Journal of Entomology and Zoology Studies

Available online at www.entomoljournal.com

E-ISSN: 2320-7078
P-ISSN: 2349-6800
JEZS 2018; 6(4): 1055-1060

Received: 29-05-2018
Accepted: 30-06-2018

Pooja Verma,
Scientist, ICAR-Central Institute for Cotton Research, Panjari, Wardha, Nagpur, Maharashtra, India

AR Rao
Principal Scientist, ICAR-Indian Agricultural Statistical Research Institute, New Delhi, India

Priyanka
Ph.D. ICAR-Indian Agricultural Research Institute, New Delhi, India

JP Singh
Ph.D. ICAR-Indian Agricultural Research Institute, New Delhi, India

Correspondence
Pooja Verma,
Scientist, ICAR-Central Institute for Cotton Research, Panjari, Wardha, Nagpur, Maharashtra, India

In silico prediction and designing of potential siRNA to control cotton whitefly *Bemisia tabaci* Gennadius vectored cotton leaf curl virus (CLCuV)

Pooja Verma, AR Rao, Priyanka and JP Singh

Abstract
Molecules or agents that silence a gene in a specific manner can be used as a powerful research tool, primarily in the area of functional genomics. siRNA is one of the important and latest addition in this repertoire of sequence specific gene silencing. The experimental result of gene silencing using siRNA highly depends on the sequence of siRNA being chosen. Here, we identified potential siRNAs against different strains of cotton leaf curl virus (CLCV), a severe threat to cotton production and productivity. Out of computationally identified 54 putative siRNAs, 27 qualified as highly potential siRNAs for targeted gene silencing.

Keywords: siRNA, cotton, CLCuV, viral gene

Introduction
Viruses are obligate intracellular parasites and known to be responsible for several devastating diseases of different classes including, food, fibre and medicinal crops which results in significant quantitative and qualitative losses. Cotton is the one of the most important kharif cash crop of India, but it has been the victim of several diseases time to time. Among the factors responsible for low production and productivity of cotton, Cotton leaf curl virus disease (CLCuD), has been found to be one of the key limiting factors. Cotton leaf curl virus, cause for this disease, belongs to the Geminiviridae family comprising of three genera *i.e.* Mastrevirus, *Curtovirus* and *Begomovirus*. A notorious group, causing major menace to cotton crop belongs to the genus *Begomovirus*, which is well notable as Cotton leaf curl virus disease (CLCuD) and is transmitted by whitefly i.e. *Bemisia tabaci* complex (including *B. argentifolii*) in a tenacious manner [1, 2]. The highly irrigated cotton belt of north India, comprising an area of around fifteen lakh hectares is the most affected area by this disease so far. Two genomic components; DNA-A and DNA-B are the main part of *Begomovirus*, which are indispensable for a disease to be transmitted by whitefly *Bemisia tabaci* as per reports [3]. The disease caused by a whitefly transmitted Gemini virus was noticed on *Gossypium peruvianum* and *G. vitifolia* [4] in Nigeria for the first time. In India, cotton leaf curl virus disease was first reported on American cotton (*G. hirsutum*) in Sriganganagar, Rajasthan state during 1993 [5]. During 1994, it started appearing in Haryana and Punjab states on hirsutum cotton as reported by different research groups [6, 7]. Overall it posed a major threat to cotton cultivation in northern India [8], but the most severe spread of CLCuD was witnessed during 2009-10 crop seasons in north zone, based on the surveys carried out during the crop season. To characterize symptoms variability, sequence diversity within the associated isolates and variability in disease pattern, regular monitoring of CLCuV affected cotton has been done. Leaf curl with and without prominent enations, upward as well as down ward curling of leaves were found to be four distinct symptom types. Different isolates having differences in the sequences of their DNA A and DNA B components associated with different symptoms showed existence of significant variation and recombination with other strains of CLCuV.

Extracting information about the factors affecting epidemiology of any disease is required to find the cure. As per the previous reports, rainfall prior to seedling may result in the development of an increased population of vector of CLCuV virus due to abundant food source [9]. Since, cotton is grown only for the part of a year, cultivated hosts and alternate

Weeds serve as virus reservoirs and primary sites of infection are established when whitefly infects these cotton fields. Secondary spread to other plants may occur from the primary sites and from additional vector which enter the field during the whole growing season [10]. Khan *et al*. 1998 [11] used regression analysis on weekly air temperature (maximum and minimum), rainfall, relative humidity and wind movement relationship with severity of plant infection by CLCuD on eight varieties of cotton. Maximum disease incidence % was recorded at 6 week old seedlings which gradually declined with increase in growth of the plants. On the other hand, some researchers found non-significant relationship between white fly population and disease incidence [12, 13, 14].

Discovery of RNA interference and their application in gene silencing has opened tremendous opportunities in all the fields of biological sciences including agriculture and insect-pest management. The strategy involves expressing or inducing the short (21-24 nt) siRNA molecules in the plants which are capable of initiating the homology-dependent gene silencing in a sequence specific manner [15]. Small interfering RNA (siRNA) is the most frequently used RNA interference (RNAi) tool for introducing short-term silencing of protein coding genes. Even chemically synthesized siRNA duplexes, when introduced into the plants, have shown great promises in achieving effective target RNA cleavage [16]. The best way for target-specific gene knockdown is offered by sequence complementarity-based mechanism underlying RNAi but the same can also result in knockdown of genes which are not intended to be directly targeted. Several strategies have been developed to overcome these so-called "off-target" effects, thus ensuring on-target activity. siRNAs, which have been chemically modified provide promising results when it comes to promote preferential loading of the intended antisense (guide) strand into the RISC complex [17, 18] and reduce sense (passenger) strand loading and activity [19, 20]. Further, to check the risk of the siRNA guide strand seed region from causing off-target effects, algorithms have been designed which can incorporate different filters to exclude high-frequency seed sequences from known mammalian microRNAs [21]. Chemical modifications or thermodynamic-based design considerations can also be applied to the siRNA seed region to discourage unintended interactions. In the present study, an *in silico* approach has been used and applied for screening and designing of efficient siRNAs to counter the diseases caused by different strains of cotton leaf curl virus, emphasizing primarily on eight viral genes from four different cotton leaf curl virus strains causing cotton leaf curl disease (CLCuD).

2. Materials and methods

2.1 Retrieval (Collection) of viral target sequences

Two different viral genes from each of the four viral strains have been selected. The complete sequences of these viral genes namely coat protein (AV1) and replication initiation proteins (AC1) were retrieved from NCBI. The complete information of the two genes for each virus, and CLCuV replicase gene used to be studied in the present investigation, is summarized in Table 1.

Table 1: List of selected genes with their accession numbers

S. No	Gene Name	Gene length (nt)	Gen Bank Accession number
1	CLCuV-S coat protein (AV1) gene	771	AY577460.1
2	CLCuV-S replication initiation protein (AC1) gene	328	AY146961.1
3	CLCuV-L coat protein (AV1) gene	770	AF465619.1
4	CLCuV-L replication initiation protein (AC1) gene	396	AY146963.1
5	CLCuV-H coat protein (AV1)gene	771	AF465620.1
6	CLCuV-H replication initiation protein (AC1) gene	388	AY146964.1
7	CLCuV-D coat protein (AV1)gene	771	AF465618.1
8	CLCuV-D replication initiation protein (AC1) gene	349	AY146962.1

2.2 Target Identification and Designing of Candidate siRNA

The first and primary step while designing the siRNA, is the identification of various target sequences for candidate siRNAs within the viral genes. The various targets were screened based on algorithms incorporated in three highly cited siRNA design tools namely, Genscript (https://www.genscript.com/ssl-bin/app/rnai), siRNA Wizard v3.1 (http://www.sirnawizard.com/design.php) and Whitehead WI siRNA selection programme using default settings. Several multi stringent criteria as recommended by Birmingham *et al*., 2007 [21] and Elbashir *et al*., 2001 [22] was employed in order to screen the best candidate siRNAs which in turn significantly improves the likelihood of identifying functional siRNA. This approach would enhance the target specificity and adapt siRNA designs for more sophisticated experimental design in RNAi mediated gene silencing.

Criteria in detail:

1. siRNA targeted sequence is usually 21 nt in length.
2. Avoid regions within 50-100 bp of the start codon and the termination codon
3. Avoid intron regions
4. Avoid stretches of 4 or more bases such as AAAA, CCCC
5. Avoid regions with GC content <30% or > 60%.
6. Avoid repeats and low complex sequence
7. Avoid single nucleotide polymorphism (SNP) sites

2.3 Cross Validation of Candidate siRNAs by BLAST Search against mRNA Database

The candidate siRNAs were selected to design putative siRNAs for efficient silencing of viral genes, where these sequences do not reveal any homology to any other gene sequences *i.e*., target site having more than 16 nt contiguous base pair with any other organism. In addition, to ensure and reduce the off-target cleavage effect of the designed siRNAs in plant gene, BLASTn search was performed against non-redundant (NR) database of NCBI of the selected 21-24 nt long candidate siRNA sequences. Cases where the candidate siRNA showed any homology to any of the plant genes from the database were discarded from the list of candidate siRNAs that could be an efficient silencing tool against the target viruses, so as to avoid any unintended effect on endogenous gene expression.

2.4 GC Content, G value and siRNA Secondary Structure Prediction

The short listed siRNAs were further screened for estimation of GC percentage and secondary structures prediction to design the efficient siRNA. The i-Score designer tool (http://www.med.nagoya-u.ac.jp/neurogenetics/i_Score/i_score.html) was used to calculate the GC % and G value of the most stable secondary structure of siRNA strand according to the RNA m fold algorithm. The secondary structure and free energy of folding of each putative siRNA was computed using RNA fold web server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

3. Results and Discussion

Plant viruses have been considered as preferred objects for understanding host-microbe interactions because they code very few proteins and are exclusively dependent on host cellular machinery for their multiplication and movement activities. Among all viral proteins, replication initiation protein (AC1/ RInP) plays an essential role in multiplication and also serves as an elicitor of the hypersensitive response (HR). In fact most of the viral proteins such as AC1, CP (coat protein) and MP (movement protein) are required for virus infection including their movement and replication. The viral gene sequences of different CLCuV viruses infecting cotton crop, downloaded from Gen Bank of NCBI have been reported in Table 1. To select various target sequence for siRNA within the viral genes, a stretch of nucleotides with 5´AA dinucleotide or triplet AAG/C with 18/19 nucleotides were scanned manually. In addition to this, a stretch of 4 T's or A's in the target sequence were avoided as it may act as a termination signal for RNA polymerase III enzyme [22]. Further, a number of filtering procedures described by Birmingham *et al*., 2007 [21] and Elbashir *et al*., [22] 2001, were employed to screen out the best possible candidate siRNAs from the number of siRNA generated by different softwares. More than 150 siRNA sequences satisfying the mixed rules of Ui-Tei, Hsieh, s-Biopredsi and i-Score were considered for further analyses. A cross homology search of candidate siRNA against non-redundant database of NCBI through BLAST was performed and 54 putative siRNAs were screened out. Cross homology search revealed that the designed siRNAs showed a similarity of 95-100% to considered plant viruses but not to any other gene of the plant genome, thus assured silencing of only selected viral plant genes can be achieved through the designed siRNA. Furthermore, the putative siRNAs were assessed for their possible folding pattern using RNA fold server, a widely used server to predict minimum free energy (MFE) structures and base pair probabilities of RNA molecules.

Out of the 54 putative siRNA, only 27 siRNA sequences showed zero free energy of folding at room temperature (37°C) (Table 2). The secondary structure of siRNAs has been shown in figure 1, 2 and 3. It has been reported earlier [23] that the RNA molecule should have minimum free energy of folding for their stability. Therefore, the RNA molecule with positive energy renders the high potential for binding and may be more accessible for target site lead towards an effective gene silencing.

Table 2: List of highly potential siRNA sequences developed for the viral coding genes with their sequence features (%GC, Free energy of binding)

S. NO	Sequence	Length (nt)	Location of siRNA Within gene	Gene name	Virus	Software used	GC %	Total free energy For binding (kcal/mol)	Secondary structure
1	AAGACGAAGAATCACACGAAT	21	381	CLCuV- S (AV1)	CLCuV- S	Genscript	38.1	-31.7	1a
2	AAGGTCCATGTAAGGTTCAGT	21	208	CLCuV- S (AV1)	CLCuV- S	Genscript	42.86	-35.3	2a
3	AATGCTTTATATGGCTTGTAC	21	683	CLCuV- S (AV1)	CLCuV- S	Genscript	33.33	-30.7	3a
4	AATACGAGAATCATACGGAA A	21	655	CLCuV- S (AV1)	CLCuV- S	Genscript	33.33	-30	4a
5	GAACAGGAAGCCCAGGATGT ACA	23	147	CLCuV- S (AV1)	CLCuV- S	SiRNA wizard	52.17	-41.5	5a
6	GGATGTACAGAAGTCCAGATG TT	23	170	CLCuV- S (AV1)	CLCuV- S	SiRNA wizard	43.48	-38.7	6a
7	AAGAACCATACGAATTCGGTG	21	387	CLCuV- L (AV1)	CLCuV- L	Genscript	42.86	-33.2	1b
8	AAGGAACAGGCGTTGGTTAA G	21	582	CLCuV- L (AV1)	CLCuV-	Genscript	47.62	-35.6	2b
9	GTTGTCCATATTGGTAAGGTA AT	23	247	CLCuV- L (AV1)	CLCuV- L	SiRNA wizard	34.78	-34.8	3b
10	AATGCTTTATATGGCTTGTAC	21	683	CLCuV- H AV1)	CLCuV- H	Genscript	33.33	-30.7	1c
11	AATACGAGAATCATACGGAA A	21	655	CLCuV- H AV1)	CLCuV- H	Genscript	33.33	-30	2c
12	GAACATTAAGACGAAGAATC ACA	23	372	CLCuV- H (AV1)	CLCuV- H	SiRNA wizard	34.78	-30.7	3c
13	AAGTCCAGATGTTCCTAGAGG	21	182	CLCuV- D AV1)	CLCuV- D	Genscript	47.62	-37.4	1d
14	AAGAAGTTTGTCAGAGTTAAC	21	600	CLCuV- D AV1)	CLCuV- D	Genscript	33.33	-30.2	2d
15	GAAGGTCCATGTAAGGTTCAG	21	156	CLCuV- D AV1)	CLCuV- D	siRNA wizard	47.62	-37.2	3d
16	AGTGATGTTACTCGTGGTACT	21	228	CLCuV- D (AV1)	CLCuV- D	siRNA wizard	42.86	-36.6	4d
17	GAGAACATTAAGACGAAGAAT	21	321	CLCuV- D (AV1)	CLCuV- D	siRNA wizard	33.3338.1	-33.2	5d
18	AGATCGTAGACCTGTTGATAA	21	374	CLCuV- D (AV1)	CLCuV- D	siRNA wizard		-36.8	6d
19	AACCAGGTCAGCACATTTCCA	21	239	RInP	CLCuV- H	Genscript	47.62	-37.4	1e
20	AATGCTCTCTAACTAAAGAAG	21	55	RInP	CLCuV- H	Genscript	33.33	-30.3	2e
21	AAATCAAGCTCAGATGTCAAG	21	279	RInP	CLCuV- H	Genscript	38.1	-32.4	3e
22	AATCAAGCTCAGATGTCAAGG	21	280	RInP	CLCuV- L	Genscript	42.86	-34.8	1f
23	GCACATTTCCATCCGAACATT	21	247	RInP	CLCuV- L	siRNA wizard	42.86	-37.3	2f
24	AACCCTCTCCCAATTACTAAA	21	75	RInP	CLCuV- D	siRNA wizard	38.1	-37.2	1g
25	AAACCTACATACACCAGTAAA	21	93	RInP	CLCuV- D	siRNA wizard	33.33	-34.9	2g
26	AAGCAGGTCAGCACATTTCCA	21	242	RInP	CLCuV- S	Genscript	47.62	-37.4	1h
27	AAATCCAGCTCCGACGTCAAA	21	282	RInP	CLCuV- S	Genscript	47.62	-36.1	2h

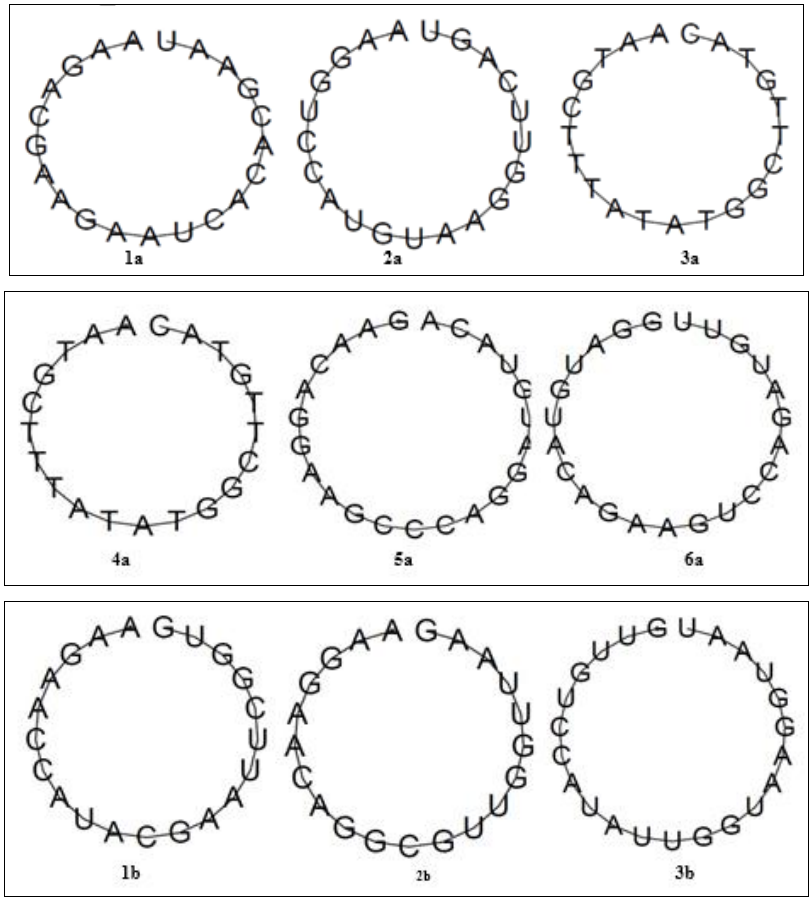

Fig 1: (1a-3b). siRNA secondary

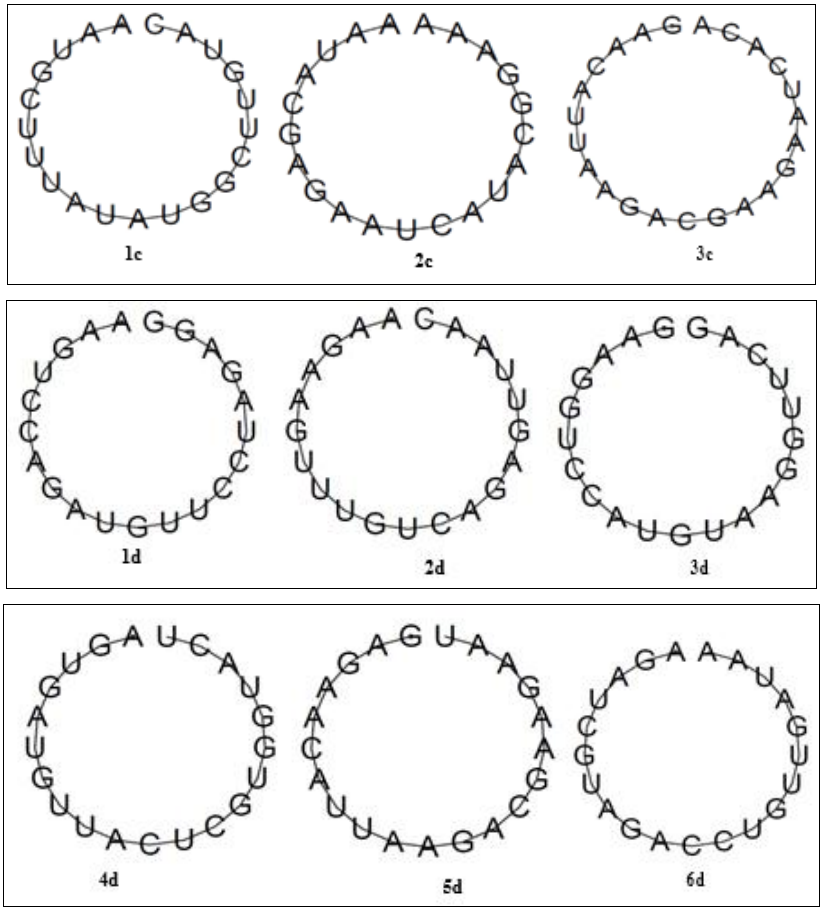

Fig 2: (1c-6d). siRNA secondary structures.

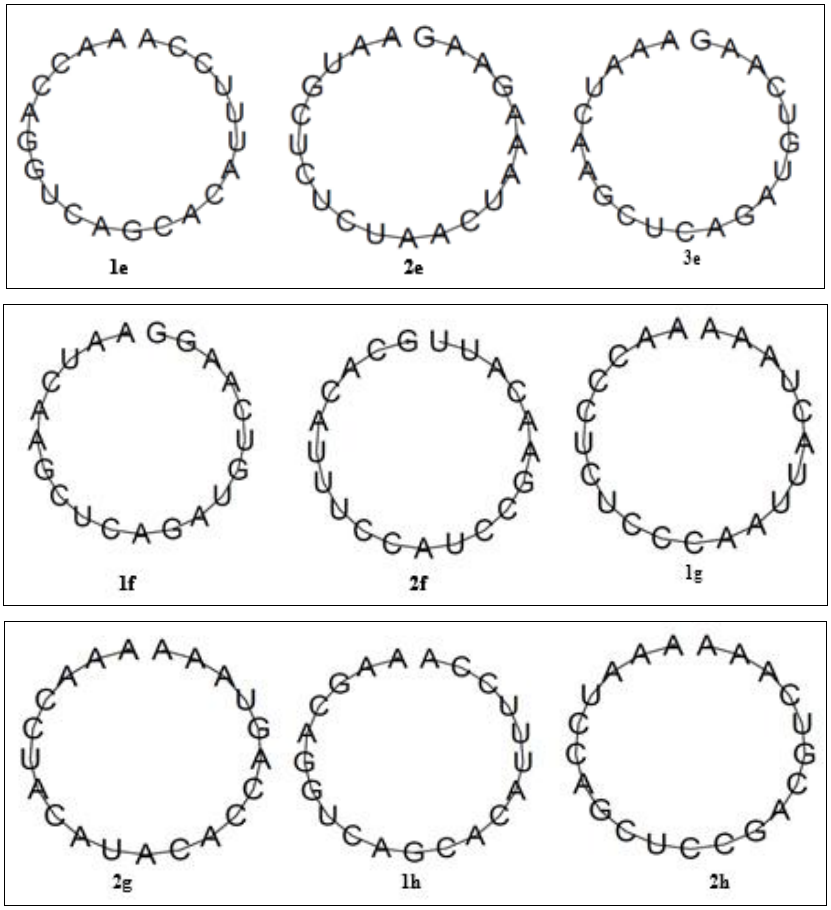

Fig 3: (1e-2h). siRNA secondary structures.

4. Conclusion

Plant viruses are considered as severe threat to agricultural practices across the globe and one of the major constrain to the development of virus resistant crop plants. The economically important cash crops are prone to various plant diseases, resulting million dollar loss every year thereby affects the global economy. As siRNA mediated gene silencing of viral genes have been proven to become a pivotal tool in the development of virus resistance crop plants, it is of interest to design effective and efficient siRNA to silence the most casual agents of virus infection and replication indifferent crops. The present study involves screening and design of 27 promising siRNA against different strains of CLCuV considering two genes from each virus, which have been obtained after qualifying all the criteria. These will build a valuable repository of siRNA for their chemical synthesis.

5. References

1. Brown JK, Frohlich DR, Rosell RC. The sweet potato or silver leaf whitefly; biotype of *Bemisia tabaci* or a species complex. Annual Review of Entomology. 1995; 40:511-534.
2. Rybicki E, Fouquet C. Geminiviridae classification: current concepts and demarcation criteria. 2nd International workshop on *Bemisia* and Geminiviral Diseases. San Juan, Puerto Rico, 1998, 65.
3. Monga D, Soni R. Detection of weeds carrying cotton leaf curl virus disease using PCR. 2011.
4. Farquharson CO. A report of the mycologist. A manual report of Agriculture Department of Nigeria. In: Siddique MA and Hungus LC (Eds) Cotton growth in Gezira environment. W Haffer and Sons Ltd, Cambridge, England, 1912, 106.
5. Ajmera BD. Occurrence of leaf curl virus on American Cotton (*G. hirsutum*) in north Rajasthan. National Seminar on Cotton Production Challenges in 21st Century, Hisar, India, 1994.
6. Rishi N, Chauhan MS. Appearance of leaf curl disease of Cotton in northern India. Journal of Cotton Research and Development. 1994; 8:174-180.
7. Singh J, Sohi AS, Mann HS, Kapoor SP. Studies on whitefly *Bemisia tabaci* (Genn.) transmitted cotton leaf curl virus disease in Punjab. Journal of Insect Science. 1994; 7:194-198.
8. Verma A, Puri SN, Raj S, Bhardwaj RP, Kannan A, Jayaswal AP *et al*. Leaf curl disease of cotton in North-West-India. Report of the ICAR Committee. 1995.
9. Bink FA. Leaf curl and mosaic diseases of cotton in central Africa. Empire Cotton Growing Review. 1975; 52:133-41.
10. Giha OH, Nour MA. Epidemiology of cotton leaf curl virus in the Sudan. Cotton Growing Review. 1969; 46:105-118.
11. Khan MA, Mirza JH and Ahmed S Relationships of environmental conditions conducive to cotton leaf curl virus disease development. Pakistan Journal of Phytopathology. 1998; 10:5-8.
12. Briddon RW, Mansoor S, Pinner MS, Markham PG. Infectious clones of cotton leaf curl virus. 2nd Int. workshop on *Bemisia* and Gemini viral disease. San Juan, Puerto Rico (Abst). 1998, 35.
13. Hameed S, Khalid S, Haq EU, Hashmi AA. Cotton leaf curl disease in Pakistan caused by a whitefly transmitted Gemini viruses. Plant Disease. 1994; 78: 528-529.
14. Iqbal M, Khan MA. Management of Cotton leaf curl virus by planting time and plant spacing. Advances in

Agriculture and Botanics. 2010; (2)1.
15. Praveen S, Mishra AK, Antony G. Viral Suppression in Transgenic Plants Expressing Chimeric Transgene from Tomato Leaf Curl Virus and Cucumber Mosaic Virus. Plant Cell Tissue Organ Culture. 2006; 84(1):49-55.
16. Tenllado F, Diaz-Ruiz J R. Double-Stranded RNA-Mediated Interference with Plant Virus Infection. Journal of Virology. 2001; 75(24):12288-12297.
17. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell. 2003; 115: 209-216.
18. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. Cell. 2003; 115:199-208.
19. Jackson AL, Burchard J, Schelter J, Chau BN, Cleary M, Lim L *et al*. Wide- spread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. RNA. 2006; 12:1179-1187.
20. Vaish N, Chen F, Seth S, Fosnaugh K, LiuY, Adami R *et al*. Improved specificity of gene silencing by siRNAs containing unlocked nucleobase analogs. Nucleic Acids Research. 2011; 39:1823-1832.
21. Birmingham A, Anderson E, Sullivan K, Reynolds A, Boese Q, Leake D, *et al*. A protocol for designing siRNAs with high functionality and specificity. Nature Protocols. 2007; 2(9):2068-2078.
22. Elbashir SM., Lendeckel W, Tuschl T. RNA interference is mediated by 21 and 22 nucleotide RNAs. Genes and Development. 2001; 15:188-200.
23. Shao Y, Chan CY, Maliyekkel A. Lawrence CE, Roninson IB, Ding Y. Effect of target secondary structure on RNAi efficiency, RNA. 2007; 13(10):1631-1640.
24. https://www.genscript.com/ssl-bin/app/rnai
25. http://www.sirnawizard.com/design.php
26. http://www.med.nagoya-u.ac.jp/neurogenetics/i_Score/i_score.html
27. http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi