ECRONICON OPEN ACCESS

EC MICROBIOLOGY
Research Article

Molecular Characterization of Cotton Infecting *Begomovirus*(es) and DNA Satellites Associated with Cotton Leaf Curl Disease in Pakistan

Malik Nawaz Shuja[1,2]* and Muhammad Tahir[1]

[1]*Department of Plant Biotechnology, Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan*

[2]*Department of Microbiology, Kohat University of Science and Technology (KUST), Kohat, Khyber Pakhtunkhwa, Pakistan*

***Corresponding Author:** Malik Nawaz Shuja, Department of Microbiology, Kohat University of Science and Technology (KUST), Kohat, Khyber Pakhtunkhwa, Pakistan.

Received: December 20, 2018; **Published:** June 13, 2019

Abstract

Cotton leaf curl disease (CLCuD) is the major biotic encumbrance to cotton production in the world, transmitted by the insect vector *Bemisia tabaci* (whitefly). The disease was early observed on cotton in 1967 but were not significant until 1986 on Indian subcontinent. The first epidemic of CLCuD in early 1990's was associated with several distinct begomoviruses along with a disease-specific beta-satellite. Resistant cotton varieties were introduced in late 1990's but soon resistance was broken and was associated with a single recombinant *Begomovirus* named Burewala strain of *Cotton leaf curl Kokhran virus* that lacks a full complement of a gene encoding a transcription activator protein (TrAP). In order to understand the ongoing changes in CLCuD complex in Pakistan, CLCuD affected plants from cotton fields were collected from cotton belt in Punjab province, Pakistan. Total nucleic acid from leaf tissues was extracted and subjected to PCR. Thirty full-length genomic components were cloned and sequenced, corresponding to *Begomovirus* and associated DNA alpha- and beta-satellite. The sample from a non-symptomatic plant did not yield the amplification product. Analysis of nine *Begomovirus* sequences showed the variant of Cotton leaf curl Kokhran virus-Burewala strains, except one clone which was a new recombinant, named Cotton leaf curl Kokhran virus -Layyah, which also lacks an intact TrAP gene. Associated alpha and beta-satellite sequences showed highest identity with Cotton leaf curl Multan alpha-satellite and Cotton leaf curl Multan beta-satellite. Further, a partial dimeric (2.1 mer) DNA-A, (0.55 mer) beta-satellite and (0.9 mer) alpha-satellite clones were constructed, containing the Nona-nucleotide region, in a binary vector pGreen0029 and used in *Agrobacterium*-mediated inoculation of *Nicotiana benthamiana, N. tabacum, Lycopercicum esculentum, Cucurbita pepo* and *Cucumis sativus.* Leaf curling, deformation and yellowing like symptoms were observed on inoculated plants, while control plants produced no symptoms. Following PCR amplification, the viral DNA as well sub-genomic components were detected in newly emerging non-inoculated leaves. Both molecular and biological studies indicate that this new isolate is a virulent strain of CLCuD and has the potential to aggravate symptomology, thus posing an alarming situation which may lead to another CLCuD outbreak in the subcontinent.

Keywords: *Begomovirus; Beta-Satellite; Cotton Leaf Curl Disease; Recombinant*

Abbreviations

CLCuD: Cotton Leaf Curl Disease; CLCuKoV-Bu: Cotton Leaf Curl Kokhran Virus- Burewala; CLCuMB: Cotton Leaf Curl Multan Beta-Satellite

Introduction

Geminiviruses (the family *Geminiviridae*) are plant pathogens with genome of small circular single-stranded DNA with geminate isometric particle morphology [1,2]. These are found in tropical to warm temperate geographical zones and infect a wide range of plants including crops, native weeds and wild plants. Geminiviruses are classified into nine genera *Begomovirus, Mastrevirus, Curtovirus, Topocuvirus, Turncurtovirus, Becurtovirus, Eragrovirus, Capulavirus* and *Grablovirus,* based on insect vector involved in transmission, plant host range (either mono or dicotyledonous hosts) and genome architecture [3]. The genus *Begomovirus* is the largest genus of geminiviruses transmitted by whitefly, *Bemisia tabaci*. Begomoviruses have emerged everywhere in the world where environmental conditions support large populations of *B. tabaci* and have become a major threat for the production of food and fiber crops [4-6].

Begomoviruses have either bipartite (two circular, single-stranded DNA components known as DNA-A or DNA-B) or monopartite (equivalent of DNA-A of bipartite *Begomovirus*) genomes. The DNA-A component encoding six proteins and the DNA-B component encoding two proteins, each component being 2.5 - 2.8 kbp in size. The DNA-A component is capable of autonomous replication and can produce virus particles that encode proteins responsible for particle encapsidation (Coat Protein), viral replication (replication-associated protein (Rep), replication enhancer protein (REn), regulation of gene expression and suppression of host defense mediated by gene silencing (transcriptional activator protein (TrAP)). Modulation of symptom development and suppressor of gene silencing was attributed to the protein encoded by C4 protein of begomoviruses originating from the Old World. The New World begomoviruses lack an ORF termed V2 but it is found in begomoviruses originating from the Old World. The V2 protein is involved in movement and pathogenicity. The ORF BV1 and BC1 located on DNA-B component encode two proteins involved in intracellular (nuclear shuttle protein- NSP) and intercellular (movement protein, MP) virus movement [7]. The two components of begomoviruses have only a little sequence homology except for an intergenic region (IR), which includes the common region (CR) of approximately 200 nucleotides with typically more than 85% identity between DNA-A and DNA-B. The CR contains a highly conserved (among geminiviruses) hairpin structure with the characteristic nonanucleotide sequence (TAATATT↑AC) that marks the origin of virion-strand DNA replication [8] and repeated sequences (collectively known as "iterons") required for the binding of the Rep protein [9-12] and regulatory regions for bidirectional transcription. The CR thus functions in integrity maintenance of the divided genome and in certifying/ensuring the replication of both components initiated by the DNA-A-encoded Rep protein [13].

Recombination and pseudorecombination are very frequent and widespread phenomenon that occur between species as well as within or across genera and are significant contributors to geminivirus evolution. Recombination is the exchange of DNA between similar DNA components, and pseudorecombination is the exchange of whole genome components between the viruses [14]. The evolution of virulent viral strains by recombination mechanism can have devastating consequences on the host population structure and skipping of evolved novel virus to the new host species [15]. High diversity among *Begomovirus* species associated with mixed infections may be the major facilitator for the recombination and pseudorecombination events, leading to the frequent emergence of new begomoviruses [16,17]. Recent studies of *Begomovirus* populations from the Indian subcontinent revealed their evolutionary divergence as a new group of begomoviruses, distinct from other begomoviruses causing similar diseases in other geographical regions of the world [18].

Pakistan, a major producer of the cotton, accounts for 60% of export value. Unfortunately, the crop is deadly infected by several pathogens and pests, and among them cotton leaf curl disease (CLCuD) is one of the major hindrance to the production of cotton across Pakistan and border link countries since the early 1990s [19]. The disease is caused by a complex of begomoviruses (family *Geminiviridae*). The disease has been early reported from Multan vicinity in 1967 [20] but its epidemic proportion was not evident until 1986. It was this year, when the disease spread and encompass a huge region of cotton belt of Pakistan and into the northwestern India. An attempt to reduce the losses, resistant varieties of cotton were introduced in the late 1990s [21] and losses due the disease were reduced its pre-epidemic threshold. After a gap of some years, disease on resistant cotton cultivars began to exhibit typical symptoms CLCuD in the vicinity of "Burewala" [22]. This attack was a gestured for the onset of a second epidemic on cotton; which was referred to as the "Burewala strain" of the disease. The previous epidemic during 1990s, referred as Multan strain, involved several begomoviruses with a single disease-specific

beta-satellite (cotton leaf curl Multan beta-satellite [CLCuMB]) and nanovirus like component known as alpha-satellite (cotton leaf curl Multan alpha-satellite [CLCuMA]). In previous studies, Burewala strain was found more infectious and devastating than Multan strain, resulting 100% crop losses in many areas at early stage of infection [23-26].

After Burewala epidemic, the dissemination of this virus with new areas and hosts has been reported as a dominant virus. Keeping this in view, samples were investigated in the cotton belt, to track virus diversity and evolution in a long-lasting host-pathogen interaction.

Materials and Methods

Sample collection

During a survey of the cotton belt in Punjab-Pakistan, cotton plant samples were collected. Plants exhibiting symptoms like leaf curling, vein darkening, vein swelling and sometimes enations were collected in zipbags. Ten to six leaves per plant were collected at different locations. The zipbags were properly labeled and stored at -80°C till further processing.

DNA extraction and virus detection

Total nucleic acid was extracted from each sample using approx. 100 mg frozen leaf tissue through CTAB (Cetyl trimethyl ammonium bromide) method [27]. DNA was finally resuspended in TE buffer or nuclease-free water and stored at -20ºC. Quantity of DNA was measured by Nano Drop machine. From the total DNA, 5 ul was electrophoresed on 1% agarose gel with a DNA marker. Initially, a set of degenerative primers WTGF/WTGR to amplify/detect all *Begomovirus* DNA-A component, covering 1.5 kb region starting from the Rep to CP including the LIR [28]. Another set of diagnostic set of primers CLCV1/CLCV2 which amplifies/detects only cotton infecting leaf curl viruses was also used.

Amplification, cloning and sequencing

To amplify full length virus sequence, two sets of abutting primers (BurXF [5´-CTCGAGAGTGTCCCCGTCCTTGTCG-3´]/BurXR [5'-CTCGAGTGGGGAGAGTTTCAGATCG-3], and BurNF [5´-CCATGGTTGTGGCAGTTGATTGACAGATAC-3´]/BurNR [5´- CCATGGATTCACGCACAGGGGAACCC-3´]) were designed with a unique internal restriction site from the partial sequenced clones. For PCR amplification of associated alpha- and beta-satellite, a primer set AlphaF/R and Beta01/02 was used [29,30] following standard procedure. The PCR amplified products were cloned into pTZ57R/T plasmid vector (InsTAclone™ PCR Cloning Kit, Fermentas) or pDrive vector (Invitrogen) Cloning Kit. Full length clones were sequenced to their entirety in both orientations (Macrogen, Korea) with no ambiguities remaining.

Sequence analysis

Complete nucleotide sequences of DNA-A, alpha- and beta-satellite were initially analyzed using BLASTn (http://www.ncbi.nlm.nih.gov/BLAST/) search. An online ORF (Open Reading Frame) finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was used to indicate ORFs in the genomes. For the sequence alignment, high percentage identical sequences from the databases following NCBI-BLAST analysis were selected and aligned using MUSCLE for DNA-A (accordingly as [31] and CLUSTAL-X for alpha-and beta-satellite sequences [32]. Phylogenetic trees were viewed, manipulated and printed using either MEGA6 or Treeview tool [33].

Recombinant sequences were identified with Plotcon, which is part of EMBOSS [34] and using the software Recombination Detection Program (RDP; http://darwin.uvigo.es/rdp/rdp.html) version 4.0 [35] for detection of potential recombinant sequences, identification of likely parental sequences, and breakpoints were carried out using the RDP, GENECONV, BOOTSCAN, MAXIMUM CHI SQUARE, CHIMAERA, SISTER SCAN and 3Seq recombination detection methods implemented in the program.

Construction of partial dimers

To investigate the possible effects of Burewala strain and Layyah strain, CLCuKoV-Bu [GenBank HF549180] and its cognates alpha-satellite (CLCuMA-[GenBank KR816016]) and beta-satellite [GenBank HF549185] and CLCuKo-La [GenBank HF549182] and its associated alpha-satellite [GenBank KR816008] and beta-satellite [GenBank HG000665] obtained from sample C28 and C49 were selected to produce constructs for *Agrobacterium*-mediated inoculation. A fragment of approximately 2.1 kb was released from C28A clone (CLCuKoV-Bu [GenBank HF549180]) utilizing *Xho*I and *Eco*RI sites which containing the nona-nucleotide region. The fragment that contained

the hairpin loop was ligated into a binary vector pGreen0029, termed as C28pGreen-A-2.1. Again, the full-length isolate was released from C28A clone and ligated into the unique *Xho*I site into C28pGreen-A-2.1, now termed as C28pGreen-A-4.9 containing a 2.1-mer partial tandem repeat. For construction of partial dimer for alpha-satellite, approximately 900 bp *Hind*III-*Pst*I fragment of the alpha-satellite C28α [GenBank KR816016] was cloned into pGreen0029, producing C28pGreen-α-0.9. Again, full length alpha-satellite from the clone C28α was released with digestion of *Hind*III and re-ligated into C28pGreen-α-0.9, producing C28pGreen-α-2.3. In a similar an approximately 550 bp *Kpn*I-*Sal*I fragment of the beta-satellite C28β [GenBank HF549185] was cloned into pGreen0029, producing C28pGreen-β-0.5. Again, full length beta-satellite from the clone C28β was released with digestion of *Kpn*I and re-ligated into C28pGreen- β-0.5, producing C28pGreen-β-1.9.

Similar strategy were also used for construction of partial tandem repeats constructs corresponding to CLCuKoV-La (named C49p-Green-A-4.9) and its cognates CLCuMA (C49pGreen-α-2.3) and CLCuMB (C49pGreen-β-1.9.). Partial tandem repeats constructs produced in pGreen0029 and an empty pGreen0029 victor were finally transferred into *Agrobacterium tumefaciens* strain C58C1 by electroporation. *Agrobacterium* cultures were prepared and the plants were inoculated as described previously [36].

Results

Begomovirus detection

During the survey, high incidence of whiteflies was observed in the field. The plants displayed symptoms of CLCuD similar to the earlier reports [36-41], like vein thickening, leaves curl downward or upward with stunted plant growth due to reduction of inter-nodal distance (in severe attack), cup shape outgrowth on the lower side of the curled leaves (leaf enation; in severe disease conditions). Overall, out of sixty samples, fifty four samples were found positive with set of primers CLCV1/2 in the diagnostic PCR. For amplification of full length virus with abutting primer sets BurNF/BurNR and BurXF/BurXR, a PCR product of approximately 2.8 kb was amplified, cloned and sequenced. Ten full length DNA-A were amplified from cotton samples collected in different district across the cotton belt (Table 1A). In genome organization, the sequences were of typical of *Begomovirus*, 2759 nucleotides in length (C28A, C32A, C47A, C50A, C52A, C55A, C58A, C62A and C64A), whereas, one clone (C49A) was determined to be 2751 bp in length [42]. PCR amplifications with a pair of universal primers for the amplification of alpha-satellite (AlphaF/AlphaR) [29] and beta-satellites (Beta01/Beta02) [30] were positive for all samples yielding the expected 1350 nucleotide product. The expected band size satellite molecules were cloned and sequenced. They were of a length typical of alpha- and beta-satellites, ranging between 1349 to 1374 nucleotides (Table 1B).

Plant Sample	Origin	Begomovirus DNA-A							
		Acc. No.	No. of Ntds	Position of gene. Predicted size (no. of amino acid)/mol. weight (kDa)						
				V2	CP	Rep	REn	TrAP	C4	C5
C28	Faisala-bad	HF549180	2759	132-488 118/13.7	292-1062 256/29.7	2596-1505 363/40.7	1463-1059 134/15.5	1608-1501 35/4.3	2682-2137 181/30.3	807-283 174/19.8
C32	Faisala-bad	HF549181	2759	132-488 118/13.7	292-1062 256/29.6	2596-1505 363/40.8	1463-1059 134/15.6	1608-1501 35/4.3	2682-2242 146/16.3	807-283 174/19.9
C47	Dhok Ali Khan	KR816002	2759	132-488 118/13.6	292-1062 256/29.7	2596-1505 363/40.7	1463-1059 134/15.7	1608-1501 35/4.4	2682-2242 146/16.3	807-283 174/19.9
C49	Layyah	HF549182	2751	118-474 118/13.7	278-1048 256/29.6	2588-1497 363/40.8	1455-1051 134/15.6	1600-1481 39/4.6	2674-2129 181/20.5	-----
C50	Bhakkar	HF549183	2759	132-488 118/13.7	292-1062 256/29.6	2596-1505 363/40.7	1463-1059 134/15.4	1608-1501 35/4.3	2682-2242 146/16.3	807-283 174/19.8
C52	Multan	KR815998	2759	132-488 118/13.6	292-1062 256/29.7	2596-1505 363/40.7	1463-1059 134/15.3	1608-1501 35/4.4	2682-2137 181/20.4	------
C55	Bahawal Nagar	KR815999	2759	132-488 118/13.6	292-1062 256/29.7	2596-1505 363/40.6	1463-1059 134/15.4	1608-1501 35/4.3	2682-2242 146/16.3	807-283 174/19.7
C58	Okara	HF549184	2759	132-488 118/13.7	292-1062 256/29.6	2596-1505 363/40.7	1463-1059 134/15.4	1608-1501 35/4.3	2682-2242 146/16.3	807-283 174/19.8
C62	Khanewal	KR816000	2759	132-488 118/13.5	292-1062 256/29.7	2596-1505 363/40.6	1463-1059 134/15.3	1608-1501 35/4.2	2682-2242 146/16.2	-----
C64	Rahim Yar Khan	KR816001	2759	132-488 118/13.6	292-1062 256/29.6	2596-1505 363/40.7	1463-1059 134/15.4	1608-1501 35/4.3	2682-2242 146/16.3	807-283 174/19.8

Table 1A: *Origins of virus isolates and features of the cotton infecting Begomovirus.*

Plant Sample	Origin	Alpha-satellite			DNA Beta-satellite		
		Acc. No.	No. of Ntds	Position of gene. Predicted size (no. of amino acid)/mol. weight (kDa) Rep	Acc. No.	No. Ntds	Position of βC1 gene. Predicted size (no. of amino acid)/ mol. weight (kDa) βC1
C28	Faisalabad	KR816016	1367	77-1024 (315/36.6)	HF549185	1351	550-194 (118/13.6)
C32	Faisalabad	-----	-----	------	HF549186	1349	551-195 (118/13.6)
C47	Dhok Ali Khan	KR816017	1367	77-1024 (315/36.6)	KR816003	1350	551-195 (118/13.7)
C49	Layyah	KR816008	1365	77-1024 315/36.7)	HG000665	1350	551-195 (118/13.7)
C50	Bhakkar	KR816009	1365	77-1024 (315/36.6)	HF549187	1351	551-195 (118/13.7)
C52	Multan	KR816010	1365	77-1024(315/36.5)	KR816004	1349	551-195 (118/13.6)
C55	Bahawal Nagar	KR816011	1366	77-1024 (315/36.6)	KR816005	1351	551-195 (118/13.7)
C58	Okara	KR816012	1364	77-1024(315/36.5)	HF549188	1371	551-195(118/13.6)
C62	Khanewal	KR816013	1369	77-1024(315/36.7)	KR816006	1372	551-195 (118/13.6)
C64	Rahim Yar Khan	KR816014	1374	77-1024 (315/36.8)	KR816007	1350	551-195 (118/13.7)

Table 1B: *Origins and features of alpha-satellite and beta-satellite molecules associated with cotton infecting Begomovirus.*

Sequence analysis

The isolates C28A, C32A, C47A, C49A, C50A, C52A, C55A, C58A, C62A and C64A showed the genome organization typical Burewala strain in the Old World - with two ORFs (V2 and CP) on virion strand and four ORFs (Rep, REn, TrAP and C4) on complementary strand. They have large intergenic region (LIR) containing a putative hairpin loop structure with highly conserved sequence, the nonanucleotide (TAATATT/AC), as per previous report [22]. The genomic coordinates of ORFs are listed in the table 1A. Further analysis revealed that all isolates have truncated TrAP gene, due to mutation at a different position making the protein terminate prematurely, encoding a smaller 35 aa product. The 2nd termination codon occurs 3 amino acids downstream of the first and again an in-frame methionine codon is present 21 amino acids before the usual termination codon for the TrAP gene. All isolates identified here are categorized in group 2S (2 stop codons in the TrAP), well described by Amaro., *et al.* [22], whereas in C49A, it encodes 39 aa product (initiating from the start codon for TrAP) [42].

Initial nucleotide pairwise sequence analysis of identified isolates showed a high percentage nucleotide identity (96.2-98.6%) among each other and 96.4 to 98.2% with the available CLCuKoV-Bu sequences in the database from Pakistan and India [22-25] (Table 2A). This showed them to be variants of CLCuKoV-Bu, except for isolate C49A, based on the present pertinent species demarcation threshold for begomoviruses (>94%) according to Brown., *et al.* [30] (Table 2A). The isolate C49A showed that 53 nucleotides are missing from ori with 3'UTR in the start of the intergenic region and a high variable region of approximately 430 nts from position 874 to 1300. The missing portion (53 nts) and high variable region (430 nts) were blast separately. They showed high percentage nucleotide identities (98-99%) with CLCuMuV-[Philp:G2:2013] KF413616 and (98%) with CLCuMuV-[CN:JS-Hi-3-2:2012] JX914662, both reported from *Hibiscus rosa-sinensis* Linn, respectively.

Isolates DNA A Components	ToLCV-[IN-Ban-Chi-08]	PaLCuV-PK[PK-Cot-02]	CLCu-GeV-EG[EG-Cai-Okr]	CLCuKoV-Sha[PK-TanA-05]	CLCu-BaV-[IN-Ban-04]	CLCuAlV-Al[PK-Ko-h802a-96]	CLCuKoV-Ko [PK-Man806b-96]	CLCuMuV-Fai[CN-Fz1-Hib-12]	CLCuKoV-Bu[IN-Sri-Gang-05	C28A, C32A, C47A, C50A, C52A, C55A, C58A, C62A, C64A	C49A
C49A	79.8	80.2	73.2	91.7	85.2	79.1	84.8	89.9	92.1	87.7-92.1	100
C28A, C32A, C47A, C50A, C52A, C55A, C58A, C62A, C64A	80.1-81.7	81.6-83.1	73.2-73.8	89.4-90.6	83.8-84.4	75.6-76.6	88.2-89.4	89.3-90.6	96.4-98.2	100	
CLCuKoV-Bu[IN-SriGang-05]	81.3	82.4	73.7	89.6	83.7	75.8	87.5	90.4	100		
CLCuMuV-Fai[CN-Fz1-Hib-12]	76.8	75.6	73.7	85.5	85.2	83.1	79.1	100			
CLCuKoV-Ko[PK-Man806b-96]	85.2	86.2	75.9	91.4	81.2	74.6	100				
CLCuAlV-Al[PK-Koh802a-96]	75.1	79.2	70.4	77.9	79.2	100					
CLCuBaV-[IN-Ban-04]	79.5	76.7	74.3	85.4	100						
CLCuKoV-Sha[PK-TanA-05]	86.6	80.6	74.7	100							
CLCuGeV-EG[EG-Cai-Okr]	72.9	71.6	100								
PaLCuV-PK[PK-Cot-02]	83.4	100									
ToLCV-[IN-Ban-Chi-08]	100										

Table 2A: *Percentage nucleotide sequence identity between the complete sequences of the DNA As of isolates C28, C32, C47, C49, C50, C52, C55, C58, C62 and C64 with selected begomoviruses in the databases. Sequences were selected of the most similar species from the Blast comparison.*

Comparing the ten isolates, C49A was distinct from the others. The BLAST of its full length sequence hit into two separate portions, minor portion (from position 54 to 874) showed a high percentage nucleotide identity with available CLCuKoV (97%). The major portion of isolate C49 was 98% identical to CLCuMuV-[PK:62:05] and CLCuMuV-[PK:62:05], which were the virulent strain during the first epidemic on cotton in Pakistan. Further insights were to identify the TATA box and the iterons position in the virus genome, where Rep protein binds for the initiation of replication. Nine out of ten isolates showed the position of TATA box at 26462-2665 and iterons at 2645-2648, while TATA box position in the isolate C49A was found at 2657-2657 and iterons at 2637-2640 at the 5' of the TATA Box.

Along with the helper viruses, the associated beta-satellites component were also sequenced, showing characteristics of cotton leaf curl Multan beta-satellite, previously reported to be associated with CLCuD. Pairwise similarities analysis reveals high percentage nucleotides identities (87.6-99%) of all beta-satellite molecules with each other, as well as, 89.6 to 92.2 with the sequence of CLCuMB reported in the database (Table 2B). Furthermore, the results also showed that the beta-satellite molecules here are the variants of CLCuMB [29] having one ORF of 118 amino acid in the complementary strand required for pathogenicity, A-rich region and a putative hairpin loop with the conserved nonanucleotide (TAATATTAC) sequence.

Following, the complex were also associated with an alpha-satellite. An initial comparison of the these sequences obtained with sequences in the databases using BAST indicated them to be most similar to begomoviruses associated alpha-satellite previously identified in cotton and other hosts of *Begomovirus* in the old world [43]. Nucleotide sequences identity with selected alpha-satellites from databases, showed them to be a variants of CLCuMA, having a single ORF of 315 amino acid on the virion stand required for replication, A rich region and a putative hairpin loop with a conserved TAGTATTAC sequence (Table 2C).

Isolates of beta-satellite	PaLCuB-[IN-Coi-BG-CBE-12]	ChiLCB-[IN-Joh-04]	CrYVMB-[IN-Bhub-OYB-HU-06]	AYLCB-[IN-WS-FBI-10]	MaYVB-[CN-Yun-Y217-05]	ToYLCThB-[IN-Ramg-07]	CLCuMB-[IN-bt18]	C28β, C32β, C47β, C49β, C50β, C52β, C55β, C58β, C62β, C64β
C28β, C32β, C47β, C49β, C50β, C52β, C55β, C58β, C62β, C64β	39.3-40.6	39.0-40.5	37.8-42.7	36.7-38.8	60.1-64.2	39.4-42.3	89.6-92.2	87.6-99.0
CLCuMB-[IN-bt18]	38.3	38.9	41.1	37	60.7	39.3	100	
ToYLCThB-[IN-Tamg-07]	45.4	47.4	51.6	47.9	38.4	100		
MaYVB-[CN-Yun-Y217-05]	36.1	38.7	40.1	36.3	100			
AYLCB-[IN-WSFBI-10]	45.7	55.7	47.9	100				
CrYVMB-[IN-Bhub-OYB-HU-06]	52.7	45.6	100					
ChiLCB-[IN-Joh-04]	53.1	100						
PaLCuB-[IN-Coi-BG-CBE-12]	100							

Table 2B: *Percentage nucleotide sequence identity between the complete sequences of the beta-satellite of isolates C28β, C32β, C47β, C49β, C50β, C52β, C55β, C58β, C62β and C64β with selected beta-satellites in the databases. Sequences were selected of the most similar species from the Blast comparison.*

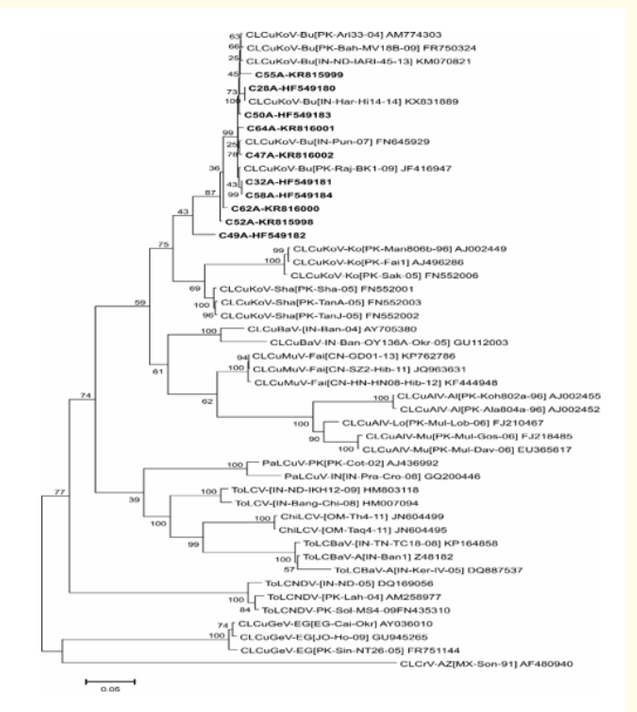
Isolates	ToLCA-[IN-LN-11]	SiYVCNA-[CN-Y340-10]	MalYVMA-[CN-Hn39-06]	GDaSA-[PK-Tom-2-06]	MeYMA-[IN-10b-RCA-al-F-07]	CLCShA-[PK-DS C07]	CLCuMA-[PK-AS1-01]	C28α, C47α, C49α, C50α, C52α, C55α, C58α, C62α, C64α
C28α,C47α,C49α, C50α, C52α, C55α, C58α, C62α, C64α	69.7-72.1	67.2-70.8	67.6-70.7	67.1-71.0	74.0-78.1	75.1-79.0	83.9-88.9	87.5-100
CLCuMA-[]PK-AQ16-11	71.3	70.4	69.8	69.9	76.6	77.7	100	
CLCuShA-[PK-DS C07]	65.0	76.2	73.9	77.9	80.9	100		
MeYMA-[IN-10b0R-CA-al-F-07]	69.3	70.2	73.3	72.6	100			
GDaSA-[PK-Tom-2-06]	64.9	73.8	68.5	100				
MalYVMA-[CN-Hn39-06]	65.6	74.6	100					
SIYVCNA-[CN-Y340-10]	76.2	100						
ToLCA-[IN-LN-11]	100							

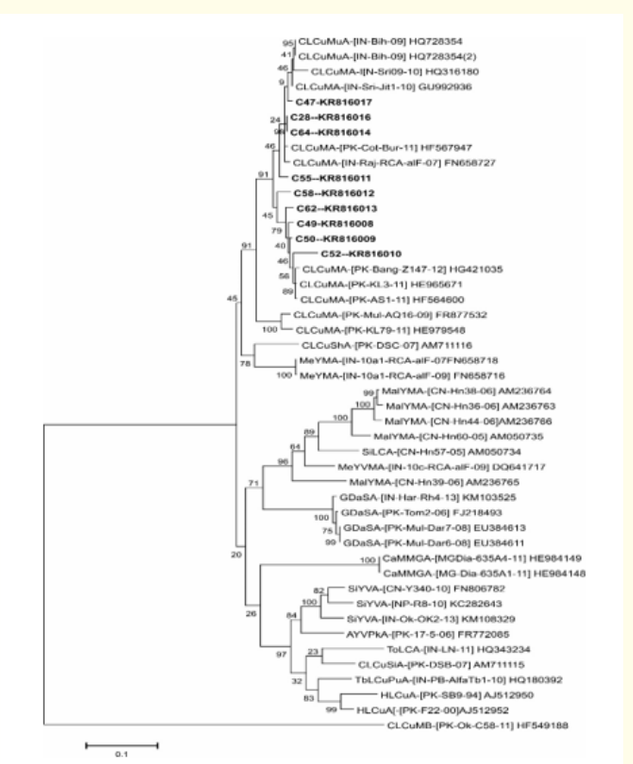
Table 2C: *Percentage nucleotide sequence identity between the complete sequences of alpha-satellite isolates C28α, C47α, C49α, C50α1, C502α, C52α, C55α, C58α, C62α and C64α with selected alpha-satellites in the database. Sequences were selected of the most similar species from the blast comparison.*

Phylogenetic relationship

Cotton is infected by several begomoviruses. Phylogenetic tree was constructed, using full length DNA A sequences with other selected sequences from databases. The most similar sequences in the databases were included as separated clades, indicating species confirmation (Figure 1A).

Furthermore, alpha- and beta-satellites with other available similar satellite sequences in the databases that were found to infect cotton and other economical crops in the surrounding were aligned. Based on high identity with the available sequences in the databases, different clades were made in coherence with species cut-off demarcation for alpha-satellite and beta-satellite [29,44] (Figure 1B and 1C).





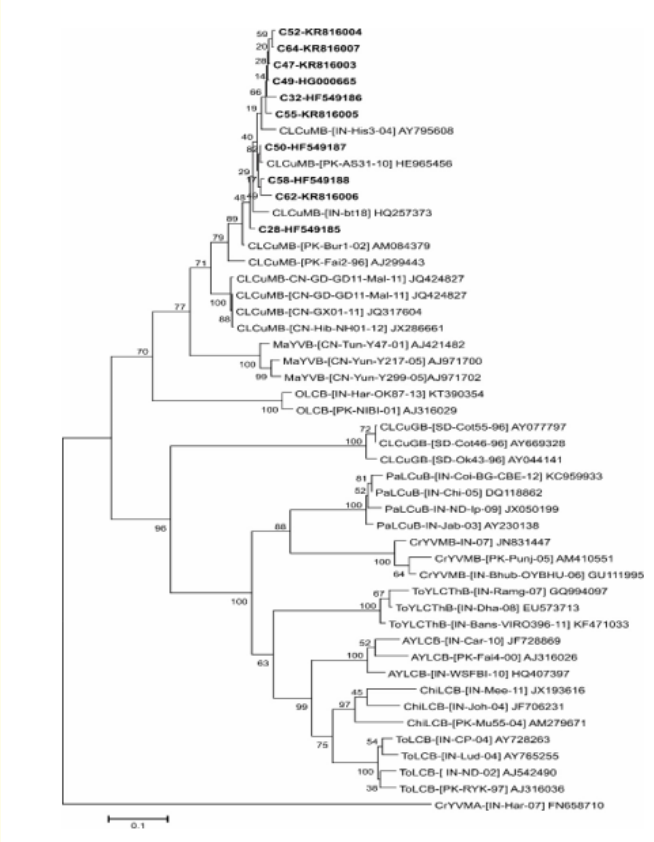


Figure 1: *Phylogenetic analysis of begomoviruses, and its associated alpha-satellites and beta-satellites. A, B and C: Cotton infecting Begomviruses, alpha- and beta-satellite identified here with similar sequences from databases were aligned using MUSCLE and tree was constructed with help of Maximum Likehood was used, embedded in Omega 6 In each case, the isolate descriptor and accession number are given. Begomoviruses and its associated alpha and beta-satellites cloned and characterized here are shown in bold. The numbers at the nodes indicates the bootstrap confidence valves (1000 replicates). The alignment was arbitrary rooted on an out-group the complete Begomovirus of Cotton leaf crumple virus (CLCrV-AZ[MZ-Son-91]), alpha-satellite of Croton yellow vein mosaic alpha-satellite (CrYVMA-[IN-Har-07]), and beta-satellite sequence of Cotton leaf curl Multan beta-satellite (CLCuMB-[PK-Fai-C58-11]) a distantly related Begomovirus associated satellite originating from the New and Old World*

Infectivity studies

Two complete clones (helper viruses with cognate alpha- and beta-satellite), that are, CLCuKoV-Bu[PK-C28-11] HF549180, CLCuMA-[PK-C28-11] KR816016, CLCuMB-[PK-C28-11] HF549185 isolated from Faisalabad region and CLCuKoV-La[PK-C49-12] HF549182, CLCuMA-[PK-C49-12] KR816008 and CLCuMB-[PK-C49-12] HG000665 isolated from Layyah region were proceeded for partial tandam repeat constructs. Clones were used to inoculate *Nicotiana benthamiana, N. tabacum, Lycopersicum esculentum, Cucurbita pepo* and *Cucumis sativus* plants using *Agrobacterium tumefaciens*. After 15dpi (Days post inoculation), light symptoms were observed like downward curling of leaves and yellowing of veins (Figure 2). As the more days passed, these symptoms were appeared more clearly and severely.

New leaves were highly curl and deformed with small size and overall growth of the plant in comparison with the healthy plants remained smaller and shunted. Inoculated plants with C28pGreen-A-4.9, C28pGreen- induced more severe symptoms and ceased the growth than that of C49 partial dimers. The possible reason may be short trap which is 35 amino acid in C28 while 39 amino acid in C49 or, maybe, there is some role of intergeneric sequences, as discussed earlier for C49 [42].

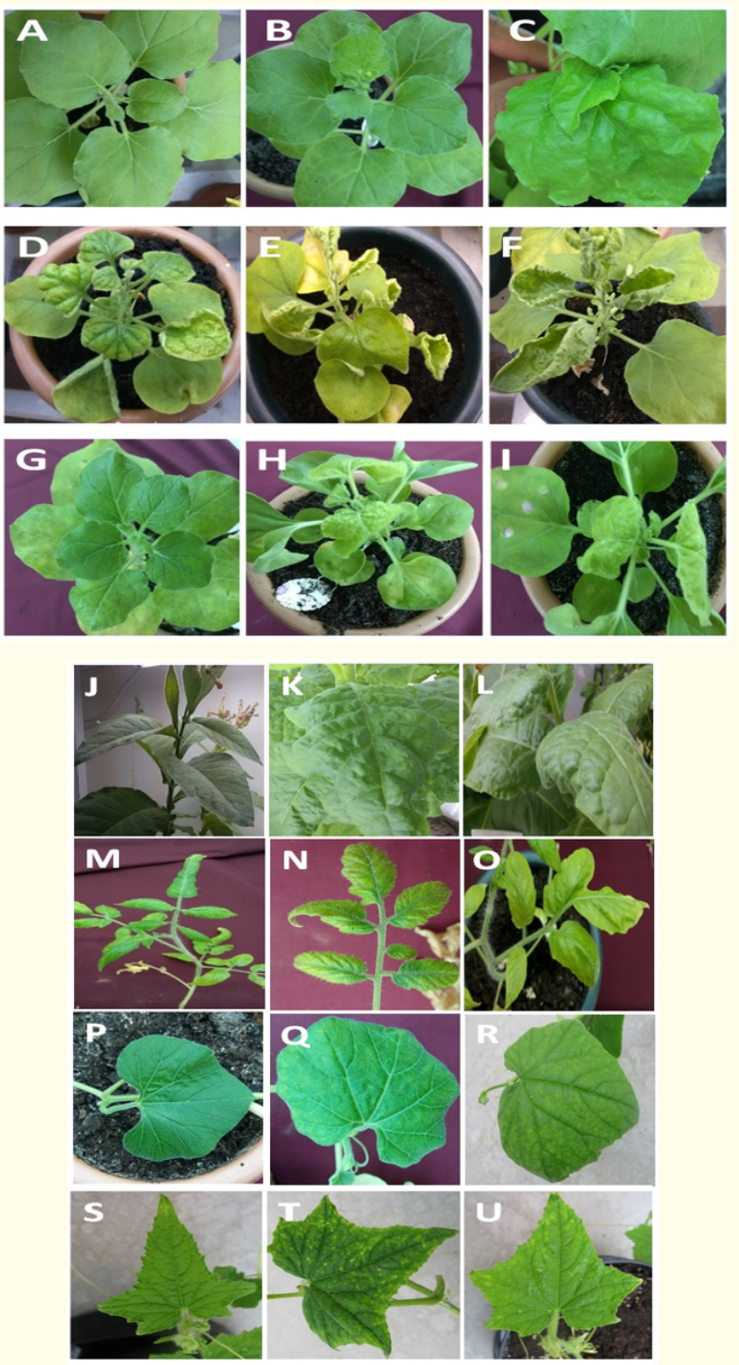
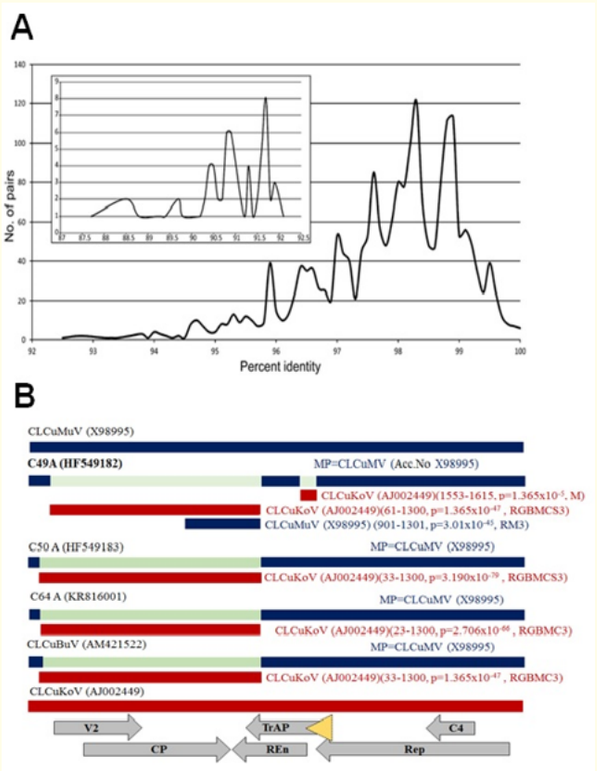
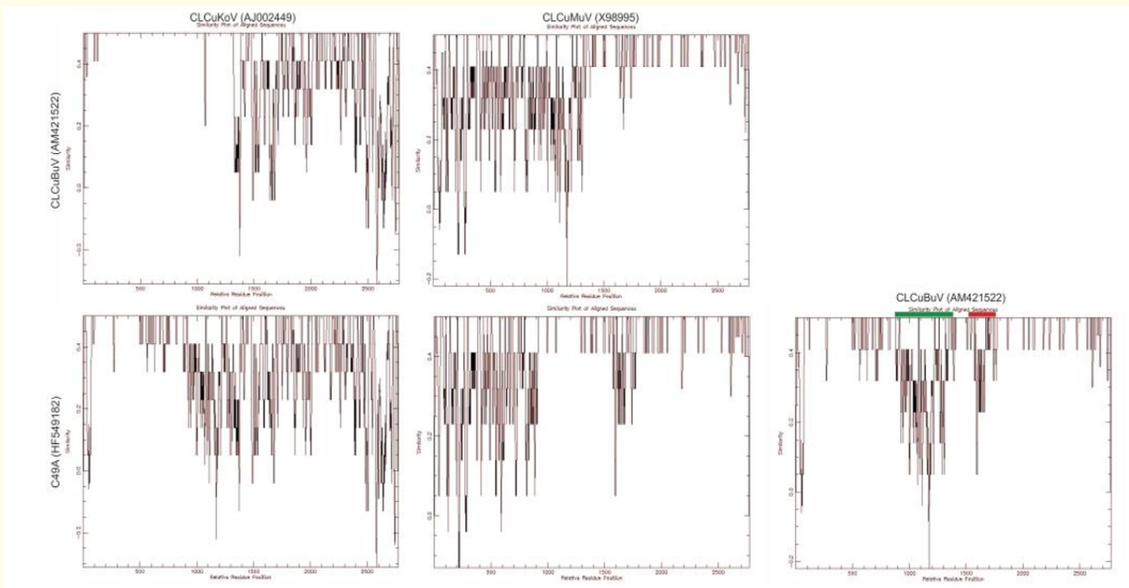


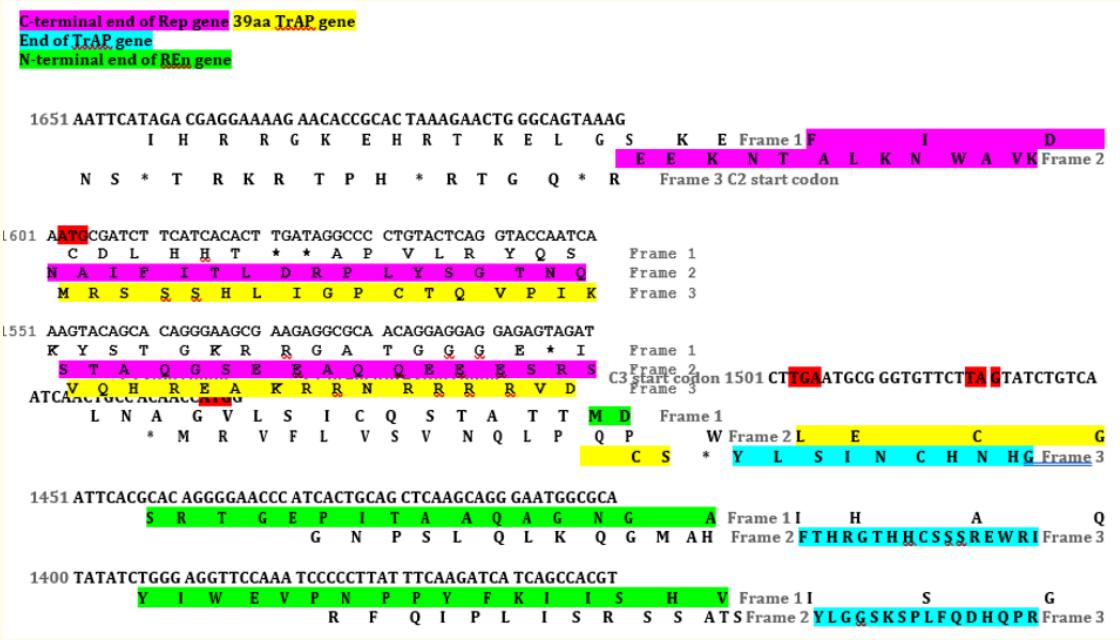
Figure 2: *Infectivity analysis of partial dimers of Burewala virus isolates inoculating Nicotiana benthamiana, Nicotiana tabacum, Lycopersicum esculentum, Cucurbita pepo and Cucumis sativus plants. Agro-inoculation of C28pGreen-A-4.9 (C, J, M, P and S), C28pGreen-A-4.9 and C28pGreen-β-1.9 (D, E, H, K, N, Q, and T), C28pGreen-A-4.9, C28pGreen-β-1.9 and C28pGreen-α-2.3 (F, I, L, O, R and U), C49A (G), C49pGreen-A-4.9, and C49pGreen-β-1.9 (H) C49pGreen-A-4.9, C49pGreen-β-1.9 and C49pGreen-α-2.3 (I), mock plant (B) and control healthy plant (A).*



Supplementary Figure 1: *Sequence relatedness of C49A to CLCuBuV and recombination analysis. A. Graphical illustration of pairwise nucleotide sequence identities of 66 CLCuBuV isolates and (inset) pairwise comparison of the sequence of C49A to 66 CLCuBuV isolates. B. RDP analysis of C49A, C50A, C64A and CLCuBuV. For each isolate the database accession number is given. The major parent (MP) for each recombinant sequence is indicated. For each recombinant fragment the origin (minor parent), position (nucleotide coordinates of the 5' and 3' breakpoints), highest p-value and methods supporting the event are indicated. The methods are RDP (R), GENECONV (G), Bootscan (B), Max Chi (M), Chimera (C), SiScan (S) and 3Seq (3).*



Supplementary Figure 2: *Pairwise Plotcon comparisons of the sequence relatedness of C49A, CLCuMuV, CLCuKoV and CLCuBuV. The database accession number of each isolate used in the comparison is given. For the C49A comparison with CLCuBuV the green bar highlights the sequences differing between the two isolates that suggest the recombinant fragment originating from CLCuMuV is smaller for C49A than for CLCuBuV. The red bar indicates the possible second recombinant fragment which RDP3 suggests originates from CLCuMuV.*



Supplementary Figure 3: *Sequences surrounding the overlap of the Rep,TrAP and REn genes showing truncated TrAP gene. Sequence of C49A surrounding the overlap of the Rep, TrAP and REn genes highlighting coding sequences, start and stop codons (as indicated by the key at the top). Note the truncated (predicted 39 amino acid coding capacity) TrAP gene.*

Note: Sequence is in complementary sense orientation.

Discussion

From the ancient times, cotton is grown on Indian subcontinent used for food, feed, fiber and other purposes. *Gossypium barbadense* has been grown native to this region. However, the production of this local cotton was not of high ginning quality. For a good production/high yielding, a new cotton cultivar was introduced from the New World. Although yield was increased into so many folds, it got susceptible to diseases like CLCuD. The disease was ignored until it appeared as epidemic and the production was affected badly. Till the first epidemic, 6 - 7 species of begomoviruses were thought to infect cotton as single species or in mix infection. Now the situation has become more complicated in Pakistan and northwestern India, as the number of species has increased up to ten. The increase in newly emerging virus species may be due to point mutations and specifically due to genetic recombination such as CLCuBuV, CLCuShV [22,45] and PMLCuV [46] and the virus from other dynasties, i.e. CLCuGeV infecting cotton in south of Pakistan [47]. This elucidated the conducive environment for begomoviruses and the insect carrier whiteflies.

During our survey, a high load of whitefly and severe symptoms on cotton farms has been observed. This elucidated a much bigger vista of CLCuD. Maximum plants were found to be infected with whitefly transmitted begomoviruses displaying disease symptoms like stunted growth, upward/downward leaf curl, with a prominent thick and dark veins and one to several leaf enations. Interestingly, virus sequenced from the infected sampled plants were all "Burewala strain" infecting cotton in Punjab. More interestingly, Multan strain, Rajasthan strain, Kokhan strain were not found in any plant sample. Along with, a new recombinant Layyah strain was found. From all symptomatic cotton, showing the dominancy of resistance breaking CLCuKoV-Bu species in this region at the moment. Previous pre-dominant species like CLCuMuV and CLCuKoV were not found from last so many years, reflecting the virus is not present anymore or switched to other plant species. Similar reports were also published regarding CLCuKoV-Bu from India [23,25].

To date, CLCuKoV-Bu is considered to be the resistant breaking since 2001. Previous reports highlighted Burewala strain to be the major concern for the low production of cotton from Pakistan and India. Every year a high reduction in number of cotton bales are reported which effects the value added by cotton to the total GDP. In the line with, no resistant cotton variety has been produced. Many attempts to produce transgenic plants through RNAi techniques have been flourished but could not yield a complete success to date. The TrAP gene, which has a multifunctional potential, is involved to enhance the transcription activity or can override the defense mechanism of the host. The most striking feature of Burewala strain is the intact TrAP gene to transcribe 35aa of a bona-fide codon from the start instead of 150aa of the total gene. This sort of mutation in the gene divert the concept to believe how it perform it's compete activity with this short and mutated protein to control host defense system during systemic infection, is still a big mystery. May be the sequence that has been deleted made the virus escape from the silencing screens of their hosts. Our analysis of the diversity of begomoviruses associated with cotton obviates a recombinant (C49 isolate), consisting of sequences derived from CLCuMuV and CLCuKoV, two viruses that were prevalent prior to 2001. This provides further evidence of the importance of recombination in the evolution of geminiviruses [14,48-54]. The virus is sufficiently distinct from all other begomoviruses for it to be considered a new stain [42].

Among cotton infecting viruses, CLCuKoV-Bu isolates are the most notorious in terms of severity of the disease and are the most widespread isolates across Pakistan and northwestern India [6]. Infectivity studies of this new strain reveals to be infectious like Burewala strain and also found compatible with CLCuMA and CLCuMB. The incidence of this new strain in cotton and other associated crops is under way.

Conclusion

Based on present delineated results showed the higher dominancy of a single strain CLCuKoV-Bu to CLCuD across Pakistan, and solely responsible for the outbreak in 2001. After second epidemic in Pakistan, CLCuKoV-Bu continuously spreading to other regions of cotton belt, while predominant species like CLCuMV and CLCuKoV switching from cotton to other crops. Likewise, CLCuKoV-Bu strain, CLCuKoV-La is also a recombinant with distinct recombinant origins and represents a newly identified strain in the Kokhran species. Furthermore, this new strain is also efficient to aggregate symptoms in *Agrobacterium* mediated inoculation.

Bibliography

1. Gutierrez C. "Replication and gene expression of DNA viruses". In: Khan JA, editors DJ, editors. Handbook of plant virology. New York: Food Products (2006): 75-89

2. Navas-Castillo J., *et al*. "Emerging virus diseases transmitted by whiteflies". *Annual Review of Phytopathology* 49 (2011): 219-248.

3. Varsani A., *et al*. "Capulavirus and Grablovirus: two new genera in the family Geminiviridae". *Archives of Virology* 162.6 (2017): 1819-1831.

4. Varma A and Malathi VG. "Emerging geminivirus problems: A serious threat to crop production". *Annals of Applied Biology* 142.2 (2003): 145-164.

5. Sattar N., *et al*. "Cotton leaf curl disease - an emerging threat to cotton production worldwide". *Journal of General Virology* 94.4 (2013): 695-710.

6. Qadir R., *et al*. "Diversity and recombination analysis of Cotton leaf curl Multan virus: a highly emerging begomovirus in northern India". *BMC Genomics* 20 (2019): 274.

7. Lazarowitz SG and Shepherd R. "Geminiviruses: genome structure and gene function". *Critical Review in Plant Sciences* 11.4 (1992): 327-349.

8. Hanley-Bowdoin L., *et al*. "Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation". *Critical Review in Plant Sciences* 18.1 (1999): 71-106.

9. Argüello-Astorga GR., *et al*. "Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication". *Virology* 203.1 (1994): 90-100.

10. Arguello-Astorga GR., *et al*. "High frequency reversion of geminivirus replication protein mutants during infection". *Journal of Virology* 81.20 (2007): 11005-11015.

11. Fontes E., *et al*. "Interaction between a geminivirus replication protein and origin DNA is essential for viral replication". *Journal of Biological Chemistry* 269.11 (1994): 8459-8465.

12. Fontes E., *et al*. "Geminivirus replication origins have a modular organization". *Plant Cell* 6.3 (1994): 405-416.

13. Briddon., *et al*. "Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses". *BMC Evolutionary Biology* 10 (2010): 97.

14. Polston JE and Anderson PK. "The emergence of whitefly-transmitted geminiviruses in tomato in the Western Hemisphere". *Plant Disease* 81.12 (1997): 1358-1369.

15. Padidam M., *et al*. "Possible emergence of new geminiviruses by frequent recombination". *Virology* 265.2 (1999): 218-225.

16. Jeske H., *et al*. "DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus". *EMBO Journal* 20.21 (2001): 6158- 6167.

17. Schnippenkoetter WH., *et al*. "Forced recombination between distinct strains of Maize streak virus". *Journal of General Virology* 82.12 (2001): 3081-3090.

18. Lefeuvre P., *et al*. "Begomovirus 'melting pot' in the south-west Indian Ocean islands: molecular diversity and evolution through recombination". *Journal of General Virology* 88.12 (2007): 3458-3468.

19. Briddon RW and Markham PG. "Cotton leaf curl disease". *Virus Research* 71.1-2 (2000): 151-159.

20. Hussnain T and Ali M. "A review of cotton diseases of Pakistan". *Pakistan Cotton* 19 (1975): 77-78.

21. Rahman M., *et al*. "Genetics of resistance against cotton leaf curl disease in Gossypium hirsutum". *Plant Pathology* 54.6 (2005): 764-772.

22. Mansoor S., *et al*. "Breakdown of resistance in cotton to cotton leaf curl disease in Pakistan". *Plant Pathology* 52.6 (2003a): 784.

23. Amrao L., *et al*. "Cotton leaf curl disease in resistant cotton is associated with a single begomovirus that lacks an intact transcriptional activator protein". *Virus Research* 152.1-2 (2010b): 153-163.

24. Rajagopalan P., *et al*. "Dominance of resistance-breaking cotton leaf curl Burewala virus (CLCuBuV) in northwestern India". *Archives of Virology* 157.5 (2012): 855-868.

25. Ashraf MA., *et al*. "Molecular characterization and phylogenetic analysis of a variant of highly infectious cotton leaf curl Burewala virus associated with CLCuD from Pakistan". *Australian Journal of Crop Science* 7.8 (2013): 1113-1122.

26. Zaffalon V., *et al*. "A survey of geminiviruses and associated satellite DNAs in the cotton-growing areas of northwestern India". *Archives of Virology* 157.3 (2012): 483-495.

27. Doyle JJ and Doyle JL. "Isolation of plant DNA from fresh tissue". *Focus* 12 (1990): 13-15.

28. Mansoor S., *et al*. "Association of a begomoviruse and nanovirus like molecules with Ageratum yellow vein disease in Pakistan". *Plant Disease* 84.1 (2000a): 101.

29. Zia-ur-Rehman M., *et al*. "First Detection of Cotton leaf curl Burewala virus and Cognate Cotton leaf curl Multan betasatellite and Gossypium darwinii symptomless alphasatellite in Symptomatic Luffa cylindrical in Pakistan". *Disease Note* 97.8 (2013): 1122.

30. Briddon RW., *et al*. "Universal primers for the PCR-mediated amplification of DNA _; a molecule associated with some monopartite begomoviruses". *Molecular Biotechnology* 20.3 (2002): 315-318.

31. Brown JK., *et al*. "Revision of Begomovirus taxonomy based on pairwise sequence comparisons". *Archives of Virology* 160.6 (2015): 1593-1619.

32. Thompson JD., *et al*. "The Clustal X windows interface; flexible strategies for multiple sequence alignment aided by quality analysis tools". *Nucleic Acids Research* 25.24 (1997): 4876-4882.

33. Page RDM. "TREEVIEW: An application to display phylogenetic trees on personal computers". *Computer Applications in the Biosciences* 12.4 (1996): 357-358.

34. Rice P., *et al*. "EMBOSS: the European Molecular Biology Open Software Suite". *Trends in Genetics* 16.6 (2000): 276-277.

35. Martin DP and Rybicki EP. "RDP: detection of recombination amongst aligned sequences". *Bioinformatics* 16.6 (2000): 562-563.

36. Hussain M., *et al*. "The hypersensitive response to tomato leaf curl New Delhi virus nuclear shuttle protein is inhibited by transcriptional activator protein". *Molecular Plant-Microbe Interactions* 20.12 (2007): 1581-1588.

37. Mansoor S., *et al*. "Molecular characterization of a Geminivirus associated with cotton leaf curl disease in Pakistan". Proceeding of International Symposium on Biotechnology and Sustainable Development. NIBGE, Faisalabad (1993).

38. Harrison BD., *et al*. "Detection and relationship of cotton leaf curl virus and allied white fly transmitted Geminiviruses accruing in Pakistan". *Annals of Applied Biology* 130.1 (1997) 61-75.

39. Rehman HU., *et al*. "Stability of cotton cultivars under leaf curl virus epidemic in Pakistan". Field Crop Research 69.3 (2001): 251-257.

40. Briddon RW., *et al*. "Identification of DNA Components Required for Induction of Cotton Leaf Curl Disease". *Virology* 285.2 (2001): 234-243.

41. Qazi J., *et al*. "Contribution of the satellite encoded gene β C1 to cotton leaf curl disease symptoms". *Virus Research* 128.1-2 (2007): 135-139.

42. Monga D., *et al*. "Cotton leaf Curl Disease in India-recent status and management strategies". Presented in 5th meeting of Asian Cotton Research and Development Network Held in Lahore (2011).

43. Shuja MN., *et al*. "Identification of a distinct strain of Cotton leaf curl Burewala virus". *Archives of Virology* 159.10 (2014): 2787-2790.

44. Briddon RW., *et al*. "Diversity of DNA 1: a satellite-like molecule associated with monopartite begomovirus - DNA h complexes". *Virology* 324.2 (2004) 462-474.

45. Mubin M., *et al*. "Complete nucleotide sequence of chili leaf curl virus and its associated satellites naturally infecting potato in Pakistan". *Archives of Virology* 154.2 (2009): 365-368.

46. Amrao L., *et al*. "Cotton leaf curl disease in Sindh Province of Pakistan is associated with recombinant begomovirus components". *Virus Research* 153.1 (2010a): 161-165.

47. Nawaz-ul-Rehman MS., *et al*. "Melting Pot of Old World Begomoviruses and Their Satellites Infecting a Collection of Gossypium Species in Pakistan". *PLoS ONE* 7.8 (2012): e40050.

48. Tahir MN., *et al*. "The merging of two dynasties-identification of an african cotton leaf curl disease-associated begomovirus with cotton in Pakistan". *PLoS ONE* 6.5 (2011): e20366.

49. Nateshan HM., *et al*. "Host range, vector relations and serological relationships of cotton leaf curl virus in South India". *Annals of Applied Biology* 126.2 (1996): 233-244.

50. Navas-Castillo J., *et al*. "Natural recombination between Tomato yellow leaf curl virus-Is and Tomato leaf curl virus". *Journal of General Virology* 81.11 (2000): 2797-2801.

51. Sanz AI., *et al*. "Genetic variability of natural populations of cotton leaf curl geminivirus, a single-stranded DNA virus". *Journal of Molecular Evolution* 49.5 (1999): 672-681.

52. Sanz AI., *et al*. "Multiple infection, recombination and genome relationships among begomovirus isolates found in cotton and other plants in Pakistan". *Journal of General Virology* 81.7 (2000): 1839-1849.

53. Saunders K and Stanley J. "A nanovirus-like DNA component associated with yellow vein disease of Ageratum conyzoides: Evidence for interfamilial recombination between plant DNA viruses". *Virology* 264.1 (1999): 142-152.

54. Zhou X., *et al*. "Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination". *Journal of General Virology* 78.8 (1997): 2101-2111.

55. Hou YM and Gilbertson RL. "Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination". *Journal of Virology* 70.8 (1996): 5430-5436.

Volume 15 Issue 7 July 2019