	**67.099**	**0.02629**
a.3. SCN2 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H1* candidate gene/allele					
1.	*H1*-1	183	191	309	0.593
2.	*H1*-2	254	179	433	0.587
3.	*H1*-3	315	204	519	0.606
Average:		**250.666**	**191.333**	**420.33**	**0.59533**
SEM (±)		**31.142**	**5.894**	**49.766**	**0.00457**
a.4. SCN2 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H2* candidate gene/allele					
1.	*H2*-1	205	140	345	0.594
2.	*H2*-2	283	202	485	0.583
3.	*H2*-3	248	185	433	0.573
Average:		**245.333**	**175.66**	**421**	**0.58333**
SEM (±)		**18.416**	**15.102**	**33.359**	**0.00495**
a.5. SCN2 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H3* candidate gene/allele					
1.	*H3*-1	121	150	271	0.447
2.	*H3*-2	110	156	266	0.414
3.	*H3*-3	110	151	261	0.422
Average:		**113.666**	**152.333**	**266**	**0.42766**
SEM (±)		**2.993**	**1.515**	**2.357**	**0.00811**
a.6. SCN2 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H4* candidate gene/allele					
1.	*H4*-1	205	140	345	0.595
2.	*H4*-2	283	202	485	0.584
3.	*H4*-3	248	185	433	0.573
Average:		**245.333**	**175.666**	**421**	**0.5840**
SEM (±)		**18.416**	**15.102**	**33.359**	**0.00518**

S. No.	Transgenic hairy root	Total nematodes at beyond J2 (J3+J4+adult) stage	Total nematodes at J2 stage	Total nematodes	*Resistance Index
		b.1. SCN5 infected transgenic hairy roots with *GFP* gene (negative vector control (*VC*))			
1.	CR-*GFP*-2	211	125	336	0.6279
2.	CR-*GFP*-4	148	84	232	0.6379
3.	CR-*GFP*-5	323	193	516	0.6259
Average:		**227.33**	**134**	**361.33**	**0.63062**
SEM (±)		**41.783**	**25.952**	**67.733**	**0.00301**
		b.2. SCN5 infected transgenic hairy roots with *GUS* gene (positive vector control (*VC*))			
1.	CR-*GUS*-1	201	133	334	0.6017
2.	CR-*GUS*-2	276	155	431	0.6403
3.	CR-*GUS*-3	317	208	525	0.6038
Average:		**264.666**	**165.333**	**430**	**0.61532**
SEM (±)		**27.730**	**18.174**	**45.020**	**0.01023**
		b.3. SCN5 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H1* candidate gene/allele			
1.	*H1*-3	299	192	491	0.6089
2.	*H1*-4	284	187	471	0.6029
3.	*H1*-5	232	159	391	0.5933
Average:		**271.666**	**179.333**	**451**	**0.60176**
SEM (±)		**16.575**	**8.384**	**24.944**	**0.00371**
		b.4. SCN5 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H2* candidate gene/allele			
1.	*H2*-1	263	210	473	0.5560
2.	*H2*-2	317	251	568	0.5580
3.	*H2*-5	255	176	431	0.5916
Average:		**278.333**	**212.333**	**490.666**	**0.56859**
SEM (±)		**15.897**	**17.703**	**33.0868**	**0.00942**
		b.5. SCN5 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H3* candidate gene/allele			
1.	*H3*-1	166	239	405	0.4098
2.	*H3*-3	97	149	246	0.3943
3.	*H3*-4	128	201	329	0.3890
Average:		**130.333**	**196.333**	**326.666**	**0.39774**
SEM (±)		**16.291**	**21.298**	**37.488**	**0.00510**
		b.6. SCN5 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H4* candidate gene/allele			
1.	*H4*-2	303	235	538	0.5631
2.	*H4*-3	260	197	457	0.5689
3.	*H4*-4	236	179	415	0.5686
Average:		**266.333**	**203.666**	**470**	**0.5669**
SEM (±)		**16.002**	**13.477**	**29.473**	**0.00152**

*Resistance Index = total # of nematodes developed beyond J2 (J3 + J4 + adult) stage/total # of nematodes;

Table 3.4.3.1: SCN performance in transgenic hairy roots. Counted total number of cysts in individual transgenic hairy roots with '*VC* (*GFP* and *GUS*) and '*HI4'OMT/I4'OMT-H1-H4'* gene/allele constructs; transgenic hairy roots were infected with second stage juvenile (J2) nematodes, **(c) SCN2 and, (d) SCN5 races respectively.**

S. No.	Transgenic hairy root	Total number of female cysts	Transgenic hairy root	Total number of female cysts
	c.1. SCN2 infected transgenic hairy roots with *GFP* gene (negative vector control (*VC*))		**c.2. SCN2 infected transgenic hairy roots with *GUS* gene (positive vector control (*VC*))**	
1.	CR-*GFP*-1	121	CR-*GUS*-1	107
2.	CR-*GFP*-2	131	CR-*GUS*-2	117
3.	CR-*GFP*-3	94	CR-*GUS*-4	110
4.	CR-*GFP*-5	100	CR-*GUS*-7	110
5.	CR-*GFP*-6	93	CR-*GUS*-8	166
6.	CR-*GFP*-8	110	*CR-*GUS*-9	100
7.	**CR-*GFP*-12	280	**CR-*GUS*-11	195
8.	CR-*GFP*-13	93	CR-*GUS*-13	147
9.	CR-*GFP*-14	217	CR-*GUS*-14	115
10.	CR-*GFP*-15	103	CR-*GUS*-15	133
11.	*CR-*GFP*-17	89	CR-*GUS*-16	180
12.	CR-*GFP*-19	114	CR-*GUS*-18	124
Average		**128.75**	**Average**	**133.67**
SEM (±)		**16.272**	**SEM (±)**	**8.66**

S. No.	Transgenic hairy root	Total number of female cysts	Transgenic hairy root	Total number of female cysts
	c.3. SCN2 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H1* candidate gene/allele		**c.4. SCN2 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H2* candidate gene/allele**	
1.	*H1*-R-1	142	*H2*-R-1	140
2.	*H1*-R-2	103	**H2*-R-2	179
3.	**H1*-R-3	196	*H2*-R-3	81
4.	*H1*-R-4	145	*H2*-R-7	173
5.	*H1*-R-5	64	*H2*-R-8	151
6.	*H1*-R-6	149	*H2*-R-9	96
7.	*H1*-R-12	61	*H2*-R-10	90
8.	**H1*-R-14	45	*H2*-R-13	96
9.	*H1*-R-15	119	*H2*-R-16	111
10.	*H1*-R-16	80	*H2*-R-17	71
11.	*H1*-R-17	83	*H2*-R-19	136
12.	*H1*-R-18	161	**H2*-R-20	29
Average		**112.34**	**Average**	**112.75**
SEM (±)		**12.96**	**SEM (±)**	**12.25**

S. No.	Transgenic hairy root	Total number of female cysts	Transgenic hairy root	Total number of female cysts
	c.5. SCN2 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H3* candidate gene/allele		**c.6. SCN2 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H4* candidate gene/allele**	
1.	*H3*-R-2	42	*H4*-R-2	124
2.	*H3*-R-4	63	*H4*-R-4	133
3.	*H3*-R-5	66	*H4*-R-5	90
4.	*H3*-R-6	45	*H4*-R-6	94
5.	*H3*-R-7	60	*H4*-R-7	184
6.	**H3*-R-8	79	*H4*-R-8	109
7.	*H3*-R-10	61	**H4*-R-9	86
8.	**H3*-R-11	35	*H4*-R-10	135
9.	*H3*-R-12	39	*H4*-R-11	99
10.	*H3*-R-14	40	*H4*-R-13	101
11.	*H3*-R-19	53	**H4*-R-14	193
12.	*H3*-R-20	61	*H4*-R-16	108
Average		**53.67**	**Average**	**121.34**
SEM (±)		**3.88**	**SEM (±)**	**10.15**

*Transgenic hairy roots with minimum number of developed cysts;

**Transgenic hairy roots with maximum number of developed cysts;

S. No.	Transgenic hairy root	Total number of female cysts
d.1. SCN5 infected transgenic hairy roots with *GFP* gene (negative vector control (*VC*))		
1.	CR-*GFP*-27	121
2.	CR-*GFP*-28	131
3.	CR-*GFP*-30	94
4.	CR-*GFP*-31	100
5.	CR-*GFP*-32	93
6.	CR-*GFP*-33	110
7.	**CR-*GFP*-34	280
8.	CR-*GFP*-35	93
9.	CR-*GFP*-36	217
10.	CR-*GFP*-37	103
11.	*CR-*GFP*-38	89
12.	CR-*GFP*-39	114
Average		**134.42**
SEM (±)		**11.138**

Transgenic hairy root	Total number of female cysts
d.2. SCN5 infected transgenic hairy roots with *GUS* gene (positive vector control (*VC*))	
CR-*GUS*-24	109
CR-*GUS*-25	149
CR-*GUS*-26	133
CR-*GUS*-27	110
CR-*GUS*-28	117
CR-*GUS*-29	132
**CR-*GUS*-31	190
CR-*GUS*-33	180
CR-*GUS*-34	135
CR-*GUS*-35	123
CR-*GUS*-36	178
*CR-*GUS*-38	94
Average	**137.5**
SEM (±)	**8.532**

S. No.	Transgenic hairy root	Total number of female cysts
d.3. SCN5 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H1* candidate gene/allele		
1.	*H1*-R-25	92
2.	*H1*-R-26	133
3.	*H1*-R-30	176
4.	**H1*-R-32	195
5.	*H1*-R-33	74
6.	*H1*-R-34	154
7.	*H1*-R-35	161
8.	*H1*-R-36	145
9.	*H1*-R-37	69
10.	*H1*-R-38	89
11.	*H1*-R-39	183
12.	**H1*-R-40	61
Average		**127.67**
SEM (±)		**13.332**

Transgenic hairy root	Total number of female cysts
d.4. SCN5 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H2* candidate gene/allele	
H2-R-28	90
H2-R-29	160
H2-R-30	98
H2-R-31	128
H2-R-32	145
**H2*-R-34	169
H2-R-35	89
H2-R-36	160
H2-R-37	113
**H2*-R-38	77
H2-R-39	120
H2-R-40	129
Average	**123.17**
SEM (±)	**8.536**

S. No.	Transgenic hairy root	Total number of female cysts
d.5. SCN5 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H3* candidate gene/allele		
1.	*H3*-R-26	32
2.	*H3*-R-27	25
3.	*H3*-R-28	28
4.	*H3*-R-29	49
5.	*H3*-R-33	66
6.	*H3*-R-34	27
7.	**H3*-R-35	21
8.	*H3*-R-36	38
9.	*H3*-R-37	49
10.	*H3*-R-38	42
11.	*H3*-R-39	58
12.	**H3*-R-40	67
Average		**41.84**
SEM (±)		**4.623**

Transgenic hairy root	Total number of female cysts
d.6. SCN5 infected transgenic hairy roots with *HI4'OMT/I4'OMT-H4* candidate gene/allele	
H4-R-25	104
H4-R-26	143
**H4*-R-27	190
H4-R-28	96
**H4*-R-29	78
H4-R-30	100
H4-R-33	96
H4-R-34	115
H4-R-35	79
H4-R-36	90
H4-R-37	113
H4-R-40	158
Average	**113.5**
SEM (±)	**9.767**

*Transgenic hairy roots with minimum number of developed cysts;

**Transgenic hairy roots with maximum number of developed cysts;