Transcriptomics reveals multiple resistance mechanisms against cotton leaf curl disease in a naturally immune cotton species, *Gossypium arboreum*

Rubab Zahra Naqvi1,2,3, Syed Shan-e-Ali Zaidi1,2,3,8, Khalid Pervaiz Akhtar4, Susan Strickler3, Melkamu Woldemariam3, Bharat Mishra3, M. Shahid Mukhtar5, Brian E. Scheffler6, Jodi A. Scheffler7, Georg Jander3, Lukas A. Mueller3, Muhammad Asif5 & Shahid Mansoor1

Cotton leaf curl disease (CLCuD), caused by cotton leaf curl viruses (CLCuVs), is among the most devastating diseases in cotton. While the widely cultivated cotton species *Gossypium hirsutum* is generally susceptible, the diploid species *G. arboreum* is a natural source for resistance against CLCuD. However, the influence of CLCuD on the *G. arboreum* transcriptome and the interaction of CLCuD with *G. arboreum* remains to be elucidated. Here we have used a RNA-Seq based study to analyze differential gene expression in *G. arboreum* under CLCuD infestation. *G. arboreum* plants were infested by graft inoculation using a CLCuD infected scion of *G. hirsutum*. CLCuD infested asymptomatic and symptomatic plants were analyzed with RNA-seq using an Illumina HiSeq. 2500. Data analysis revealed 1062 differentially expressed genes (DEGs) in *G. arboreum*. We selected 17 genes for qPCR to validate RNA-Seq data. We identified several genes involved in disease resistance and pathogen defense. Furthermore, a weighted gene co-expression network was constructed from the RNA-Seq dataset that indicated 50 hub genes, most of which are involved in transport processes and might have a role in the defense response of *G. arboreum* against CLCuD. This fundamental study will improve the understanding of virus-host interaction and identification of important genes involved in *G. arboreum* tolerance against CLCuD.

Cotton (*Gossypium spp.*) is a leading source of natural textile fiber, oil and protein meal, making it one of the most important cash crops in many countries1. The major cotton producing countries include China, India, United States, and Pakistan which contribute more than 70% of the total cotton production around the globe (https://apps.fas.usda.gov/). Among the 50 *Gossypium* species, two diploid species (*G. herbaceum* and *G. arboreum*) and two tetraploid species (*G. hirsutum* and *G. barbadense*) are cultivated worldwide, but *G. hirsutum* (upland cotton) accounts for more than 90% of overall cotton production2.

Tetraploid cotton, *G. hirsutum*, evolved about 1–1.5 million years ago (Mya) as a result of a polyploidization event with an A-genome species *G. arboreum* and a D-genome species *G. raimondii* the putative progenitors3. *G. hirsutum* cultivars have superior lint yield and produce very high-quality fiber, however, these tetraploid cultivars are often more susceptible to several abiotic and biotic stresses4,5. In contrast, diploid cotton species, particularly

---

1Agricultural Biotechnology Division, National Institute for Biotechnology and Genetic Engineering (NIBGE), Jhang Road, Faisalabad, Punjab, Pakistan. 2Pakistan Institute of Engineering & Applied Sciences (PIEAS), Nilore, Islamabad, Pakistan. 3Boye Thompson Institute, 533 Tower Road, Cornell University, Ithaca, NY, USA. 4Nuclear Institute for Agriculture & Biology (NIAB), Jhang Road, Faisalabad, Punjab, Pakistan. 5Department of Biology, University of Alabama at Birmingham, Birmingham, AL, USA. 6Genomics and Bioinformatics Research Unit (USDA-ARS), Stoneville, MS, USA. 7Crop Genetics Research Unit, United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Stoneville, MS, USA. 8Present address: AgroBioChem Department, Gembloux Agro-Bio Tech, University of Liège, 5030, Gembloux, Belgium. Rubab Zahra Naqvi and Syed Shan-e-Ali Zaidi contributed equally to this work. Correspondence and requests for materials should be addressed to S.M. (email: shahidmansoor7@gmail.com)
*G. arboreum*, are more tolerant to many biotic and abiotic stresses including CLCuD. *G. arboreum* is a valuable source for novel genes to enhance genetic diversity as well as for specific traits.

*Bemisia tabaci*, commonly known as silverleaf whitefly, is one of the sap sucking pests that causes damage to more than 500 plant species including cotton. Among the list of sap-sucking insects that cause damage to a cotton crop, whitefly is responsible for more than 50% crop losses. Whitefly serves as a vector for viruses causing cotton leaf curl disease (CLCuD), which is a major threat to cotton crops in several countries including Pakistan and India. Viruses causing CLCuD, collectively referred to as cotton leaf curl viruses (CLCuVs), belong to the family **Geminiviridae** and the genus **Begomovirus**, which is the most important genera of this family. Other members of this genus also infect important crops including tomato, beans, chillies, cucurbits, mungbean and other vegetables. Plants infected with CLCuD show typical begomovirus infection symptoms such as leaf curling, vein yellowing, leaf enations and stunted growth. CLCuD is among the most devastating of viral diseases, responsible for serious crop losses annually that negatively impact a country's economy.

As their defense system, plants mobilize a variety of intrinsic mechanisms to deal with pathogens and it is necessary to understand the mechanisms that plants use to cope with different stresses at the genetic level. Abiotic and biotic stresses activate a relay network of plant gene expression mechanisms that lead to the reprogramming of a variety of physiological and metabolic processes in accordance with the stress response. Early studies mainly used model plant species to identify a wide spectrum of genes that are involved in different levels of metabolism, signal transduction, osmotic regulation and stress response. Microarray gene expression profiling has been widely used to study differential gene expression in plants under different stresses. Advances in the field of transcriptomics has made RNA-Sequencing (RNA-Seq) a major "-omics" technology for studying global gene expression of plants under different stress conditions. Several RNA-Seq studies have been done to reveal mechanisms in cotton under different biotic stresses. Transcriptional analysis of cotton boll weevil highlighted the diversity of genes and pathways regulated by pest infestation, such as kinase cascades, transcription factors (TFs), and phytohormone-signaling pathways. Significant changes in the expression of transcripts were found to be associated with sugar and amino acid metabolism in cotton following aphid and whitefly infestation. Another recent transcriptomic study reported several genes for control of phloem-feeding pests.

Viral infection initiates a complex interaction between the virus and the host. Understanding host responses during viral infection can help in the development of effective strategies for virus control. In recent years, RNA-Seq studies have been applied extensively to uncover the responses of plant hosts to viral infection. However, the comprehensive molecular mechanisms underlying CLCuD-cotton interactions remain poorly defined.

Therefore, to elucidate the mechanisms involved in cotton defense against CLCuD, we aimed to identify important differentially expressed genes in cotton under CLCuD infestation and to elucidate the resistance mechanisms of the naturally immune cotton species *G. arboreum* at a molecular level. We infested *G. arboreum* with CLCuD by grafting, then performed RNA-Seq followed by data analysis, weighted gene co-expression network analysis (WGCNA), gene ontology enrichment and functional annotation.

**Results and Discussion**

Cotton is an economically important crop and CLCuD causes drastic losses in crop production. Compared to normally susceptible tetraploid *G. hirsutum*, the diploid species *G. arboreum* is a natural source for resistance against CLCuD. Typically, whitefly is the vector that transmits CLCuD in tetraploid cotton, but *G. arboreum* remains unaffected when infected with viruliferous whiteflies. Recent studies have revealed evidence that *G. arboreum* plants support CLCuD replication and its long-distance spread only if CLCuD is transmitted by graft inoculation. In the presence of CLCuV/CLCuMB*on*, mild symptoms of CLCuD have been shown to be induced in *G. arboreum* by its graft inoculation with scions from CLCuD infested *G. hirsutum* plants. Interestingly, only a few leaves of *G. arboreum* have been observed to develop symptoms of the disease following grafting. The nature of this resistance pattern remains unknown and might be correlated with gene expression levels or other indigenous immune mechanisms of *G. arboreum*. The present study aims to detect changes in *G. arboreum* gene expression in response to CLCuD infestation by graft inoculation, even in the absence of severe symptoms. This study will help in understanding the mechanisms involved in resistance of *G. arboreum* to the begomoviruses causing CLCuD.

**Grafting and transcriptome sequencing.** The standard method for CLCuD, delivery into *G. hirsutum* is by whitefly-mediated virus transmission. However, in most studies, *G. arboreum* remained asymptomatic and thought to be a non-host against CLCuD. As an alternative, a graft infestation method was developed for systemic delivery of the CLCuD complex (cotton leaf curl geminivirus with its associated betasatellite) into *G. arboreum*. In this study, we used the graft-infestation method for delivery of CLCuD complex to *G. arboreum*. Characteristic viral infection symptoms were observed on *G. arboreum* grafted with CLCuD-infected *G. hirsutum* scions (Fig. 1). Out of 30 plants grafted with scions, 27 plants established a successful graft. Among these 27 *G. arboreum* plants, 20 plants showed characteristic symptoms of CLCuD on the leaves near the grafted scion, while 7 plants remained asymptomatic for CLCuD (Fig. 1). These results are comparable with previous observations where most of the graft-infested *G. arboreum* developed mild symptoms with a few plants remaining asymptomatic. RNA was extracted from three biological replicates of graft-infested, symptomatic and asymptomatic *G. arboreum* plants. This RNA was processed for library preparation followed by RNA sequencing.

**RNA-Seq data analysis.** All six libraries were sequenced using a HiSeq™ 2500 platform. On average, 10 million total reads were obtained from each replicate (Table 1). Quality of individual sequences was evaluated using FastQC analysis, including per base sequence quality analysis which plots the Phred quality score distribution on the y axis for each read generated per sample and for each nucleotide base call plotted on the x axis.
Figure 1. Experimental design, methodology and RNA-Seq pipeline used in study. Panel A and B show the graphical representation of grafting experiments (A) scion of cotton leaf curl disease (CLCuD) infected *Gossypium hirsutum* was used on a rootstock of *G. arboreum*; asymptomatic leaves of CLCuD were collected for RNA extraction and sequencing. (B) Leaves showing symptoms of CLCuD were collected for RNA extraction and sequencing. (C) Shows an asymptomatic CLCuD free leaf of *G. arboreum*. (D) Shows a leaf with very mild symptoms of CLCuD including vein swelling and darkening, highlighted with black arrows. (E) Shows the workflow of the RNA-Seq experiment and the tools used at each step.

(Figure S1). All FASTQ sequencing files obtained in this study have an average per base Phred score of 64, a conventional threshold denoting high quality base calls. High quality sequence reads were aligned to the *G. arboreum* reference genome with 13 chromosomes and 41,330 gene models (Table S1). The percentage of mapped reads obtained from the alignment summary of HISAT2 is provided in Table 1 (Fig. 2A).

Differential gene expression was quantified and grafted CLCuD symptomatic *G. arboreum* (RZN7-RZN9) were compared to the grafted asymptomatic *G. arboreum* (RZN10-RZN12). To assess the biological reproducibility of the experiment, each sample was treated as an individual biological replicate using the cuffdiff package of
Transcription factors and R-gene mediated response. During pathogen attack, transcription factors play an important role in plant innate immunity. MYB transcription factors act as master regulators of cellular defense response. Based on these studies, the upregulation of the boron transporter gene in *G. arboreum* may also be involved in host-viral interaction which modulates the R-gene mediated response in the host to combat virus infection.

### Differential expression of membrane transporters and channel proteins.

Among the DEGs, some transporter genes were observed such as a boron transporter gene that was upregulated while a SWEET 17 gene was downregulated in *G. arboreum* asymptomatic plants. An interaction of host boron transporter with viral coat protein can regulate boron transport to induce necrosis, or the interaction can modulate an R gene mediated defense response. Based on these studies, the upregulation of the boron transporter gene in *G. arboreum* depicts the protection of the plant from boron toxicity. In *G. arboreum*, the boron transporter may also be involved in host-viral interaction which modulates the R-gene mediated response in the host to combat virus infection. SWEETs are bidirectional vacuolar fructose transporters in plants predicted to be involved in maintaining sugar homeostasis in plant organs during favorable as well as abiotic and biotic stress conditions. SWEET transporters are found to be associated with plant pathogen interactions during a pathogen attack in plants. A reduced level of SWEET gene expression in tobacco and rice was associated with enhanced resistance to pathogens, suggesting a role in pathogen growth and plant disease resistance. The downregulation of *G. arboreum* SWEET transporter suggests its role in sugar homeostasis under unfavorable conditions and possible involvement in reduced pathogen growth and disease resistance.

Alteration in expression levels of channel proteins in plants has been reported as a response to environmental stimuli. Significantly enhanced soybean mosaic virus resistance was observed in soybean by the overexpression of GmAKT2. The induction of aquaporin channel proteins encoding genes has been reported with expression profiling of soybean under *Pseudomonas syringae* infection. Studies have also shown the involvement of aquaporins in plant-viral interactions. For example, aquaporins interact with a cucumber mosaic virus (CMV) replication protein that potentially affects CMV replication in the host plant. Thus, the upregulation of *G. arboreum* AKT2 and aquaporin genes might have a role in CLCuD.

### Putative defense related genes involved in resistance of *G. arboreum* to CLCuD infection.

Plants have evolved with sophisticated and well-established defense mechanisms to cope with pathogens, such as insects, fungi, bacteria and viruses. Besides physical barriers (like trichomes), toxic compounds or secondary metabolites, plants have numerous defense pathways to defend themselves against diverse pathogens. An initial gene ontology (GO) term analysis of the top one hundred most significantly expressed genes present in the RNA-Seq dataset indicated that several GO terms were involved in regulation of responses related to stress, external stimuli, biotic stress and plant defense. Further dissection of expression levels of individual genes in symptomatic and asymptomatic *G. arboreum* plants revealed that many groups of genes were differentially expressed in response to CLCuD infection.

### Table 1. Summary of RNA-Seq runs used in this study.

| Sample       | Condition                  | No. of Reads | % Mapping |
|--------------|---------------------------|--------------|-----------|
| RZN7         | Grafted - CLCuD symptomatic | 11304210     | 76.25     |
| RZN8         | Grafted - CLCuD symptomatic | 983891       | 66.97     |
| RZN9         | Grafted - CLCuD symptomatic | 10201621     | 62.17     |
| RZN10        | Grafted - asymptomatic     | 10049099     | 69.77     |
| RZN11        | Grafted - asymptomatic     | 10941016     | 68.2      |
| RZN12        | Grafted - asymptomatic     | 9667217      | 77.69     |

Cufflinks. We compared individual runs from each condition and calculated the transcript abundances of each independent run (Fig. 2B). A quality check of compared transcripts provided standard gene density and dispersion (Fig. 2C,D). The expression levels of mapped genes were normalized with a FPKM (fragments per kilobase of exon per million fragments mapped) value. To confirm the quality of RNA-Seq, the eight highest-ranking genes were selected to evaluate gene expression. Based on comparisons of our samples, none of these reference genes were significantly differentially expressed (Table 2), suggesting that the sequences obtained and the transcript expression levels qualified for further transcriptome analysis. In the comparison of gene expression levels of symptomatic and asymptomatic *G. arboreum* plants, an absolute value of log2 fold change >1 and the False Discovery Rate (FDR) <0.05 was set to declare differentially expressed genes (DEGs) involved in the response of CLCuD infestation. Overall, 1062 genes were differentially expressed in this comparison, out of which 563 and 499 were up and downregulated, respectively (Fig. 2E; Fig. 3).

The transcriptome data of all replicates from this experiment is publicly available under BioProject accession number PRJNA380937 and a comprehensive list of up- and downregulated DEGs is provided in Table S2. This data is a valuable resource for understanding CLCuD tolerance in *G. arboreum*. Furthermore, the DEGs identified were narrowed down by log2 fold-change, p values and q values for further experimentation. Many functional classes of genes related to pathogen defense were found to be significantly affected in CLCuD-infected symptomatic plants compared to that of the asymptomatic plants.

For the validation of our transcriptomic data we selected 17 significant DEGs with a probable role in disease resistance. We designed primers and performed qPCR on cDNA of independent biological replicates of *G. arboreum*. Our qPCR data strongly correlated with the RNA-Seq expression data (Fig. 4). The implication of these genes has been discussed with reference to published data (Supplementary Discussion; Table 3).
responses and are involved in plant development, secondary metabolism, hormone signal transduction, abiotic stress tolerance and disease resistance. These transcription factors have been found to be involved in mechanisms of disease resistance in several plants through regulation of defense genes. In the present data set, differential gene expression of different MYB transcription factors, including RADIALIS and REVEILLE, indicate their role in plant defense against CLCuD. A recent study on genome-wide analysis of bHLH transcription factors in Solanum lycopersicum shows that bHLH transcription factors are involved in the plant’s defense under infection by Tomato yellow leaf curl virus and upregulation of bHLH is associated with disease resistance. G. arboreum bHLH transcription factor UNE10 was upregulated during CLCuD stress and suggests a role in disease resistance. Transcription factors, particularly ethylene response factor (ERF), lead to the transcriptional regulation of several jasmonate and ethylene responsive defense genes under pathogen attack. In our experiment, seven DEGs were found to be ethylene responsive genes (Table S5) that might have a role in G. arboreum defense against CLCuD.

R gene-triggered resistance in plants is associated with a rapid defense response known as hypersensitive response (HR) that can bring localized cell death at the site of infection and trigger a series of downstream defense pathways. R-genes have been found to be induced in response to several pathogens and diseases in plants, including tobacco mosaic virus, rice blast, Arabidopsis downy mildew, tomato leaf mold and Verticillium wilt-resistance in cotton. In our experiment, eleven putative R genes were differentially expressed indicating their role in plant defense against CLCuD (Table S6). Protein kinases along with R genes are well known in disease resistance both in a positive and a negative manner. Forty-six kinase genes were identified as DEGs and most of them were downregulated in asymptomatic G. arboreum (Table S7).

Table 2. Expression levels of housekeeping control genes in G. arboreum.
Phytohormone signaling. CLCuD infestation in *G. arboreum* also revealed the induced expression of genes associated with phytohormone signaling pathways. Phytohormones are the key players in signaling cascades induced by pathogens. In this study auxin, cytokinin, abscisic acid and brassinosteroid related genes were upregulated while genes involved in ethylene and salicylic acid pathways were downregulated, highlighting their role in plant defense. Taken together the complex cascade in phytohormone related gene expression demonstrates the interconnection of the signaling pathways regulated by phytohormones in response to CLCuD infection.

Figure 3. Hierarchical clustering of differentially expressed genes. A heat map of *G. arboreum* differentially expressed genes in response to cotton leaf curl disease with respect to hierarchical clustering. Log10 expression values were used for the analysis and negative values were set to zero. Clustering and the heat map were performed using heatmap 2.0 package in R.
Figure 4. Validation of gene expression with quantitative RT-PCR. (A) Heat map of 17 selected DEGs for qRT-PCR in G. arboreum in response to cotton leaf curl disease with respect to hierarchical clustering. Log10 expression values were used for the analysis and negative values were set to zero. Clustering and the heat map were performed using heatmap 2.0 package in R. (B) Quantitative RT-PCR was used to measure the relative expression levels of seventeen pathogen resistance related genes with 18 S as an internal reference. Values were expressed as fold changes of transcript levels in the CLCuD infested symptomatic leaf samples with respect to the transcript levels in CLCuD infested asymptomatic leaf samples. Error bars represented standard error (SE) of three biological replicates.
as potential translocation factors and are involved in regulation of viral cell-to-cell movement. Deposition of β−1,3-glucan or callose in plasmodesmata initiates a hypersensitivity response (HR) in hosts, consequently leading to blocking of plasmodesmata to hinder cell-to-cell viral movement. Viral proteins like the TGB2 protein of begomoviruses, interacts with host plant SUMO-conjugating enzyme 1 (SCE1), which is an important component of the CUL1-based SCF protein complex. SKP1 is another potent component of the CUL1-based SCF protein complex.

The ubiquitin proteasome system (UPS) has a very important role in virus host interactions and plant defense against viruses. There are several reports showing the complex interaction of host plant UPS components with viral proteins, suggesting that the ubiquitin pathway is probably a conserved pathway involved in plant virus interaction. UPS in plants is comprised of ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin-ligase (E3). These three enzymes make an E3 ubiquitin ligase complex which is required for polyubiquitination of cellular proteins, followed by 26 S proteasome targeted degradation. During viral infection, the virus takes over the plant host’s UPS machinery regulating its gene expression to enhance pathogenesis, while in parallel the plant uses UPS as a strong layer of defense, mainly by the getting of viral proteins. Different components of UPS, from ubiquitin to 26 S proteasome have been found to be involved in plant defense mechanisms. Host UPS machinery is involved in targeted degradation of TMV and Turnip yellow mosaic virus movement proteins and decreased viral pathogenesis suggest a role for UPS in plant antiviral defense.

Like other plant viruses, geminiviruses also interact with the different components of host UPS. Rep, a replication protein of geminiviruses, interacts with host plant SUMO-conjugating enzyme 1 (SCE1), which is an essential host factor for viral infection in the plant. Altered expression of SCE1 leads to unsuccessful geminiviral replication and suppression of infection. SKP1 is another potent component of the CUL1-based SCF protein complex.

| Gene                                      | Probable role in defense                        | References |
|-------------------------------------------|-------------------------------------------------|------------|
| Probable zinc metallopeptidase EGY3       | Development and stress response                  | 94         |
| Defensin-like protein 1                   | Bacterial and fungal pathogens as well as herbivorous insects | 82–87     |
| Phytosulphokines 3                       | Pattern-triggered immunity against pathogens, Leucine Rich Repeat family (LRR) | 98–100    |
| Chaperone protein Dnaj                      | Pathogen defense, antiviral defense              | 101–103    |
| Transcription factor UNE10                | Antiviral defense                                | 96         |
| Protein REVEILLE 7                       | Plant growth, stress, pathogen response          | 104–105    |
| Small heat shock protein C2                | Antiviral and antibacterial stress response      | 59,106     |
| Transcription factor RADIALIS             | Myb encoding genes, plant defense response       | 43,107     |
| Putative disease resistance protein RGA3   | R-gene mediated pathogen and disease response    | 45,108,109 |
| Bidirectional sugar transporter SWEET17    | Abiotic stress tolerance, pathogenesis related protein | 39       |
| Protein ECERIFERUM 1                      | Biotic and abiotic stresses                      | 110        |
| Probable aquaporin PIP2-2                 | Biotic and abiotic stresses, plant immunity      | 111–113    |
| Potassium channel AKT2/3                  | Plant development, stress responses, antiviral defense | 41-42    |
| 50S ribosomal protein L39-3               | Pathogen and disease resistance, antiviral defense | 114,115  |
| Boron transporter 4                       | R-gene mediated viral defense                    | 39         |
| Vignain                                   | Plant immunity, pathogenesis and plant defense   | 116,117    |
| Nitrate reductase [NADH]                  | Pathogen signal-induced NO production            | 118–120    |

Table 3. Selected differentially expressed genes for qRT-PCR and their probable functions in plant pathogen defense.

Protein processing in endoplasmic reticulum and ubiquitin proteasome systems are involved in plant-viral interaction. Under virus infection, the plant’s cellular machinery is hijacked by viruses for DNA replication of viral genomes and translation of viral proteins. Like other pathogens including bacteria, phytopathogenic viruses activate endoplasmic reticulum (ER) stress signaling machinery. Viruses act as “pirates” and use host machinery for their spread across the plasmodesmata to neighboring cells using viral movement proteins (MP). The ER signaling mechanism has been found to be involved in cell to cell movement and spread of viruses through plasmodesmata. In Arabidopsis, a broad range of viruses induce gene expression of ER stress marker genes such as Calreticulins (CRTs). Calreticulin has been shown to interact with movement of viral proteins including MP of tobacco mosaic and turnip crinkle virus (TCV). CRT helps in redirecting virus movement through plasmodesmata which leads to symptom development, however, its overexpression leads to delayed cell to cell movement of virus. Many other well studied host proteins interact with viral MPs for regulation of their cell to cell movement including β−1,3-glucanase-interacting proteins. DnaJ (HSP 40) is an important co-chaperon of HSP70 and is important for its function. HSP70 is proposed to be associated with coat protein interaction, viral transportation and replication of many geminiviruses. Downregulation of HSP70 showed decreased viral load and reduced viral movement in plants infected with tomato yellow leaf curl virus (TYLCV). In G. arboreum symptomatic plants, the induction of many genes involved in protein processing in ER including CRT, lectin like gene OS9, glucanase interacting gene OST and heat shock proteins HSP70 and HSP 40 indicate the possible role of this pathway interacting with CLCuD proteins and helping viral cell to cell movement in the plant, making it susceptible to establishment of mild symptoms from graft inoculation with CLCuD. The downregulation of these genes in asymptomatic plants might have role in impairment of viral movement proteins, restricting viral movement in plasmodesmata (Figure S3).
ubiquitin E3 ligases that recruits proteins for polyubiquitination. Geminiviral C2 protein interferes with the different components of host UPS including SKP1[323]. Suppression of SKP1 in plants abolishes the downstream HR and resistance response in plants[36]. JCI from Cotton leaf curl Multan virus (CLCuMV) interacts with the host ubiquitin-conjugating (UBC) enzyme, SIUBC3, leading to symptom induction that is possibly due to downregulation of the UPS[38]. The upregulation of UPS pathway components including UBC3 (COP1), SCE1 (UBE21) and SKP1 in the G. arboreum asymptomatic plants indicates that ubiquitin mediated proteolysis could be a defense response against symptom development of CLCuD (Figure S4).

**Weighted co-expression gene network analysis (WGCNA) identifies two novel modules within the network.** With the availability of large-scale transcriptome datasets, co-expression network analysis allows identification of a cohort of genes with similar expression patterns in response to a given stimulus or physiological condition within a cell[365,366]. Thus, co-expression networks can identify a set of genes, which might participate in a common biological process. To determine CLCuD-responsive common gene signatures, we performed a weighted gene co-expression network analysis[367,368] on 468 selected DEGs. Using the WGCNA platform, we created a topological overlap mapping metric (TOM) plot, a measure of neighborhood proximity that calculates the similarity of gene expression between two nodes[369]. TOM also features hierarchical clustering dendrograms possessing a range of weighted correlations[370]. These analyses led us to generate an undirected weighted network with scale-free topology, a network with power-law degree distribution. This weighted co-expression network encompasses two different modules (Fig. 5A, B; Table S8) that are denoted with two different colors. The turquoise and blue modules comprise of 252 and 216 genes, respectively.

Centrality measures can reveal the most influential vertices in a network. To decipher the most important nodes within this co-expression network, we calculated, degree (number of connections of a node) and clustering coefficients (degree to which a node is connected in a neighborhood) as well as information centrality (the flow of information between any two nodes in a connected network)[366,367,368]. Degree distribution revealed a total of 50 hubs, nodes with ≥ 5 connections within the co-expression network (Fig. 5B; Table S8).

Out of these identified hub genes, co-expression network upregulated hub genes included Aquaporin TIP4-1, Nitrate/peptide (NRT1/PTR) transporter, Ammonium transporter 1 member, vacuolar calcium ion transporter, CGDGH iron-sulfur cluster binding protein, 3-isopropylmalate dehydratase and nitrate reductase, while downregulated genes included sugar transport protein 13, NAT3 (Nucleobase-acrosbate transporter 3 ion symporter) and zinc transporter genes. GO term analysis of these hub genes revealed their involvement in transport processes. Plants use a wide array of chemicals as weapons for their defense against pathogens. These chemicals or compounds can accumulate at high levels in the infected or attacked tissue[372]. Proteins involved in transport processes are vital for long distance transport of these defense compounds to the other tissues and for plant sustainability under viral attack[373]. The differentially expressed transporters identified here might have a role in transport of secondary metabolites and defense related compounds that further activate other mechanisms involved in plant defense responses under viral attack.

Moreover, we determined that the average degree of turquoise and blue modules was constant. We did not observe any change in the degree for blue and turquoise modules compared to the entire co-expression network or blue module. Intriguingly, we also found that the turquoise module exhibited a significantly heightened average clustering coefficient (Fig. 5C, D). Finally, our information centrality measure in the largest component with 92 nodes determined that the turquoise module displayed significantly increased information centrality compared to the entire network (Fig. 5E; Table S9). These data suggest that nodes in the network may transmit the information via DEGs faster.

**Oxidative stress related genes under CLCuD infection.** Light has a major role in photosynthesis and meeting energy demands of the plant under pathogen attack. It provides energy as well as elicits protection to plants against invading microbes[374]. We found GO enrichment of light and radiation based gene expression under a module after co-expression network analysis (Figure S5). Under plant microbe attack, one protective strategy in plants is to produce reactive oxygen species (ROS)[375]. Nicotinamide adenine dinucleotide (reduced form NADH) and cytochrome oxidase (COX) enzymes are involved in synthesis of glutathione S-transferase a major scavenger of ROS. However, another mechanism involved in generation of ROS under biotic stress includes light dependent ROS production from chloroplasts[376]. Therefore, upregulation of COX1, COX3, NADH and genes related to light response in asymptomatic G. arboreum plants suggests modulation of gene regulation to combat CLCuD by ROS scavenging.

**Role of primary metabolic pathways in plant immunity.** Cellular energy demands have been found to be increased under pathogen attacks in plants. Primary metabolic pathways along with secondary metabolic processes support these increased requirements of the plants. Secondary metabolites including terpenes and phenolics are required by plants for protection against microbial pathogens. However, plant secondary metabolic pathways create a large carbon flux by influencing and reconfiguring the primary metabolism for resistance responses[377,378]. We identified upregulation of genes related to the metabolic pathway specifically involved in steroid and lipid metabolism in asymptomatic plants (Figure S6). These genes include flavonoids, brassinosteroid, sphingolipids and terpenoids, suggesting a role in the defense response to CLCuD. Carbohydrates or sugars act as a fuel to boost energy levels in plants during plant microbe interaction. These are studied as potent players in coordinated plant metabolism with development as well as dealing with different stress responses[379]. The concept of “sweet immunity” has been established based on their importance in plant immunity. In the present dataset, several genes involved in carbohydrate or saccharide metabolism were found to be highly expressed including phosphoensolpyruvate carboxikinase, fructose-1-6-bisphosphatase and phosphoglucomutase (Figure S6). We detected several genes related to cation or meta ion channel activity by WGCNA (Figure S6). Ca2+ signaling

---

**References:**

[365]Centrality measures can reveal the most influential vertices in a network.

[366]To decipher the most important nodes within this co-expression network, we calculated, degree (number of connections of a node) and clustering coefficients.

[367]Degree distribution revealed a total of 50 hubs, nodes with ≥ 5 connections within the co-expression network.

[368]Out of these identified hub genes, co-expression network upregulated hub genes included Aquaporin TIP4-1, Nitrate/peptide (NRT1/PTR) transporter, Ammonium transporter 1 member, vacuolar calcium ion transporter, CGDGH iron-sulfur cluster binding protein, 3-isopropylmalate dehydratase and nitrate reductase, while downregulated genes included sugar transport protein 13, NAT3 (Nucleobase-acrosbate transporter 3 ion symporter) and zinc transporter genes.

[369]GO term analysis of these hub genes revealed their involvement in transport processes.

[370]Plants use a wide array of chemicals as weapons for their defense against pathogens. These chemicals or compounds can accumulate at high levels in the infected or attacked tissue.

[371]Proteins involved in transport processes are vital for long distance transport of these defense compounds to the other tissues and for plant sustainability under viral attack.

[372]The differentially expressed transporters identified here might have a role in transport of secondary metabolites and defense related compounds that further activate other mechanisms involved in plant defense responses under viral attack.

[373]Moreover, we determined that the average degree of turquoise and blue modules was constant.

[374]We found GO enrichment of light and radiation based gene expression under a module after co-expression network analysis.

[375]Under plant microbe attack, one protective strategy in plants is to produce reactive oxygen species (ROS).

[376]Nicotinamide adenine dinucleotide (reduced form NADH) and cytochrome oxidase (COX) enzymes are involved in synthesis of glutathione S-transferase a major scavenger of ROS.

[377]However, another mechanism involved in generation of ROS under biotic stress includes light dependent ROS production from chloroplasts.

[378]Therefore, upregulation of COX1, COX3, NADH and genes related to light response in asymptomatic G. arboreum plants suggests modulation of gene regulation to combat CLCuD by ROS scavenging.

[379]The concept of “sweet immunity” has been established based on their importance in plant immunity.

[380]In the present dataset, several genes involved in carbohydrate or saccharide metabolism were found to be highly expressed including phosphoensolpyruvate carboxikinase, fructose-1-6-bisphosphatase and phosphoglucomutase.
Figure 5. Weighted gene co-expression network construction and analysis for CLCuD-responsive cotton genes in CLCuD symptomatic and asymptomatic leaves. (A) Transcriptomic data from symptomatic and asymptomatic leaves was analyzed by WGCNA using 468 DEGs. Genes were clustered as per expression arrangements characterized by the dendrogram and topological overlap mapping metric (TOM) heat map. Each line of the dendrogram corresponds to a gene. Clusters of similarly-regulated genes are grouped as modules by corresponding color (blue and turquoise) with a threshold minimum module size of 70 genes. The intensity of pink coloring in heatmap specifies high strength and green as no strength of correlation between pairs of genes on a direct scale and features hierarchical clustering dendrograms possessing a range of weighted correlations. (B) Weighted network illustrates correlations (edges) among the nodes (genes) with a weighted correlation threshold of ≥0.85. The network is composed of 468 connections and 348 genes organized in two different modules. The node color corresponds to modules identified via WGCNA. Nodes with high connectivity (≥5 connections, hubs) among different modules are indicated with increased node size. (C) Highly connected nodes (hubs) within the co-expression network are shown. Average degree of turquoise and blue modules is not changed in the entire co-expression network (2.689655172). (D) Clustering coefficient (degree to which a node is connected in a neighborhood) for genes within each node is illustrated in a box plot. Turquoise module
Plant growth and virus inoculation by grafting. Plants of *G. arboreum*, variety Ravi, were grown in an insect-free glasshouse. *G. hirsutum*, variety CIM 496, plants were maintained in a separate glasshouse and were inoculated with CLCuD by allowing viruliferous whiteflies to feed on these plants. Severely CLCuD infected CIM496 branches were then used as scions to graft on Ravi plants. *G. arboreum* plants were grafted using a “bottle shoot” grafting method. In this technique, a graft-scion is placed in a tube filled with distilled water to support plant growth at the higher temperatures maintained for this experiment (Fig. 1). The water in the tube was refreshed daily till the graft union between root stock and scion was established. The tube of water was removed at day 9 after grafting. The temperature of the glasshouse was maintained between 38–45 °C for day time and 25–30 °C for night time. The leaf tissue was collected from CLCuD symptomatic and asymptomatic plants at day 25 post graft infestation. Three biological replicates were processed independently.

RNA extraction. Total RNA was extracted from CLCuD infested symptomatic and asymptomatic leaves using TRIzol reagent according to the manufacturer’s instructions (Invitrogen, Carlsbad, CA, USA). Three biological replicates from each sample were used for this experiment. The quality and quantity of RNA were assessed by electrophoresis on 1% agarose gels and by a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, USA). The integrity of RNA samples was examined using Bio-analyzer 2100 equipment (Agilent Technologies, Germany).

Library construction and RNA sequencing. The extracted total RNA samples with a concentration of 10 µg were used for cDNA synthesis and strand specific RNA-Seq libraries were constructed as described earlier.Poly (A) mRNA was isolated using oligo-dT beads (Qiagen, Hilden, Germany). The mRNA was broken into short fragments (~300 nt). First-strand cDNA was synthesized using random hexamer-primed reverse transcription. Second-strand cDNA was generated using RNase H and DNA polymerase I. The cDNA fragments were purified and washed for end repair and ligated to sequencing adapters. The cDNA fragments of suitable size were purified and enriched by PCR to obtain the final cDNA library. The integrity of each cDNA library was examined using Bio-analyzer 2100 equipment (Agilent Technologies, Germany). The cDNA libraries were then sequenced using single-end mode of HiSeq™ 2500 equipment (Illumina, San Diego, CA, USA).

RNA-Seq data analysis. Clean reads were selected after preprocessing with Trimmomatic to remove low-quality sequences (i.e., reads containing adaptor sequences, and reads with more than 5% unknown bases). After preprocessing the RNA-Seq data, the quality of reads was checked by FastQC. The cleaned reads were then mapped to the reference *G. arboreum* genome using HISAT2-build and HISAT2 aligner. Default HISAT2 parameters, which allow up to two mismatches and report up to 20 alignments for reads mapping at multiple positions, were used. The sequence alignment/map files generated by HISAT2 were used as the input to the software Cufflinks, which assembles the alignments in the sequence alignment/map file into transfrags. Cufflinks does this assembly independently of the existing gene annotations and constructs a minimum set of transcripts that best describes the RNA-Seq reads. The unit of measurement used by Cufflinks to estimate transcript abundance is FPKM (Fragments Per Kilobase of transcript per Million mapped reads). The Cufflinks statistical model probabilistically assigns reads to the assembled isoforms. Cuffdiff was used to find differentially expressed genes (DEGs). The read coverage of one gene was used to calculate the gene expression level, which was measured with
the fragment per kilobase of exon model per million mapped reads (FPKM) method. A q value cutoff of 0.05 was used to determine whether a gene had differential expression between samples.

**Quantitative real time PCR.** To verify the differential expression detected by the Illumina RNA-Seq data, quantitative real-time RT-PCR (qRT-PCR) was performed on a new set of symptomatic and asymptomatic samples. A set of 17 genes was chosen, including 10 upregulated and 7 downregulated genes (Table S4). Primers for qPCR were designed with the program Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/primer3) with default settings. All primer sequences are provided in Table S3. qRT-PCR was performed using a QUANTSTUDIO 6 flex qRT-PCR instrument and the light cycler fast start DNA Master SYBR Green 1 kit (Roche, Basel, Switzerland). Reactions were performed in triplicate, and contained 100 ng of cDNA, 0.5 μL of each primer (10 μM/μL), and 10 μL SYBR Green Master Mix in a final volume of 20 μL. The amplification reactions were performed under the following conditions: 95 °C for 5 min, followed by 40 cycles of 95 °C for 15 s, 55 °C for 20 s, and 72 °C for 30 s. Melting curve analysis, performed by increasing the temperature from 55 to 95 °C (0.5 °C per 10 s), and gel electrophoresis of the final product confirmed the presence of single amplicons. Relative fold differences for each sample in each experiment were calculated using the ΔΔ Ct method and the amplification of 18S rRNA was used as an internal control to normalize all the data. To corroborate the expression levels measured by RNA-Seq, the ratio of expression as measured by qRT-PCR was compared to the ratio of expression levels between samples using RNA-Seq.

**Construction of co-expression network and network analyses.** To construct a co-expression network, we processed transcripts with FPKM count ≥10 and removed all the outliers, which yielded a total number of 468 DEGs. We implemented the R-based Weighted Gene Co-expression Network Analysis (WGCNA)\(^\text{67,68}\) package to construct a co-expression network. The dendrogram was constructed using the cutreeDynamicTree algorithm\(^\text{68}\), with a threshold of minimum module size of 70 genes. A weighted correlation threshold of ≥0.85 was set that resulted in the identification of two modules representing the entire network. Python-based Network\(^\text{X}\)\(^\text{90}\) and Cytoscape v. 3.5.1 Plugins\(^\text{91}\) were utilized for network analyses including information centrality, degree and cluster coefficients. The weighted co-expression network was visualized using “Attribute Groups Layout”, a feature of Cytoscape v. 3.5.1\(^\text{91}\). Box plot was used to display the distribution of nodes within each module for information centrality, degree and cluster coefficients. Students t-test describes statistical significance.

**Gene ontology and functional annotation of DEGs.** To determine the main biological functions and pathways of the DEGs, all DEGs were mapped to terms in the Kyoto Encyclopedia of Genes and Genomes (KEGG: www.kegg.jp/kegg/kegg1.html) database\(^\text{89}\) using AgriGO and KOBAS 3.0 tools respectively. Functional analysis of DEGs was done by searching the gene IDs of DEGs in the Swisspro and TeEMBL blast alignment hits available on Cottongen database (https://www.cottongen.org/). Furthermore, the GO term analysis of the identified nodes and hub genes in WGCNA was performed to determine their possible roles in plant defense to CLCuD.

**Data availability.** RNA-Seq data from this study have been deposited at the NCBI under the BioProject accession No. PRJNA380937 and BioSample Nos. SAMN06628122, SAMN06628123, SAMN06628124, SAMN06628125, SAMN06628126 and SAMN06628127.

**References**

1. Lee, J. J., Woodward, A. W. & Chen, Z. J. Gene expression changes and early events in cotton fibre development. *Annals of botany* **100**, 1391–1401 (2007).

2. Chen, Z. J. *et al*. Toward sequencing cotton (*Gossypium*) genomes. *Plant physiology* **145**, 1303–1310 (2007).

3. Paterson, A. H. *et al*. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**, 423–427 (2012).

4. Maqbool, A. *et al*. *Gossypium arboreum* GHSP26 enhances drought tolerance in *Gossypium hirsutum*. *Biotechnology progress* **26**, 21–25 (2010).

5. Robinson, A. F. Reniform in *US Cotton*: When, Where, Why, and Some Remedies 1. *Annual review of phytopathology* **45**, 263–288 (2007).

6. Li, R. & Erpeding, J. E. Genetic diversity analysis of *Gossypium arboreum* germplasm accessions using genotyping-by-sequencing. *Genetica* **144**, 335–354 (2016).

7. Gong, Q. *et al*. Salicylic acid-related cotton (*Gossypium arboreum*) ribosomal protein GaRPL18 contributes to resistance to *Verticillium dahliae*. *BMC plant biology* **17**, 59 (2017).

8. Razaq, M., Abbas, G., Farooq, M., Aslam, M. & Athar, H.-U.-R. Effect of Insecticidal Application on Aphid Population, Photosynthetic Parameters and Yield Components of Late Sown Varieties of *Canola*, Brassica napus L. *Pakistan journal of zoology* **46**, 661–668 (2014).

9. Zubair, M., Zaidi, S. S., Shakir, S., Amin, I. & Mansoor, S. An insight into Cotton leaf curl Multan betasatellite, the most important component of cotton leaf curl disease complex. *Virus Research* **19**, 280 (2017).

10. Zaidi, S. S., Bridgon, R. W. & Mansoor, S. Engineering dual Begomovirus-Bemisia tabaci resistance in plants. *Trends in Plant Science* **22**, 6–8 (2017).

11. Zaidi, S. S., Martin, D. P., Amin, I., Farooq, M. & Mansoor, S. Tomato leaf curl New Delhi virus: a widespread bipartite begomovirus in the territory of monopartite begomoviruses. *Molecular Plant Pathology* **18**, 901–911 (2017).

12. Zaidi, S. S.-e.-A., Tashkandi, M., Mansoor, S. & Mahfouz, M. M. Engineering Plant Immunity: Using CRISPR/Cas9 to Generate Virus Resistance. *Frontiers in plant science* **7** (2016).

13. Zaidi, S. S. *et al*. Frequent occurrence of Tomato leaf curl New Delhi virus in tomato leaf curl disease affected cotton in Pakistan. *Plos one* **11**, e0155520 (2016).

14. Zubair, M. *et al*. Multiple begomoviruses found associated with cotton leaf curl disease in Pakistan in early 1990 are back in cultivated cotton. *Scientific reports* **7**, 680 (2017).

15. Gatehouse, J. A. Plant resistance towards insect herbivores: a dynamic interaction. *New phytologist* **156**, 145–169 (2002).

16. Park, W., Scheffler, B. E., Bauer, P. J. & Campbell, B. T. Genome-wide identification of differentially expressed genes under water deficit stress in upland cotton (*Gossypium hirsutum* L.). *BMC plant biology* **12**, 90 (2012).
56. Taguwa, S.
57. Gorovits, R. & Czosnek, H. The involvement of heat shock proteins in the establishment of Tomato yellow leaf curl virus infection.

54. Stange, C. Plant-virus interactions during the infective process. Cien. Inv. Agr. (in English) 33 (1): 1–18.

52. Kørner, C. J., Du, X., Vollmer, M. E. & Pajerowska-Mukhtar, K. M. Endoplasmic reticulum stress signaling in plant immunity—at
53. Chen, M.-H., Tian, G.-W. , Gafni, Y. & Citovsky, V. Effects of calreticulin on viral cell-to-cell movement.

50. Rowland, O.
47. Huang, P.-Y. , Catinot, J. & Zimmerli, L. Ethylene response factors in Arabidopsis immunity.

48. Thomma, B. P., Nürnberger, T. & Joosten, M. H. Of PAMPs and effectors: the blurred PTI-ETI dichotomy.

49. Chen, J.-Y.
46. Wang, J.
45. Katiyar, A.
44. Kim, M. J., Kim, H. R. & Paek, K.-H. Arabidopsis tonoplast proteins TIP1 and TIP2 interact with the cucumber mosaic virus 1a
43. Henry, E., Yadeta, K. A. & Coaker, G. Recognition of bacterial plant pathogens: local, systemic and transgenerational immunity.

42. Zhang, J.
41. Antony, G.
40. Chandran, D. Co-option of developmentally regulated plant SWEET transporters for pathogen nutrition and abiotic stress tolerance. IUBMB life 67, 461–471 (2015).

39. Lim, H.-S.
38. Lim, H.-S. et al. The coat protein of Alternanthera mosaic virus is the elicitor of a temperature-sensitive systemic necrosis in Nicotiana benthamiana, and interacts with a host boron transporter protein. Virology 452, 264–278 (2014).

37. Zhang, C. et al. Overexpression of a novel peanut NBS-LRR gene AhRRS5 enhances disease resistance to Ralstonia solanacearum in tobacco. Plant biotechnology journal (2016).

36. Henry, E., Yadeta, K. A. & Coaker, G. Recognition of bacterial plant pathogens: local, systemic and transgenerational immunity. New phytologist 199, 908–915 (2013).

35. Artico, S., Nardeli, S. M., Brilhante, O., Grossi-de-Sa, M. F. & Alves-Ferreira, M. Identification and evaluation of new reference
34. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols 7, 562–578 (2012).

33. Azhar, M. T., Amin, I., Bashir, A. & Mansoor, S. Characterization of resistance gene analogs from Gossypium arboreum and their evolutionary relationships with homologs from tetraploid cottons. Euphytica 178, 351–362 (2011).

32. Ullah, R.
31. Akhtar, K. et al. Evaluation of Gossypium species for resistance to cotton leaf curl Burewala virus. Annals of applied biology 157, 135–147 (2010).

30. Zhang, Z.
29. Zhang, Z. et al. De novo transcriptome sequencing in Frankliniella occidentalis to identify genes involved in plant virus transmission and insecticide resistance. Genomics 101, 296–305 (2013).

28. Yang, J., Zhang, F. , Li, J., Chen, J.-P. & Zhang, H.-M. Integrative analysis of the microRNAome and transcriptome illuminates the response of susceptible rice plants to rice stripe virus. PLoS one 11, e0146946 (2016).

27. Yang, C.
26. Lu, J. et al. Transcriptome analysis of Nicotiana tabacum infected by Cucumber mosaic virus during systemic symptom development. PLoS one 7, e43447 (2012).

25. Du, P.
24. Zhang, F.
23. Zhang, F. et al. Rice xa13 recessive resistance to soybean mosaic virus.

22. Zhang, Z.
21. Zhang, Z. et al. First symptomatic evidence of infection of Gossypium arboreum with Cotton leaf curl Burewala virus through grafting. International Journal of Agricultural Biology 15, 157–160 (2013).

20. Zhang, Z.
19. Zhang, Z. et al. Identification and evaluation of new reference genes in rice infected by Cucumber mosaic virus during systemic symptom development. BMC plant biology 20, 358–370 (2019).

18. Zhang, Z.
17. Zhang, Z. et al. Specific downregulation of photosynthesis.

16. Zhang, Z.
15. Zhang, Z. et al. Overexpression of A. thaliana gene Xa13 enhances disease resistance to bacterial leaf blight of soybean.

14. Zhang, Z.
13. Zhang, Z. et al. Overexpression of a novel peanut NBS-LRR gene AhRRS5 enhances disease resistance to Ralstonia solanacearum in tobacco. Plant biotechnology journal (2016).

12. Zhang, Z.
11. Zhang, Z. et al. Overexpression of A. thaliana gene Xa13 enhances disease resistance to bacterial leaf blight of soybean.

10. Zhang, Z.
9. Zhang, Z. et al. Overexpression of A. thaliana gene Xa13 enhances disease resistance to bacterial leaf blight of soybean.

8. Zhang, Z.
7. Zhang, Z. et al. Overexpression of A. thaliana gene Xa13 enhances disease resistance to bacterial leaf blight of soybean.

6. Zhang, Z.
5. Zhang, Z. et al. Overexpression of A. thaliana gene Xa13 enhances disease resistance to bacterial leaf blight of soybean.

4. Zhang, Z.
3. Zhang, Z. et al. Overexpression of A. thaliana gene Xa13 enhances disease resistance to bacterial leaf blight of soybean.

2. Zhang, Z.
1. Zhang, Z. et al. Overexpression of A. thaliana gene Xa13 enhances disease resistance to bacterial leaf blight of soybean.
105. Zhong, Y. et al. Digital gene expression analysis of Ponkan Mandarin (Citrus reticulata Blanco) in response to Asia citrus psyllid-vectored Huanglongbing infection. *International journal of molecular sciences* **17**, 1063 (2016).

106. Maimbo, M., Ohnishi, K., Hikichi, Y., Yoshioka, H. & Kiba, A. Induction of a small heat shock protein and its functional roles in Nicotiana plants in the defense response against *Ralstonia solanacearum*. *Plant physiology* **145**, 1588–1599 (2007).

107. Stevenson, C. E. et al. Crystal structure of the MYB domain of the RAD transcription factor from *Antirrhinum majus*. *Proteins: structure, function, and bioinformatics* **65**, 1041–1045 (2006).

108. Van Der Vossen, E. et al. An ancient R gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora* infestans in cultivated potato and tomato. *The plant journal* **36**, 867–882 (2003).

109. Song, J. et al. Gene RB cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. *Proceedings of the national academy of sciences* **100**, 9128–9133 (2003).

110. Bourdenx, B. et al. Overexpression of Arabidopsis *ECERIFERUM1* promotes wax very-long-chain alkane biosynthesis and influences plant response to biotic and abiotic stresses. *Plant physiology* **156**, 29–45 (2011).

111. Meng, D., Walsh, M. & Fricke, W. Rapid changes in root hydraulic conductivity and aquaporin expression in rice (*Oryza sativa* L.) in response to shoot removal-xylem tension as a possible signal. *Annals of botany* **118**, 809–819 (2016).

112. Matsumoto, T. et al. Role of the aquaporin PIP1 subfamily in the chilling tolerance of rice. *Plant and cell physiology* **50**, 216–229 (2009).

113. Tian, S. et al. Plant aquaporin AtPIP1; 4 links apoplastic H2O2 induction to disease immunity pathways. *Plant physiology* pp. 01237.02015 (2016).

114. Nagaraj, S., Senthil-Kumar, M., Ramu, V. S., Wang, K. & Mysore, K. S. Plant ribosomal proteins, RPL12 and RPL19, play a role in nonhost disease resistance against bacterial pathogens. *Frontiers in plant science* **6**, 1192 (2016).

115. Yang, C., Zhang, C., Dittman, J. D. & Whitham, S. A. Differential requirement of ribosomal protein S6 by plant RNA viruses with different translation initiation strategies. *Virology* **390**, 163–173 (2009).

116. Figueiredo, A., Monteiro, F. & Sebastiana, M. Subtilisin-like proteases in plant–pathogen recognition and immune priming: a perspective. *Frontiers in plant science* **5**, 739 (2014).

117. Xia, Y. Proteases in pathogenesis and plant defence. *Cellular microbiology* **6**, 905–913 (2004).

118. Fagard, M. et al. Nitrogen metabolism meets phytopathology. *Journal of experimental botany* **65**, 5643–5656 (2014).

119. Yamamoto-Kato, A., Katou, S., Yoshioka, H., Doke, N. & Kawakita, K. Nitrate reductase is responsible for elicitor-induced nitric oxide production in Nicotiana benthamiana. *Plant and cell physiology* **47**, 726–735 (2006).

120. Modolo, L. V. et al. Decreased arginine and nitrite levels in nitrate reductase-deficient Arabidopsis thaliana plants impair nitric oxide synthase and the hypersensitive response to *Pseudomonas syringae*. *Plant science* **171**, 34–40 (2006).

**Acknowledgements**

The authors highly acknowledge USDA, ICARDA and HEC for funding this research project. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the USDA or ICARDA. This research was supported by grants from the USDA-ICARDA Cotton Productivity Enhancement Project and International Research Support Initiative Program of Higher Education Commission (HEC) of Pakistan. This work was also supported by the National Science Foundation (IOS-1557796) to MSM.

**Author Contributions**

R.Z.N. and S.S.Z. contributed equally to this work. M.A., S.M., B.E.S., J.A.S., L.A.M., conceived and designed the current project. R.Z.N. and K.P.A. were involved in grafting experiments. R.Z.N., S.S.Z. and M.A. performed the experiments. G.J. and M.W. contributed in real time experiments and data analysis. R.Z.N., S.S.Z., S.S., B.M., M.S.M. analyzed the data. R.Z.N. wrote the first draft of the paper. M.A., M.S.M., S.M. and J.A.S. critically revised the manuscript.

**Additional Information**

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-15963-9.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017