**Supplementary file 1**

**Transcriptome Profiling of induced susceptibility effects on soybean- soybean aphid interaction**

Surendra Neupane1, Adam J. Varenhorst2, and Madhav P. Nepal1\*

1Department of Biology and Microbiology, South Dakota State University, Brookings, SD, USA 57007

surendra.neupane@sdstate.edu (S.N.); madhav.nepal@sdstate.edu (M.P.N.)
2Department of Agronomy, Horticulture and Plant Science, South Dakota State University, Brookings, SD, USA 57007

adam.varenhorst@sdstate.edu (A.J.V.)

\*Correspondence: madhav.nepal@sdstate.edu; Tel.: +1-605-688-5971

**Material and Methods**

**Plant material and aphid colonies**

Two genotypes of soybean were used: susceptible soybean cultivar was LD12-15838R and the resistant cultivar was LD12-15813Ra. The resistant cultivar contains the *Rag1* gene. These genotypes were infested with two SBA populations defined by their inability (biotype 1) or ability (biotype 2) to colonize soybean with *Rag1* resistance [1]. The biotype 1 and biotype 2 populations originated from a colony maintained by Iowa State University (Ames, IA). Among them, colonies of biotype 1 originated from field populations in Ohio and were maintained in a colony at the Ohio State University. At South Dakota State University, SBA colonies were maintained using susceptible cultivar SD01-76R for biotype 1 and resistant cultivar LD12-15813Ra for biotype 2. The aphid populations used in this study were randomly selected removing the leaves from the soybean plants used for maintaining the colonies.

**Induced susceptibility experiment**

To characterize induced susceptibility effects, randomized complete block design (RCBD) greenhouse experiment was conducted using eight treatments, three replications (plants) in three blocks (nine experimental units per treatment). We followed the treatments as explained by the procedure by [2]. The initial feeding population of SBA was termed as an inducer population and the subsequent feeding population of SBA was termed as a response population. Three seeds of LD12-15838R and LD12-15813Ra were planted into damp soil (Professional Growing Mix, Sun Gro Horticulture, MA, USA) in each pot of dimension of 10.1 cm by 8.89 cm (500 ml; Belden Plastics, MN, USA). Pots were placed onto plastic flats. The soybean plants were watered filling the flats when top soil began to dry. The plants were thinned down to one plant per pot upon reaching the V1 developmental growth stage. Per methods previously developed methods, at the V2 growth stage, soybean plants (Day 0) were infested with either no aphids (untreated control), 50 biotype 2 SBA that were placedon the ventral side of a middle leaf of first trifoliate using a fine tip paintbrush [2]. The infested trifoliate was covered with a small no-see-um mesh net (Quest Outfitters, Sarasota, FL) and secured with the paper clip and tangle trap to confine within the first trifoliate of the plants. After 24 hrs (Day 1), leaves from second trifoliate were collected from one replication set of each block and snap frozen in the liquid nitrogen. After sample collection, response population of 15 biotype 1 SBA (avirulent)were added upon the middle leaf of second trifoliate (except on sampled and control plants). The whole plants were covered with the large no-see-um mesh net (Quest Outfitters, Sarasota, FL) to confine movement of aphids between the plants. The response population was allowed to move freely about the plant with the exception of first trifoliate. This ensures the spatial isolation of inducer and response populations. The response populations were counted on each plant to confirm the colonization by the response populations on day 5. On day 11, the response population of aphids was counted and the day 11 leaf samples from the one replication sets of each block were collected and snap frozen in the liquid nitrogen. The samples were kept at -80 ºC for further analysis. The greenhouse conditions were maintained approximately 24-25 °C and a 16-hour photo period (16 light: 8 dark). An overview representing experimental methods used for the experiment is shown in Figure S1.

**RNA Extraction, Library construction, and sequencing**

RNA was extracted from the leaf samples that were collected at day 1 and day 11 from resistant and susceptible cultivars that were non infested, infested with no inducer: biotype 1 response, and biotype 2 inducer: biotype 1 response. Briefly, leaf samples from each treatment were ground in liquid nitrogen with mortar and pestle to a fine powder followed by their processing for total RNA extraction using PureLink RNA mini kit (Invitrogen, USA). RNA samples were treated with TURBOTM DNase (Invitrogen, USA) to remove any DNA contamination following the manufacturer’s instructions. Assessment of the isolated RNA integrity was performed by 1% agarose gel electrophoresis, and RNA concentration was measured by Nanodrop 2000 (Thermo Fisher Scientific, USA). Three replicates from these treatments in resistant and susceptible cultivars were pooled in equimolar concentration. The cDNA libraries were constructed and sequenced at South Dakota State University Sequencing Facilities. RNAseq library construction was prepared using Illumina’s TruSeq Stranded mRNA Kit v1 (San Diego, CA). The libraries were quantified by QuBit dsDNA HS Assay (Life Technologies, Carlsbad, CA). The libraries were sequenced on an Illumina NextSeq 500 using a NextSeq 500/550 High Output Reagent Cartridge v2 (San Diego, CA) at 75 cycles. Fastq files were generated and demultiplexed on Illumina’s BaseSpace cloud network (San Diego, CA).

**RNA-Seq analysis**

Quality control of reads was assessed using FastQC program (version 0.11.3) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [3]. The FastQC results were visualized using MultiQC v1.3 [4]. Low quality bases (QC value < 20) and adapters were removed by trimming using the program Trimmomatic (version 0.36) [5] (options: PE -phred33 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 HEADCROP:8 MINLEN:30). High-quality single-end reads were mapped against the primary coding sequences of *G. max*. The coding sequences (*Gmax: Gmax\_275\_Wm82.a2.v1.transcript\_primaryTranscriptOnly.fa.gz*)wereobtained from the Phytozome database and aligned using Salmon ver.0.9.1 [6] accessed from Bioconda [7]. A flow chart showing the RNA-seq data analysis pipeline is shown in Figure S2. The downstream analyses were done using iDEP 0.82 [8]. The read quants were filtered with 0.5 counts per million (CPM) in at least one sample. The quantified raw reads were transformed using regularized log (rlog) which is implemented in the DESeq2 package [9] (Data file 2). The transformed data were subjected to exploratory data analysis such as hierarchical clustering (Figure S3; Data file 3) and correlation between samples (Figure S4).

**References**

1. Kim K-S, Hill CB, Hartman GL, Mian M, Diers BW. Discovery of soybean aphid biotypes. Crop Sci. 2008, 48(3):923-928. doi:10.2135/cropsci2007.08.0447

2. Varenhorst A, McCarville M, O’Neal M. An Induced Susceptibility Response in Soybean Promotes Avirulent *Aphis glycines* (Hemiptera: Aphididae) Populations on Resistant Soybean. Environ Entomol. 2015;22;44(3):658-67.doi: 10.1093/ee/nvv05

3. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

4. Ewels P, Magnusson M, Lundin S, Käller M: MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-3048. doi: 10.1093/bioinformatics/btw354

5. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30(15):2114-2120. doi: 10.1093/bioinformatics/btu170

6. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417. https://doi.org/10.1038/nmeth.4197

7. Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, Valieris R, Köster J, Bioconda T. Bioconda: sustainable and comprehensive software distribution for the life sciences. Nat Methods. 2018;15(7):475. https://doi.org/10.1038/s41592-018-0046-7

8. Ge SX, Son EW, Yao R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. BMC Bioinformatics. 2018;19(1):534. https://doi.org/10.1038/s41592-018-0046-79.

9. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol Evol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8