In silico prediction of cotton (*Gossypium hirsutum*) encoded microRNAs targets in the genome of *Cotton leaf curl Allahabad virus*

Shweta & Jawaid A Khan*

Plant Virus Laboratory, Department of Biosciences, Jamia Millia Islamia, New Delhi-110025, India; Jawaid A Khan - Email: jkhan1@jmi.ac.in; *Corresponding author

Abstract:
Cotton leaf curl Allahabad virus (CLCuAV) belongs to genus *Begomovirus*, family *Geminiviridae*. It has single stranded monopartite DNA genome transmitted by whitefly (*Bemisia tabaci*). MicroRNAs (miRNAs) belong to class of endogeneous small RNAs which suppress expression of genes following cleavage or translational inhibition of target messenger RNAs. They are demonstrated to be involved in a number of plant processes such as, development, biotic and abiotic stresses. Employing *in silico* approach, high scoring miRNA-target pairs satisfying rules of minimum free energy and maximum complementarity were selected to investigate if they possess the potential to bind the genome CLCuAV. Our results revealed that miRNA species viz., ghr-miR2950 can target all the viral genes, ghr-miR408 targets overlapping transcripts of *AC1* and *AC2* genes; while ghr-miR394 and ghr-miR395a and miR395d could bind overlapping transcripts of *AC1* and *AC4* genes. This is the first report of prediction of cotton miRNAs which have the potential to target CLCuAV genes including *AC1* and *AC4*, involved in viral replication and gene silencing suppression, respectively.

Keywords: MicroRNA, Gene regulation, *Begomovirus*, Cotton leaf curl disease, RNA Interference.

Background:
Cotton leaf curl Allahabad virus is a single stranded DNA virus causing devastating cotton leaf curl disease (CLCuD) in cotton (*Gossypium hirsutum*). It belongs to family *Geminiviridae*, genus *Begomovirus* and transmitted by whitefly *Bemisia tabaci* (Genn) [1]. CLCuD is a serious threat in the Indian subcontinent for several plant species in the family Malvaceae and most importantly for the cultivation of cotton (*Gossypium* sp.). In India, CLCuD was appeared in an epidemic form during the year 1997 in the state of Rajasthan affecting an area of 0.1 million hactares. This disease caused significant losses to cotton crop estimating total value of > $ 5 billion from 1992 to 1997 in Pakistan. Though losses were reduced following the introduction of resistant cultivars, resistance was broken down due to the emergence of new recombinant CLCuV species identified as *Cotton leaf curl Burewala virus*. Now it has become

the limiting factor for cotton productivity in Pakistan and north western India [1]. Therefore, it is of utmost important to devise effective CLCuD resistance strategy.

The genome of *Cotton leaf curl Allahabad virus* is circular, ~ 2.7 kb in size. It transcribes ORFs from virion and complementary strands. The virion sense produces ORFs *AV1* and *AV2*, while complementary strand includes *AC1*, *AC2*, *AC3*, *AC4* and *AC5* **(Figure 1)**. *AC1* encodes the replication- associated protein (Rep) that plays a key role in rolling circle replication and regulation of gene expression of a begomovirus. AC1 is also involved in inhibiting expression of complementary-sense gene. *AC2* encodes the transcriptional activator protein, a multifunctional protein, which acts as a transactivational regulatory factor that regulates transcription of coat protein (CP) and movement protein (MP) encoding genes, and suppresses gene silencing.

AC3 encodes the replication enhancer Protein (REn) leads accumulation of viral DNA followed by symptom development and interacts with Rep to increase Rep mediated ATPase activity. AC4 protein is involved in characteristic vein swelling symptom development in *Arabidopsis* and suppresses miRNAs and siRNAs mediated RNA silencing. AC5 protein is very uncommon protein for begomoviruses and it has been reported that it may contribute in viral DNA replication but not essential for viral infectivity. Coat Protein (AV1), a structural multifunctional protein is involved in genome packaging, insect transmission and systemic spread of virus. Pre-coat protein (AV2), interacts with suppressors of gene silencing viz AC2 or AC4 and maintains high level of viral DNA in infected tissues [2]. Even though several defense mechanisms against virus infections are operative in plants, miRNA- mediated RNAi responses are suggested to be decisive factors in combating the pathogens [3, 4]. RNA Interference (RNAi) is a sequence specific process of post-trancriptional regulation of gene expression. It involves sequence-specific degradation of target messenger mRNA transcripts employing 21-23 nucleotide long microRNA (miRNA) or small interfering RNA (siRNA) [5]. The mechanism is evolutionarily conserved and occurs in plants, animals, fungi and viruses. MiRNAs are involved in a variety of plant processes like organ morphogenesis [6], signaling pathway [7], abiotic stresses [8] and biotic stresses [8].

In ribonucleoprotein complex miRNA binds to the complementary mRNA transcript and results cleavage or translational repression of target. Till now several experimentally validated results of miRNA-mRNA transcript binding are available [9]. These findings led to generalised rules of *in silico* miRNA-target prediction. Most important feature of miRNA-target pair is octameric 'seed' 1–8, 1–7, or 2–8 nucleotide position of the miRNA having Watson-Crick complementarity [10, 11, 12]. Other characteristic that favours mRNA cleavage includes perfect complementarity and most favourable minimum free energy (MFE) of miRNA-mRNA complex [13].

Animal viruses have been found to encode miRNAs that effectively regulate virus gene expression and amend the host's small RNA metabolism. In animals it has been demonstrated that host encoded miRNAs have the potential to target viral mRNA transcripts directly or indirectly by interfering viral replication through miRNA-mediated regulation. Over the past years, several studies have revealed complex host specific miRNA- virus interactions in mammals. It has been found that the host miRNA pathway in mammals brings about powerful effects on virus replication. For example, hsa-miR-210, hsa-miR-199-3p, hsa-miR-125a-5p and hsa-miR-151-5p have been shown to affect HBV gene expression and replication in cultured cells following direct binding to the viral transcripts [14]. Likewise, human miRNAs (viz. hsa-miR-28, hsa-miR-125b, hsa-miR-150, hsa-miR-223 and hsa-miR-382) over-expressed in resting CD4+T cells, were reported to target the 3'end of the human HIV-1 RNA, thus silencing almost all transcribed viral proteins viz. *Gag*, *pol*, *env* [15]. Critical scrutiny of analogous miRNA in the genome of plant infecting viruses provide little evidence towards the existence of virus-encoded miRNAs [3]. However growing number of studies demonstrate that predicted miRNA target sites can be efficiently manipulated to generate artificial miRNA (amiRNA) based antiviral defense strategies in plants.

Niu *et al.*, (2006) were among the first to report miRNA-based defense strategy against viruses. They have shown that these small miRNAs can be engineered to protect plants against the infection by tymoviruses and potyviruses. They engineered miRNA precursor of abundant natural plant miR159 a 273-nt backbone with complementary sequences of viral suppressors of RNA silencing (VSR), viz. P69 of *Turnip yellow mosaic virus* and HC-Pro of *Turnip mosaic virus*. *Arabidopsis thaliana* plants transformed with these recombinant artificial miRNA precursors acquired immunity against infection of these viruses. Transgenic *Nicotiana tabacum* expressing an amiRNA derived from miR171 precursor of *A. thaliana* targeting another VSR, 2b of *Cucumber mosaic virus* (CMV), showed resistance against CMV. This study clearly demonstrated that, expression of miRNA was more effective than expression of short hairpin sRNA.

In another experiment similar protection was achieved against *Potato virus X* (PVX) in *N. tabacum* expressing amiRNAs (developed from *A. thaliana* miR159a, miR167b and miR171a precursors) targetting TGBp1 / p25 domains of PVX. Later, tomato plants expressing amiRNAs against highly conserved 3' untranslated region (UTR) and the overlapped coding sequence of RNA-dependent RNA polymerase 2a protein and the suppressor 2b protein genes of CMV demonstrated resistance even following infection with *Tobacco mosaic virus*, *Tomato yellow leaf curl virus* or CMV [3]. Silencing suppressor genes, such as *AC2* or *AC4*, impart pathogenic attribues to begomoviruses appear to be the important targets for designing antiviral strategy. Transgenic tomato plants expressing amiRNA targeting silencing suppressor genes, *AC2* or *AC4* of *Tomato leaf curl New Delhi virus* led to the successful production of resistant transgenic tomato lines offering a great degree of resistance towards the challenge of Tomato leaf curl virus. Similarly, *N. benthamiana* transformed with cotton miRNA169a, sharing homology with the *AV2* gene of *Cotton leaf curl Burewala virus* (CLCuBuV) showed good resistance against CLCuBuV as well as *Cotton leaf curl Kokhran virus* [4].

The rising number of experimental validations warrants attention towards competitive struggle for survival between hosts and viruses that takes place *in vivo* around the RNA silencing system. It seems that miRNA-mediated plant defence responses have emerged as relevant components of the innate immune system and some miRNA families [3, 4] appears to be playing crucial role in developing plant resistance against virus infections. In view of these findings, we have attempted to explore cotton miRNAs that have the potential to generate defence against CLCuAV disease. In the present study a combinatorial approach, which makes use of more than one computational program for predicting CLCuAV targets of cotton host microRNA set, was applied. Bioinformatics approaches were explored to check the possibility whether cotton host miRNAs have any potential to target CLCuAV genome. This is likely to provide an effective way for exploring complex host miRNA-virus target relationship and could serve as the basis to engineer resistance against CLCuAV infection in cotton (*G. hirsutum*). It will further increase our understanding of the complex virus-host interactions. This is the first report of the identification of cotton miRNAs which potentially could target crucial CLCuAV genes.
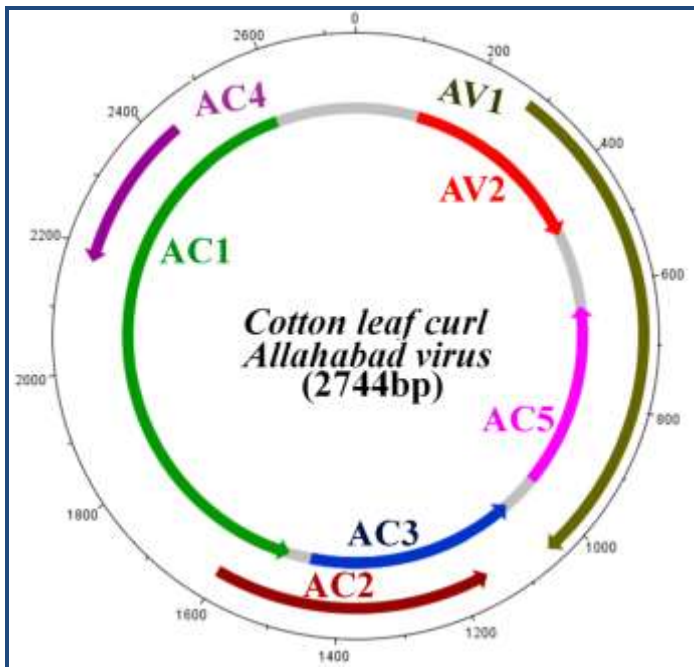
Figure 1: Schematic diagram of *Cotton leaf curl Allahabad virus* (Accession: NC_014897).

Methodology:

Source of data
The Cotton microRNA mature sequences were downloaded from The Plant microRNA database **[16].** For predicting probable targets in the CLCuAV genome (Accession: NC_014897), nucleotide sequences were retrieved from NCBI GenBank (http://ncbi.nlm.nih.gov).

Target prediction
MiRNA target prediction software ─miRanda **[17]** and RNAhybrid **[18]** were used to predict CLCuAV targets against the sixty nine numbers of cotton miRNAs in the genome of CLCuAV using detailed statistical study of minimum free energies (MFEs). Cotton miRNAs targets in CLCuAV genome were initially predicted using miRanda with default parameters (Gap Open Penalty: -9.0; Gap Extend: -4.0; Score Threshold: 50.0; Energy Threshold: -20.0 kcal/mol). With a view to increase the stringency for selection of miRNA-target pairs, a cut-off score 70 was derived by running the same program on a shuffled sequence of CLCuAV with the same set of miRNAs.

Results:
This bioinformatics analysis provided a repertoire of miRNAs targeting respective CLCuAV genes. Only top scoring miRNA-target pairs with maximum complementarity and optimal MFE were selected as summarised in **Table 1 (see supplementary material).** A total number of 18 cotton miRNA targets in the CLCuAV genome were identified above the threshold values. Among them, a total number of 11, 3, 3, 4, 4, 6 and 2 targets were recognized for *AC1*, *AC2*, *AC3*, *AC4*, *AC5*, *AV1* and *AV2* genes of CLCuAV genome respectively. Potential regulatory targets, having five or fewer mismatches and without any gap, was identified by four miRNA family viz. (miR394, miR395, miR2949 and miR3476). *AC1* gene has maximum targets against six cotton miRNAs (viz. ghr-miR394, ghr-miR395a, ghr-miR395d, ghr-miR408, ghr-miR2949 and ghr-miR2950). The MFE (kcal/mol) of these miRNA-target pairs as calculated by

RNAhybrid was claimed to be -25.1,-28.1,-28.1,-25.1,-26.2, and -24.3, respectively. Among them ghr-miR395a, ghr-miR395d and ghr-miR2950 showed multiple loci interaction within nucleotide positions 795-815, 210-230 and 230-250, 1007-1027 in CLCuAV genome, respectively. It is worth to mention that ghr-miR2949 and ghr-miR3476 exclusively target the *AC1* and *AV1* genes while ghr-miR2950 binds to the potential genes of CLCuAV genome. It was observed that *AC2* gene was targeted by three miRNAs (ghr-miR157, ghr-miR408 and ghr-miR2950).*AC3* and *AC5* gene were targeted by single miRNA (ghr-miR2950). The analysis further revealed that *AC4* gene was targeted by ghr-miR394, ghr-miR395a, ghr-miR395d, and ghr-miR2950. *AV1* gene possess potential targets for miRNA *viz.* ghr-miR164, ghr-miR168, ghr-miR394, ghr-miR779, ghr-miR2950 and ghr-miR3476 with MFE of -23.3,-25.2,-25.4,-22.6,-23.2 and -25.0, respectively. *AV2* gene showed putative targets for ghr-miR164 while, ghr-miR168, ghr-miR394 ghr-miR395a, ghr-miR395d, ghr-miR408, ghr-miR2949, ghr-miR2950, and ghr-miR3476 seem to cleave target sites, ghr-miR157, ghr-miR164, ghr-miR394, ghr-miR408 and ghr-miR779 exhibit propensity to result in translation inhibition of the targeted sequences.

Discussion:
This bioinformatics analysis revealed that there is high probability that CLCuAV genome is targeted by abundant and conserved miRNA families in the ORF regions coding for *AC1*, *AC2*, *AC4* and *AV1* genes. However, it requires experimental validation (studies are underway). Fate of a miRNA target site is determined by the degree of complementarity, it predicts with miRNA. In a similar study, miRNA families (viz. miR171, miR156, miR395, miR834 and miR398) were reported to be targeting Scarecrow-like-IV, Squamosa Promoter Binding Protein-Like3, ATP sulfurylases, COP-1 interactive partner and copper/zinc superoxide dismutase2 respectively. It was concluded that if there is perfect complementarity between miRNA and mRNA target, the deserved target will be endonucleolytically cleaved while imperfect complementarity may leads to translational repression or destabilization of mRNA **[3].** In our study, *AC1* gene that encodes a multifunctional protein was targeted by eight different conserved miRNA families. Among them, complementarity (%) of miRNA-target alignment for ghr-miR395a, ghr-miR395d, ghr-miR2949a, ghr-miR2949b and ghr-miR2949c was found to be significantly high (100, 100, 87, 87 and 87, respectively). Sequence complementarity of silencing suppressor *AC4* gene with ghr-miR395a and ghr-miR395d was also observed to be 100%. Artificial miRNA (amiRNA) constructs designed and developed from these miRNA families can serve as a novel strategy to engineer resistance against CLCuAV infection in cotton plants.

Multiple binding of a miRNA on a same target gene leads to cooperativity of binding and hence increases the efficiency and stringency of RNA silencing **[19].** In present computational study, five miRNAs (ghr-miR168, ghr-miR395a, ghr-miR395d, ghr-miR2950 and ghr-miR3476) were predicted to interact with *AC1*, *AC4* and *AV1* genes at multiple loci. The success rate on implication of these miRNAs against respective targets is assumed to be high. Multiple miRNAs binding to multiple locations of target sites reported as effective mean to achieve target repression either by cleavage or translational inhibition **[13].** Overlapping transcripts of various genes appear to be hot-

spots of miRNA-target interaction. Overlapping segment of *AC1* and *AC4* genes were observed to be targeted by ghr-miR394, ghr-miR395a, ghr-miR395d and ghr-miR2950; overlapping transcripts of *AC1* and *AC2* genes shown to be targeted by ghr-miR408 and ghr-miR2950 while overlapping segment of *AV1* and *AV2* genes retrieved to be targeted by ghr-miR164 and ghr-miR2950. These overlapped transcripts emerged as favoured sites of action for the repression of gene expression. Understanding the mechanism of repression of these genes involved in CLCuD onset and progression by miRNAs would be helpful in developing a strategic plan to control pathogenecity of CLCuAV. During miRNA: target binding process *in vivo* free energy changes occur and stability of this miRNA:target duplex is governed by thermodynamic stability [3]. This thermodynamic stability of binding sites is termed as Minimum Free Energy (MFE). Previous studies have established that weak miRNA-target seed binding stability resulted in lower downregultion of targets in plant system [3, 4]. Overlapped transcript pairs of *AC1* and *AC4* genes acquired very high MFE value (-28.0 kcal/mol) with ghr-miR395a and ghr-miR395d suggesting the key role of cotton miRNAs in the process of virus infection and replication. Contiguous binding at central region is very important as pairing between miRNA and target propagate towards central position of miRNA, only after disruption of original argonaute binding with miRNA central and 3' regions [9]. This disruption releases part of energy gained from central pairing, favouring Watson-Crick pairing of adjacent seed region. Ghr-miR157, miR164, ghr-miR168, ghr-miR394, ghr-miR395a, ghr-miR395d, ghr-miR408, ghr-miR779, ghr-miR2949, ghr-miR2950, and ghr-miR3476 were turned up to be the most potential miRNA families which could target CLCuAV genes with perfect and near-perfect complementarity. So the perfect and near perfect complementarity observed between these miRNAs and cognate genes, reflect a functional relationship that the protein-encoding genes of CLCuAV are putative regulatory targets of the cotton miRNAs.

Conclusion:
In this study, we have shown that ten cotton miRNA families (namely miR157, miR164, miR168, miR394, miR395, miR408, miR779, miR2949, miR2950 and miR3476) have the potential to down regulate the expression of *Cotton leaf curl Allahabad virus* genes and could be used to develop virus resistant cotton plant. Several studies have demonstrated the role of miRNA in inhibiting virus infection in mammals [14, 15]. Similarly, artificial-miRNA [amiRNAs] targeting specific genes of plant viruses also have been successfully used to generate resistance against devastating begomoviruses not only in model plants but also in agricultural crops [3, 4]. MiRNAs have least probability to undergo recombination with virus sequence because of their short length as compared to siRNAs. Apart from that it has been testified that begomovirus may escape from siRNA derived RNA interference [20]. Our work is coherent with the view that host miRNAs may act as first line of defence in plants against infecting viruses, the efficacy of this defence against *Cotton leaf curl Allahabad virus,* however need to be carefully evaluated *in vivo.*

References:
[1] Sattar MN *et al. J Gen Virol.* 2013 **94**: 695 [PMID: 23324471]
[2] Fondong VN, *Mol Plant Pathol.* 2013 **14**: 635 [PMID: 23615043]
[3] Ramesh SV *et al. Virus Genes.*2014 **48**: 1 [PMID: 24445902]
[4] Sahu PP *et al. Funct Integr Genomics.* 2013 doi: 10.1007/ s10142-013-0346-z [PMID: 24233104]
[5] Khvorova A *et al. Cell* 2003 **115**: 209 [PMID: 14567918]
[6] Chen X, *Science.*2004 **303**: 2022 [PMID: 12893888]
[7] Guo HS *et al. Plant Cell.* 2005 **17**: 1376 [PMID: 15829603]
[8] Khraiwesh B *et al. Biochim Biophys Acta.* 2012 **1819**: 137 [PMID: 21605713]
[9] Bartel DP *et al. Cell*. 2009 **136**: 215 [PMID: 19167326]
[10] Doench JG & Sharp PA, *Genes Dev.* 2004 **18**: 504 [PMID: 15014042]
[11] Ellwanger DC *et al. Bioinformatics* 2011 **27**: 1346 [PMID: 21441577]
[12] Fabian MR *et al. Annu Rev Biochem.* 2010 **79**: 351 [PMID: 20533884]
[13] Brodersen P *et al. Science* 2008 **320:** 1185 [PMID: 18483398]
[14] Coppola N *et al. PLoS One* 2013 **8**: e65336 [PMID: 23843939]
[15] Huang J *et al. Nat Med.* 2007 **13**: 1241 [PMID: 17906637]
[16] Zhang Z *et al. Nucleic Acids Res*. 2010 **38:** D806 [PMID: 19808935]
[17] Enright AJ *et al. Genome Biol*. 2003 **5**: R1 [PMID: 14709173]
[18] Kruger J & Rehmsmeier M, *Nucleic Acids Res*. 2006 **34**: W451 [PMID: 16845047]
[19] Doench JG & Sharp PA, *Genes Dev.* 2004 **18**: 504 [PMID: 15014042]
[20] Noris E *et al. J Gen Virol.* 2004 **85**: 1745 [PMID: 15166460]

Supplementary material:

Table 1: Potential miRNA-Target pairs derived from miRanda above the threshold values. MFE* is calculated from RNAhybrid Program and Inhibition** is predicted according to the findings of Brodersen *et al.* (2008).

Cotton miRNA	miRNA-Target pair		Target binding position	Score	MFE* (kCal/mol)	% Complem-entarity	Inhibition**
AC1 Gene							
ghr-miR395a,d	Query:	3' uucAAGGGGGUUUGUGAAGUu 5' | | : | | : | | | | | | | : | Ref: 5' gcaTTTCCTCAAACACTTTAa 3'	795-815	166	-28.1	100.00	Cleavage
ghr-miR395a,d	Query:	3' uucaagGGGGUUUGUGAAGUu 5' | | | | | : | | | : | | | : Ref: 5' cagaaaCCCCAGACATTTCGa 3'	210-230	139	-28.0	100.00	Cleavage
AC4 Gene							
ghr-miR395a,d	Query:	3' uucaagGGGGUUUGUGAAGUu 5' | | | | | : | | | : | | | : Ref: 5' cagaaaCCCCAGACATTTCGa 3'	53-73	139	-28.0	100.00	Cleavage
AV1 Gene							
ghr-miR168	Query:	3' ggGUGCUGGACGUGGUUcgu 5' | | : | : | : | | : | | : | Ref: 5' taCATGGCTTGTACTCAtgc 3'	688-707	116	-25.2	93.33	Cleavage
ghr-miR779	Query:	3' uacucgucgUUGUACUUCUgu 5' | | : | | | | | | Ref: 5' agtacggcgACCGTGAAGAac 3'	490-510	128	-22.6	90.00	Translation
ghr-miR3476	Query:	3' cguCGGUUGUUUGGGUCAAGu 5' | : | | | : | : | | | | | | Ref: 5' tttGATAACGAGCCCAGTACg 3'	475-495	143	-25.0	88.24	Cleavage