Identification of *Gossypium hirsutum* miRNA targets in the genome of *Cotton leaf curl Multan virus* and betasatellite

M Sarwar Baig and Jawaid A Khan*

Plant Virus Laboratory, Department of Biosciences, Jamia Millia Islamia, New Delhi 110 025, India

Received 21 November 2012; revised 2 January 2013; accepted 25 February 2013

*Cotton leaf curl Multan virus* (CLCuMV.) in association with betasatellite causes devastating cotton leaf curl disease in cotton plants. Bioinformatics approaches were applied to search cotton (*Gossypium hirsutum*) miRNA targets in the genomes of CLCuMV (DNA-A) and betasatellite (DNA β). A total of 18 nucleotide sequences representing full-length genome (DNA-A) of CLCuMV and 58 nucleotide sequences of full-length DNA β were screened against a set of 60 mature miRNAs of *G. hirsutum*. Efficacy of cotton miRNAs against putative viral mRNA targets having an antiviral activity was analyzed on the basis of complementarity of miRNA-mRNA target pairings, leading to either translational inhibition or endonucleolytic mRNA cleavage, or both. This study revealed 34 putative miRNA targets in DNA-A encoded proteins loci and 2 putative miRNA targets in βC1 gene of DNA β above threshold values. miRNAs, viz., miR168, miR169, miR390, miR395, miR399, miR414, miR779, miR2948, miR2950 and miR3476, were found to be the most potential and could target DNA-A with perfect or nearly perfect complementarity at multiple loci. Similarly, βC1 was targeted by 2 miRNAs, viz., miR398 and miR2950. Among them, miR168, miR169, miR398, miR399, miR779, miR2948 and miR2950 strongly cleaved the mRNA target sites, while miR390, miR414, and miR3476 were probably translational inhibitors. Though miRNA targets were available in different genes of DNA-A, majority of them were found in AC1 gene of DNA-A and satellite conserved region in DNA β. AC1 gene was significantly targeted by 14 miRNAs and βC1 by 2 miRNAs. Interestingly, miR2950 was capable of targeting mRNAs representing both DNA-A and DNA β. Artificially designed miR168, miR169, miR390, miR395, miR398, miR399, miR414, miR779, miR2948, miR2950 and miR3476 targeting DNA-A of CLCuV and βC1 gene of DNA β may have the potential to confer effective resistance against CLCuD infection in transformed cotton.

Keywords: Begomovirus, Cotton leaf curl Multan virus, miRNA

Introduction

Cotton leaf curl disease (CLCuD) is a major constraint to cotton (*Gossypium hirsutum*) production in the Indian subcontinent. It is caused by a monopartite genome (termed as DNA-A) of a cotton leaf curl virus species (CLCuV) in association with symptom-modulating, single-stranded (ss) DNA betasatellite molecule (termed DNA β). An additional satellite-like component, viz., DNA 1, commonly occurs in diseased plants, although it is not required for disease development. The ssDNA-A is approx 2751 nucleotides (nt) long containing 6 open reading frames (ORFs): *AV1* encoding coat protein (CP), *AV2* encoding pre-coat/movement protein, *AC1* replication-associated protein (Rep), *AC2* transcriptional activator protein (TrAP), *AC3* replication enhancer protein (REn) and the *AC4* encoding silencing protein (Fig. 1). These ORFs are transcribed bi-directionally from the large intergenic region (LIR) which contains motifs required for the control of gene expression and replication. The betasatellite encodes a single protein, which is a pathogenicity determinant and suppressor of host defense. During the last two decades, CLCuD has seriously hampered the cultivation and production of cotton crop in Pakistan. CLCuV, which infects a number of diverse hosts in the families Malvaceae and Solanaceae, has also become a serious threat in the irrigated cotton production belt in India. The disease is prevalent in all the primary or secondary cotton-growing areas of the Indian subcontinent. The epidemic of CLCuD occurring across Pakistan and western India has a complex genetic structure, involving numerous distinct CLCuV species. Considerable breeding efforts have been made to develop cotton varieties resistant to CLCuD. However, there has been no comprehensive assessment of the inheritance of this resistance in cotton varieties. Unfortunately, lack of effective
genetic resources of resistance against begomovirus is a limiting factor in developing begomovirus resistance varieties through conventional breeding. Effective measures are, thus, required to combat CLCuD. Among the potential strategies, RNAi (RNA interference) based approach has shown remarkable performance against begomovirus infection. Several studies investigating the use of RNAi for obtaining resistance against begomovirus have been reported with different level of success. Most plant micro- (mi)RNAs show extensive homology to target mRNAs. They are responsible for the degradation of invading viral RNA through the pre-activated RNA silencing machinery (RNAi), following translational inhibition or endonucleolytic viral mRNA cleavage or both. Therefore, the preexistence of virus-specific miRNAs in plants is crucial determinant for plants to acquire virus resistance. However, the low abundance of some miRNAs makes experimental miRNA identification difficult.

In this article, a non-conventional miRNA based approach is designed to develop CLCuD resistance in *G. hirsutum*. As a first step to chemically synthesize miRNA against *Cotton leaf curl Multan virus* (CLCuMV)/ *Cotton leaf curl Multan betasatellite* (CLCuMB) genomes, we have applied bioinformatics approaches to identify potential miRNA regulated genes in the respective genomes. Bioinformatics tools reliably identified the sequence and location characteristics of miRNA recognition elements to predict *G. hirsutum* miRNA targets in DNA-A and DNA β. The designed artificial miRNA (amiRNA) constructs could presumably be used to transform *G. hirsutum* to combat CLCuD.

**Materials and Methods**

**Data Collection**

A set of 69 known miRNA sequences of *G. hirsutum* were downloaded from the plant microRNA database (http://bioinformatics.cau.edu.cn/PMRD/). Further, 18 nucleotide sequences of CLCuMV genome (DNA-A) and 58 nucleotide sequences of associated betasatellite (DNA β) were retrieved from NCBI GenBank (http://www.ncbi.nlm.nih.gov/).

**Prediction of miRNA Targets**

To identify miRNA target sites in DNA-A and DNA β, miRanda target prediction algorithm was applied (http://www.microrna.org/). All the calculations were performed on Linux OS (Operating System) based computer having Intel (R) core i3, 3.20 GHz processor and 2 GB RAM. The miRanda algorithm aligns a miRNA to the target mRNA using a scoring scheme based on the complementarities of nucleotides. It places variable weights on features, such as, (i) complementarity to miRNA in seed region, (ii) thermodynamic stability or free energy of miRNA-mRNA heteroduplex (Vienna package), and (iii) presence of multiple target sites. Seed pairing is weighed more strongly than pairing elsewhere, but seed G:U wobbles and mismatches are allowed. High-scoring targets are then filtered on a secondary criterion of heteroduplex free energy (ΔG). As miRanda does not require exact seed pairing, it predicts sites which contain either a bulge or a G:U wobble in the seed region. For the best scanning of true miRNA targets, initially first eight positions of the miRNA from 5′ ends were focused and different octameric seed types were analyzed.
Identification of True miRNA Targets

In order to find the true miRNA targets in DNA-A and DNA β, three important miRanda parameters, i.e., free energy (\(\Delta G\) kcal/mol), score and percentage of sequence complementarity between miRNA-target duplex were searched. The threshold sequence complementarity score was tuned at 50, free energy was adjusted at –20 kcal/mol and threshold percentage complementarity between miRNA-target duplex was selected at 60. Free energy is commonly used as a measure for miRNA target prediction and indicates the stability of the miRNA-target duplex and the likeliness of correct target cleavage. The miRanda score used for this analysis allows G:U wobble pairs, which are important for accurate detection of miRNA-target duplexes. These criteria have been extensively used for miRNA target prediction in various plants. Apart from these three parameters, true regulatory targets were scrutinized on the assumption that all the potential miRNA targets do not have more than five mismatches. These include one or two mismatches in the octameric seed region, not more than three mismatches in positions 13-22 and complementarity or wobble at position 10 and 11 and not more than a single gap inserted with their corresponding miRNAs.

Results

Detection of miRNA Targets

The partial genome sequences of *G. hirsutum* showed a repertoire of miRNAs having potential for targeting DNA-A and DNA β.

DNA-A

The miRNA families like miR168, miR169, miR390, miR395, miR399, miR414, miR779, miR2948, miR2950 and miR3476 were found to be the most potential to target DNA-A with perfect and nearly perfect complementarity at multiple loci. *AC1* gene was targeted by 14 different miRNAs, followed by *AC4, AV1* and LIR (Large intergenic region), which showed 9, 8 and 3 putative targets, respectively (Fig. 2). *AC1* gene showed maximum interactions with the cotton miRNAs, viz., miR168, miR169, miR390, miR395, miR399, miR414, miR2948 and miR2950 (Fig. 3). Among them, miR168, miR390, miR2948 and miR2950 showed multiple loci interaction at different nucleotide positions.

*AC4* gene was targeted by miR168, miR169, miR2950, miR390 and miR414. Among them, miR390 and miR2948 showed multiple loci interaction at different nt positions, viz., 112-133, 149-173, 153-173, 204-224 (miR390), 23-44, 104-126 and 120-141 (miR2948). It is worth mentioning that nt positions 32-51, 56-75, 80-100, 153-173 and 271-93 targeted by miR168, miR169, miR2950, miR390 and miR414, respectively, are well conserved in *AC4* gene.

*AV1* gene showed putative targets for miR2950, miR408, miR779 and miR3476. The nt positions 65-85, 122-145, 489-510 and 703-722 targeted by miR2950, miR408, miR779 and miR3476, respectively are well conserved in *AV1* gene. Notably, the nt position 65-85 targeted by miR2950 is the only target site that is highly conserved in all the 18 genomic sequences of CLCuMV studied in the present investigation.

The miR2950 unanimously showed interaction with LIR sequences. It is pertinent to mention that miR168, miR169, miR390, miR395, miR399, miR414, miR779, miR2948, miR2950 and miR3476 could target DNA-A at multiple loci in *AC1, AC4, AV1* genes and LIR. Potential regulatory targets, having 5 or fewer mismatches and with no gaps in full-length nucleotide
sequences of DNA-A, were identified by six miRNAs (miR168, miR169, miR390, miR399, miR2950 & miR3476). Further, 3 miRNAs (miR395, miR414 & miR779) identified targets after allowing a single or double nucleotide gaps (Fig. 3).

The miR168, miR169, miR399, miR2948 and miR2950 appeared to be prominent for their target-site cleavage; while miR390, miR395, miR414 and miR3476 were effective for translation inhibition of the targeted sequences. Interestingly, the interaction pattern of octameric seed region at 5′ site and the nt position 10-12 of miRNA in the middle region clearly showed that miR168, miR395 and miR414 were capable to cleave and inhibit translation.

**DNA β**

Attempts were made to search miRNA targets in 58 full-length sequences of DNA β. Among them, 29 genomic sequences showed miRNA targets in βC1 gene that were unanimously targeted by miR398 and miR2950. Both the miRNAs strongly cleaved the target sites (Fig. 4). Moreover, miR2950 showed strong interaction with βC1 gene with 100% complementarity at target site (nt position 512-532).

**Analysis of miRNA Target Sites**

To investigate a possible relationship among different miRNA target sites of DNA-A and DNA β with their corresponding miRNAs, multiple sequence alignment was done.31

**DNA-A**

Potential miRNAs targets were prominently identified in the genes rather than intergenic regions. The largest segment, i.e., AC4 and AC2 overlapping AC1; medium segment, i.e., AV1 partly overlapping AV2; and the smallest segment, i.e., AC2 partially overlapping AC3, seemed to be the abundantly targeted portions, whereas LIR was less targeted by the miRNAs from *G. hirsutum*.

Four target sites representing nt positions 246-264, 259-281, 311-331 and 688-707 in AC1 gene were identified having 5 or less nucleotides changes in the stretch of 1-22 nucleotides. Notably, it revealed a single site (nt position 65-85) targeted by miR2950, which is fully conserved in all the 18 genomic sequences of DNA-A of CLCuMV. We refer two target sites, nt positions 65-85 and 259-281, as the ‘highly conserved targets’ because of the presence of single, double or no mutation. These highly conserved sites, targeted by miRNAs, miR2950 and miR395, may be the most potential targets against DNA-A of CLCuMV isolates analyzed in the present study.

**DNA β**

Two conserved sites, nt positions 437-458 and 512-533, were identified in full-length sequences of DNA β. They were respectively targeted by miR398 and miR2950.

**Discussion**

In the present study, a bioinformatics approach was applied to identify endogenous *G. hirsutum* miRNAs having anti-CLCuD defense by targeting DNA-A of CLCuMV and associated betasatellite DNA β. DNA-A and DNA β were targeted at multiple loci by several miRNAs. The targeted regions were mainly associated with replication (AC1) and silencing suppressor genes (AC4 & βC1, respectively). The study revealed that DNA-A was targeted by abundant and conserved miRNA families in the ORF regions coding for AC1, AC4 and AV1 genes. It was observed that there were 31 target sites representing ORFs of AC1, AC4 and AV1 genes and 3 in LIR of DNA-A. The miRNA targets identified were unequally distributed over the entire sequence of DNA β, while the gene locus in βC1 was less targeted compared to non-coding regions.

The present study clearly demonstrated 34 miRNA putative targets above threshold in nucleotide sequences of DNA-A. AC1 encodes a multifunctional protein that is essential for viral DNA replication as well as acts as a transcriptional repressor of its own expression by binding to the iterated elements located in the LIR.24-25 Therefore, targeting AC1 gene (including the overlapped AC4 gene) would impact
viral replication by reducing viral DNA accumulation in host. RNA-mediated virus resistance has earlier been applied as a potent strategy to generate host resistance against RNA viruses. For begomoviruses, AC1 gene in sense and antisense orientation has also been used with various success rates against Bean golden mosaic virus, African cassava mosaic virus and Tomato yellow leaf curl Sardinia virus.

AC1 that encodes a multifunctional protein was targeted by 14 miRNAs. Among them, miR168, miR2948, miR169, miR399 and miR2950 showed maximum complementarity (94.1, 91.6, 88.2, 87.5 & 82.3%, respectively) as well as multiple loci interaction. Further, 8 miRNAs showed significant hits with AC1 suggesting that these miRNA families and corresponding amiRNA constructs can be manipulated as novel strategies to engineer anti-CLCuD defense in cotton plants. Vein swelling of cotton leaves is due to AC4 activity causing abnormal cell division in vascular bundles. Further, AC4 protein in African cassava mosaic virus and Sri lankan cassava mosaic virus acts as a suppressor of gene silencing. The degree of complementarity determines the fate of a target site. Perfect complementarity leads to endonucleolytic cleavage, while imperfect complementarity results in translational repression leading to destabilization of mRNA. In the present study, miR168, miR169, miR390, miR395, miR399, miR414, miR779, miR2948, miR2950 and miR3476 were found to be the most potential miRNA families which could target CLCuMV with perfect and some near-perfect complementarity in complete DNA-A sequences. The miR168, miR169, miR399, miR2948 and miR2950 appeared to be prominent for their target-site cleavage, while miR390, miR414 and miR3476 were effective for translation inhibition of the targeted DNA-A sequences.

Similarly, miR398 and miR2950 strongly cleaved the target βC1 gene. DNA β molecules are a diverse set of symptom modulating DNA satellites, which are completely dependent on DNA-A for their replication, encapsidation and transmission by whitefly (Bemisia tabaci) vector. These components, typically 1350 nt in length, encode only a single open reading frame (βC1 gene, 356 nt long), which is conserved in position and sequence. In addition, DNA β molecules contain an A-rich region of ~240 nt and a satellite conserved region of ~220 nt, showing high levels of sequence conservation among different diverse DNA β satellites known. Leaf enations and vein greening symptoms in cotton leaves are induced by DNA β, which is a pathogenicity determinant. Cui and associates have demonstrated that βC1 also acts as silencing suppressor following expression of βC1 gene in transformed plants, which induced developmental abnormalities. Hence, βC1 and AC4 share characteristics of pathogenicity and silencing suppression. DNA-A in conjunction with DNA β induces characteristic symptoms of CLCuD, such as, curling of leaf, vein swelling, vein darkening and enations, on the undersides of leaves, which frequently develop into leaf-like structures in cotton plants. In the present study, βC1 was targeted by two different families of miRNA. Therefore, silencing of these regions is likely to impair virus replication and symptom induction in cotton leaf curl diseased cotton plants.

Furthermore, miRNAs not only target miRNAs with high specificity but they also have capabilities of regulating multiple transcripts. In the present study, miR168, miR169, miR390, miR395, miR399, miR2948, miR2950 and miR3476 showed multiple loci interaction in AC1, AC4, AV1 and LIR of CLCuMV genome. Similarly, miR398 and miR2950 revealed multiple loci interaction in βC1 of betasatellite.

In the present study, putative miRNA targets in DNA-A and DNA β were selected following the standard norms of miRanda algorithm. Apart from seed types, perfect complementarity in miRNA and its target region along complete stretch of 1-22 nucleotide, especially no mismatch at the middle 10-11 or 12 position of the miRNA-mRNA duplex, was also focused, which is suitable for mRNA cleavage. Finally, only ungapped miRNA-mRNA predicted target pairs were considered as the best hits. Since octameric seeds show most evidence of functionality, our focus was on the 3-seed types that show strong evidence of perfect Watson-Crick complementarity with nt positions 1-8, 1-7 or 2-8 of the miRNA. Controlling viruses following degrading their mRNAs within a plant cell is a relatively straightforward process and can effectively be achieved using amiRNAs. Niu and colleagues used a 273 bp sequence of the Arabidopsis miR159a pre-miRNA transcript expressing amiRNAs against the viral suppressor genes P69 and HC-Pro to generate resistance against Turnip yellow mosaic virus and Turnip mosaic virus infection, respectively.
a dimeric construct harbouring two unique amiRNAs against both viral suppressors conferred resistance against these two viruses in transformed Arabidopsis plants. Application of amiRNA approach offers several advantages, such as, designing miRNAs with highly target specific and less chances of recombination with viral genome, and more stability than that of siRNA mediated resistance\(^5\). Conclusively, co-expression of amiRNAs (miR168, miR169, miR390, miR395, miR398, miR399, miR414, miR779, miR2948, miR2950 & miR3476) targeting different domains of CLCuV genome and associated betasatellite may lead to the development of broad spectrum resistance against CLCuV infection.

Acknowledgement

Infrastructure support of the Department of Science and Technology, Ministry of Science and Technology, Government of India, New Delhi under FIST programme is gratefully acknowledged.

References


