Australian Journal of Crop Science AJCS
AJCS 7(8):1113-1122 (2013) ISSN:1835-2707

Molecular characterization and phylogenetic analysis of a variant of highly infectious cotton leaf curl Burewala virus associated with CLCuD from Pakistan

Muhammad Aleem Ashraf*[1], Ahmad Ali Shahid[1,2], Bahaeldeen Babiker Mohamed[1,4], Abdelhafiz Adam Dahab[1,4], Kamran Shehzad Bajwa[1], Abdul Qayyum Rao*[1], Muhammad Azmat Ullah Khan[1], Muhammad Ilyas[2,3], Muhammad Saleem Haider[2] and Tayyab Husnain[1].

[1]Centre of Excellence in Molecular Biology (CEMB), University of the Punjab, Lahore, Pakistan
[2]Institute of Agriculture Sciences (IAGS), University of the Punjab, Lahore, Pakistan
[3]International Laboratory for Tropical Agricultural Biotechnology (ILTAB), Donald Danforth Plant Science Center, St Louis, MO 63132, USA
[4]National Centre for Research (NCR), Khartoum, Sudan

*Corresponding author: maleemashraf@gmail.com, ,muhammad.aleem@cemb.edu.pk

Abstract

Cotton leaf curl disease (CLCuD) has been a considerable hindrance to producing high cotton yields and causes severe losses in the majority of cotton-growing countries, including Pakistan. Five leaves per sample per plant that exhibited characteristic CLCuD symptoms such as leaf curling, vein darkening, vein swelling and enation were collected from a farm field in Tehsil Fortabbas, district Bahawalnagar, Punjab. The genome components of CLCuBuV (DNA-A) were amplified using the rolling circle amplification (RCA) method and further characterized. The cloned genomes showed an organization typical of monopartite old world begomoviruses. Each DNA-A component of the CLCuBuV genome of the CLCuD samples was composed of 2758 bp that contained five overlapping open reading frames (ORFs) separated by a large intergenic region (LIR). The sequence analysis revealed that these are variants of the highly infectious CLCuBuV and was designated as CLCuBuV-MV12, MV13, MV14A, and MV14C clones with accession numbers FR750318, FR837932, FR837933, and FR837934, respectively. The CLCuBuV isolates under study showed an 88.5-86.5% nucleotide sequence homology to cotton leaf curl Shahdadpur virus (CLCuShV), 86.7-83.1% with cotton leaf curl Kokhran virus (CLCuKoV) and 85.5-76.2% to cotton leaf curl Multan virus (CLCuMV), begomoviruses that are associated with CLCuD in Pakistan. The phylogenetic analysis revealed that the replication associated protein (Rep), designated the AC1 ORF, the coat protein (CP), designated AV1 ORF and the single LIR of the DNA-A component of CLCuBuV genome under study, exhibited sequence identity with the CLCuMuV AC1 ORF, CLCuKoV AV1 ORF and CLCuShV LIRs respectively. The phylogenetic and recombination analysis provided conclusive evidence that the CLCuBuV isolates under consideration have a recombinant genome. It is expected that the current work will help us to understand begomovirus genetics, and evolution.

Key words: Begomovirus; *Bemisia tabaci;* Cotton leaf curl disease; Geminivirus; Recombination; Rolling circle amplification.
Abbreviation: CLCuD - cotton leaf curl disease; CLCuV -cotton leaf curl virus; CLCuBuV- cotton leaf curl Burewala virus; CLCuMuV - cotton leaf curl Multan virus; CLCuKoV - cotton leaf curl kokhran virus; CLCuAV - cotton leaf curl Alabad virus; CLCuRV - cotton leaf curl Rajasthan virus; CLCuGV - cotton leaf curl Gezira virus; PaLCuV- papaya leaf curl virus; CLCuShV - cotton leaf curl Shadadpur virus; ToLCNDV-Tomato leaf curl New Dehli virus; CLCuBV-cotton leaf curl Banglore virus; AEV-ageratum enation virus; CLCuMB-cotton leaf curl Multan betasattelite; RDP - recombination detection program; RCA - rolling circle amplification; ssDNA - single-stranded DNA; dsRF-double stranded replicative form; CR - common region; Ori-origin of replication; MEGA-Molecular evolutionary genetic analysis; NCBI-national center for biotechnology information; LIR - large intergenic region; ORF - open reading frame; RCR-rolling circle replication; CP-Coat protein; Rep-replication associated protein; REn-replication enhancer protein.

Introduction

Pakistan is the 4th largest producer of cotton in the world and a leading exporter of yarn. Cotton is susceptible to attack by several pathogens and insect pests, and among them, cotton leaf curl disease (CLCuD) is economically the most devastating to cotton production. Almost 30% of Pakistan's cotton crop is damaged and costs approximately several billion rupees annually (Asad et al., 2003). CLCuD is primarily caused by a complex of monopartite begomoviruses that belong to the geminiviridae family, which is the largest family of insect-transmitted plant viruses and is composed of 209 members with circular single-stranded (ss) genomes that are encapsidated within a characteristic twinned (geminate) icosahedral capsid. Geminiviruses are currently divided into four genera (Mastrevirus, Topovirus, Curtovirus and Begomovirus) based on genome organization, sequence identity, host plants, and

Table 1. Name, acronym, genome size and accession numbers of CLCuBuV isolates reported in this study.

Name	Acronym	Genome Size(bp) and accession numbers
		DNA-A component
Cotton leaf curl Burewala virus-MV12 (Pakistan-Fortabbas-2010)	CLCuBuV-(Pk-Fort-10)	2758 (FR750318)
Cotton leaf curl Burewala virus-MV13 (Pakistan-Fortabbas-2010)	CLCuBuV-(Pk-Fort-10)	2758 (FR837932)
Cotton leaf curl Burewala virus-MV14A (Pakistan-Fortabbas-2010)	CLCuBuV-(Pk-Fort-10)	2758 (FR837933A)
Cotton leaf curl Burewala virus-MV14C (Pakistan-Forabbas-2010)	CLCuBuV-(Pk-Fort-10)	2758 (FR837934C)

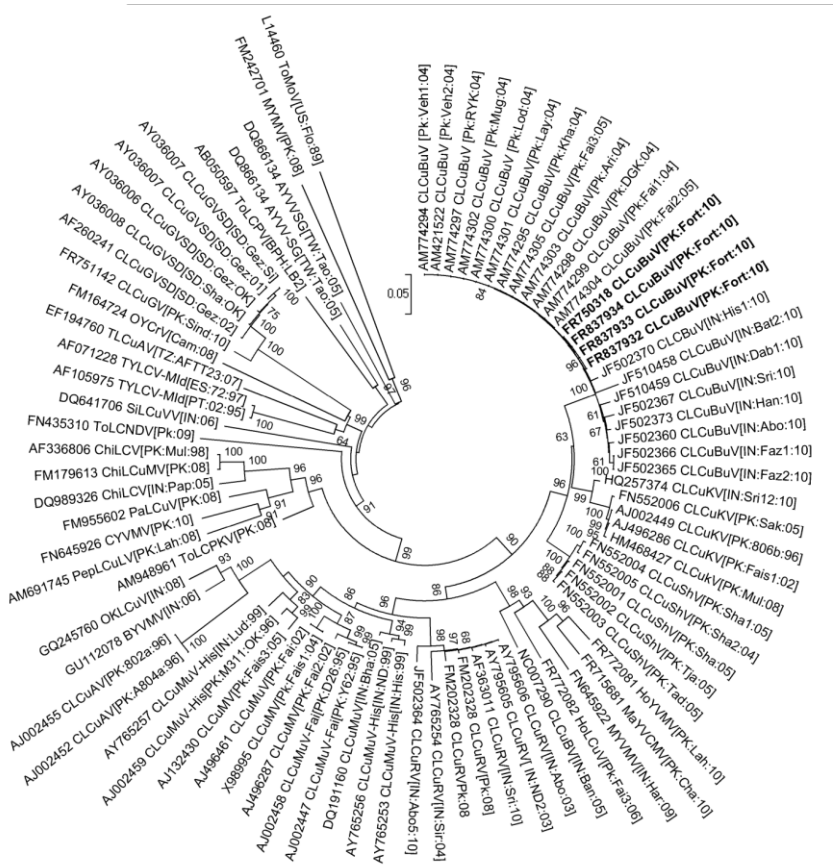


Fig 1. CLCuBuV DNA-A phylogenetic tree. A phylogenetic tree was generated using neighbor-joining analysis with MEGA software version 5.1. Numbers below branches indicate bootstrap value percentages from 1000 replicates. The scale bar represents the distance unit between sequence pairs. The sequences of CLCuBuV sequenced in present study are indicated in bold font.

insect vectors (Fauquet et al., 2008; Stanley J, 2005) The genus begomovirus is composed of 185 species of plant viruses that cause agricultural economic disasters. These viruses are transmitted exclusively by an insect vector, *Bemisia tabaci* (Gennadius), in a persistent circulative manner. Their spread may be directly linked to the inadvertent worldwide dissemination of the "B" or silverleaf biotype. Begomoviruses are classified into two groups, old world (Europe, Africa, Asia, and Australia) and new world (America) viruses, based on their genome arrangement, genetic diversity, geographical distribution, and phylogenetic analysis (Nawaz-ul-Rehman and Fauquet, 2009) Begomoviruses are composed of either a bipartite genome of two component DNAs, DNA-A and DNA-B, or a single component, DNA-A. Each component is approximately 2.7 kb and has a common region. DNA-A consists of five ORFs in the complementary strand and two ORFs in the virion strand. The ORFs are involved in replication (AC1, AC3 and AV1), transcription activation (AC2) and packaging (AV1). The role of the AC4 gene is unknown, and the AC5 gene is non-functional. DNA-B has only two major protein ORFs, BC1 and BV1, which are associated with viral movement in the plant cells (Seal et al., 2006; Stanley J, 2005) The opposing complementary and virion sense genes located on both components are disunited by a non-coding intergenic region (IR), which is highly conserved and called the common region (CR) (Lazarowitz and Shepherd, 1992).The CR consists of approximately 200 nucleotides that show 96% identity between both components and contain a highly conserved sequence of nine nucleotides (TAATATTAC) called the nonanucleotide, which contains an origin of replication (ori) within the CR (Laufs et al., 1995).Some of the monopartite Begomoviruses are also associated with additional circular ssDNA molecules, such as betasatellites or alphasatellites (previously known as DNA-1), that are nearly half of the size of DNA-A. Betasatellites have been involved

in pathogenicity, but alphasatellites have unknown function (Mansoor et al., 2003c; Saunders et al., 2002) Alphasatellites have been shown to be present in plants infected with begomoviruses that are in association with batasatellites (Mubin et al., 2007) The members of the Geminiviridae family replicate using the rolling circle replication (RCR) system, which has two steps. The first step involves the synthesis of the "minus" strand. During the synthesis of the minus strand, the "plus" strand act as a template, this results in the generation of the double stranded replicative form (dsRF). In the final step, free ssDNA is generated by the formation of the plus strand, which uses RF as a template (Hanley-Bowdoin et al., 2000). The first evidence of CLCuD was found in the vicinity of a city in the Multan district, Pakistan, in 1967 (Hussnain T. and Ali, 1975). A second CLCuD epidemic emerged in the vicinity of Burewala, in the Punjab district of Pakistan, and was a serious threat to cotton production that resulted in a 100% loss of the crop yield in the majority of the cotton growing areas in 2001. The causative agent of this epidemic was named the 'Burewala strain' and is known as a distinct begomovirus: cotton leaf curl Burewala virus (CLCuBuV) (Amrao et al., 2010b; Mansoor et al., 2003a; Mansoor and Briddon, 2007; Mansoor et al., 2006) . Several species of begomoviruses are reported to cause CLCuD. The major species of these begomoviruses that have been found in Pakistan include cotton leaf curl Alabad virus (CLCuAV), cotton leaf curl Multan virus (CLCuMuV), cotton leaf curl Burewala virus(CLCuBuV), cotton leaf curl Rajasthan virus (CLCuRV), Tomato leaf curl New Dehli virus(ToLCNDV), cotton leaf curl Shahdadpur virus (CLCuShV), cotton leaf curl Gezira virus (CLCuGV) and Papaya leaf curl virus(PaLCuV) (Mansoor et al., 2003b) All of these viruses are associated with a single betasatellite, the "Cotton leaf curl Multan betasatellite"(CLCuMB). Cotton growth has been highly affected by whitefly-disseminated begomoviruses. Cotton plants that show characteristic CLCuD symptoms such as leaf curling, vein darkening, vein swelling and enation have been reported in the large-scale cotton-growing areas in the Punjab and Sindh provinces. Cotton is grown in these regions, and the regions are very close to India. To characterize the CLCuV, diseased cotton leaves were collected. The DNA-A component of CLCuBuV was isolated, sequenced and characterized in this study. The work reported in this research article, when considered with earlier studies on CLCuV, describes the phylogenetic analysis and potential recombination of CLCuBuV isolates in comparison with other prevailing geminiviruses. It is expected that the current work will help us to understand begomovirus genetics, and evolution.

Results

Cloning, sequencing and genome organization of CLCuBuV isolates

Leaves of cotton plants that exhibited symptoms typical of begomovirus infection associated with CLCuD were collected from a farm field in South Punjab, Pakistan. The total genomic DNA was isolated, and RCA was used for the amplification of the circular DNA molecules. Restriction digestion of the RCA product with HindIII/BamH1/Sac1 resulted in 2758 bp fragments, which were cloned in the pBluescript II KS (+) plasmid vector. A total of four clones were obtained and sequenced using a primer walking strategy. The sequences are available in NCBI nucleotide sequence database under accession numbers (Table.1).The DNA-A component of the ClCuBuV genome in this study exhibited a sequence organization typical of monopartite old world begomovirus isolates from the same geographic location. The DNA-A component of each CLCuBuV isolate in this study was determined to be 2758 bp long, composed of two ORFs [AV1 (CP), AV2] in the virion-sense strand and three ORFs [AC1 (Rep), AC3, AC4] in the complementary-sense strand, separated by a large intergeinc region (LIR) (fig.5). The unique feature of these CLCuBuV isolates was determined to be a lack of one ORF [AC2 (Trap)]. The LIR was composed of 455 bp and contained cis-acting DNA elements that are involved in begomovirus replication and transcriptional regulation, including a Rep-binding site (iterons), the TATA boxes, GC boxes and stem-loop elements that contain the conserved nona-nucleotide (TAATATTAC) sequence. Genome features of DNA-A component of CLCuBuV isolates under study associated with CLCuD (Table.3).

Nucleotide sequence comparison

A comparison of sequences obtained in this study, using the Clustal W2 algorithm implemented in MegAlign, revealed that all 4 sequences share the highest (98.6-99.9%) nucleotide sequence identity with other isolates of CLCuBuV that are present in nucleotide sequence databases, indicating that these are "variants" of a single species, CLCuBuV, based on the presently applicable species demarcation criteria for International Committee on Taxonomy of Viruses (ICTV) variants (92-100%) (Fauquet et al., 2008).The CLCuBuV isolates in this study showed 86.7-83.1% nucleotide sequence homology with CLCuKoV sequences and 88.5-86.5% with CLCuShV sequences. These isolates showed the lowest nucleotide identity with CLCuAV sequences available in NCBI database. The percentage nucleotide sequence identities for pairwise comparison of the sequences of the isolates of CLCuBuV in this study with other CLCuD-associated begomoviruses (Table.4).

Phylogenetic analysis

The phylogenetic analysis was performed on the DNA-A component of CLCuBuV isolates, including the segment of DNA-A which contained the AC1, AV1, and LIR. The phylogenetic trees were developed to analyze and characterize the isolates in this study compared with other CLCuV isolates from different geographical areas. Several geminiviruses that affect nonmalvaceaus crops were also considered for the similarity search. A list of accession numbers and names of viruses (Table.2). When the DNA-A component was used for the similarity search, the CLCuBuV isolates in this study clustered together with CLCuBuV isolates that were reported from India and Pakistan, such as CLCuBuV [IN:His1:10], CLCuBuV [IN: Bat2:10], CLCuBuV [IN: Dab1:10], CLCuBuV [IN: Sri:10], CLCuBuV [PK:Lod:04], CLCuBuV [PK:RYK:04] CLCuKV [IN:Sri:10], CLCuKV [Pk:806b:96] and CLCuShV [Pk:Sha1:05] (Fig.1). The phylogenetic analysis of AC1 ORF revealed that CLCuBuV [PK:Veh1:04], CLCuBuV [IN:Sri:10], CLCuBuV [IN:Abo:10], CLCuBuV [IN:His1:10], CLCuMuV [PK:Mul], CLCuMuV [PK:Fais2:02], CLCuMuV [PK:Fais2:02], CLCuShV [Pk:Sha2:04], and CLCuShV [Pk:Sak:05] are closely related to CLCuBuV isolates in this study a (as shown inFig.2). The phylogenetic analysis of AV1 ORF of the CLCuBuV isolates in this study

Table 2. GenBank accession numbers and demographics of begomoviruses isolated from different crops.

No	Viruses	Virus acronym	Accession IDs	Host	Year	Region	Genome size (bp)
1	Ageratum enation virus	AEV	AM698011	A. conyzoides	2006	Lahore: Pakistan	2749
2	Ageratum yellow vein virus	AYVV	DQ866134	A.conyzoides	2000	Taoyuan: Taiwan	2754
3	Ageratum yellow vein China virus	AYVCV	JQ804985	S. Lycopersicum	2009	Nanning: China	2739
4	Croton yellow vein mosaic virus	CYVMV	FN645926	Acalypha sp.	2007	Punjab: India	2761
5	Hollyhock yellow vein mosaic virus	HYVMV	FR772081	A.rosea	2006	Lahore: Pakistan	2750
6	Mungbean yellow mosaic virus	MYMV	FM242701	Rcapitata	2008	Mianwali: Pakistan	2737
7	Malvastrum yellow vein Changa Manga virus	MaYVCV	FR715681	M. coromandelianum	2010	ChangaManga:Pakistan	2754
8	Okra leaf curl virus	OKLCV	GQ245760	A.esculentus	2008	New Dehli:India	2742
9	Papaya leaf curl virus	PaLCuV	FM955602	R.capitata	2008	Mianwali:Pakistan	2756
10	Tomato leaf curl Pakistan virus	ToLCPKV	AM948961	G.max	2008	Nawab Shah:Pakistan	2760
11	Tomato leaf curl New Delhi virus	ToLCNDV	FN435310	S. lycopersicum	2009	Faisalabad:Pakistan	2739
12	Tomato leaf curl Philippines virus	ToLCPhV	AB050597	S.lycopersicum	2000	Los Banos:Philippines	2755
13	Tomato yellow leaf curl virus	ToYLCV	AF071228	S. lycopersicum	1998	Spain	2791
14	Tomato yellow leaf curl virus	ToYLCV	AF105975	S.lycopersicum	1998	Purtugal	2793
15	Tomato leaf curl Arusha virus	ToLCAV	EF194760	S. lycopersicum	2006	Kilimanjaro: Tanzania	2762
16	Tomato mottle virus	ToMoV	L14460	S.lycopersicum	1989	Florida: USA	2601
17	Cotton leaf curl Gezira virus	CLCuGV	FR751142	G. hirsutum	2005	Hala:Pakistan	2763
18	Cotton leaf curl Gezira virus	CLCuGV	AY036007	S.cordifolia	2001	Gezira: Sudan	2761
19	Cotton leaf curl Gezira virus	CLCuGV	AY036010	A.esculentus	2001	Cairo: Egypt	2764
20	Cotton leaf curl Gezira virus	CLCuGV	AY036006	A.esculentus	2001	Gezira: Sudan	2761
21	Cotton leaf curl Gezira virus	CLCuGV	AY036008	A.esculentus	2001	Shambat: Sudan	2761
22	Cotton leaf curl Gezira virus	CLCuGV	AF260241	G.hirsutum	2001	Gezira: Sudan	2761
23	Cotton leaf curl Multan virus	CLCuMuV	AY765253	G.hirsutum	2004	Hissar:India	2725
24	Cotton leaf curl Multan virus	CLCuMuV	AY765256	G.hirsutum	2004	New Dehli: India	2738
25	Cotton leaf curl Multan virus	CLCuMuV	DQ191160	G.hirsutum	2005	Bhatinda:India	2753
26	Cotton leaf curl Multan virus	CLCuMuV	AJ132430	G.hirsutum	1999	Faisalabad:Pakistan	2722
27	Cotton leaf curl Multan virus	CLCuMuV	AJ496461	G.hirsutum	2002	Faisalabad:Pakistan	2727
28	Cotton leaf curl Multan virus	CLCuMuV	X98995	G.hirsutum	1996	Faisalabad: Pakistan	2747
29	Cotton leaf curl Multan virus	CLCuMuV	AJ002459	A.esculentus	1996	Multan:Pakistan	2725
30	Cotton leaf curl Multan virus	CLCuMuV	AJ002447	G.hirsutum	1995	Yazman:Pakistan	2751
31	Cotton leaf curl Multan virus	CLCuMuV	AJ002458	G. hirsutum	1995	DG Khan:Pakistan	2751
32	Cottonleaf curl Multan virus	CLCuMuV	AY765257	G. hirsutum	2004	Ludhiana:India	2725
33	Cotton leaf curl Multan virus	CLCuMuV	AJ496287	G. hirsutum	2002	Faisalabad: Pakistan	2750
34	Cotton leaf curl Shadadpur virus	CLCuShV	FN552001	G.hirsutum	2005	Shadadpur: Pakistan	2748
35	Cotton leaf curl Shadadpur virus	CLCuShV	FN552002	G.hirsutum	2005	Tandojam: Pakistan	2748
36	Cotton leaf curl Shadadpur virus	CLCuShV	FN552003	G. hirsutum	2005	Tandojam: Pakistan	2748
37	Cotton leaf curl Shadadpur virus	CLCuShV	FN552004	G.hirsutum	2005	Shadadpur: Pakistan	2748
38	Cotton leaf curl Shadadpur virus	CLCuShV	FN552005	G.hirsutum	2005	Shadadpur: Pakistan	2748
39	Pepper leaf curl Lahore virus	PepLCLV	AM691745	C. annum	2007	Faisalabad: Pakistan	2747
40	Chilli leaf curl Multan virus	ChiLCuMV	FM179613	S.tuberosum	2008	Punjab: Pakistan	2754
41	Chilli leaf curl Multan virus	ChiLCuV	AF336806	C. annum	2001	Multan: Pakistan	2754
42	Chilli leaf curl Multan virus	ChiLCuMV	FM179613	S.tuberosum	2005	Lahore: Pakistan	2754
43	Cotton leaf curl Bangalore virus	CLCuBV	NC_007290	G. hirsutum	2004	Banglore: India	2751
44	Cotton leaf curl Bangalore virus	CLCuBV	GU112003	A.esculentus	2005	Banglore: India	2758
45	Cotton leaf curl Burewala virus	CLCuBuV	AM774292	G.hirsutum	2004	Vehari: Pakistan	2758
46	Cotton leaf curl Burewala virus	CLCuBuV	AM421522	G.hirsutum	2004	Vehari: Pakistan	2759
47	Cotton leaf curl Burewala virus	CLCuBuV	AM774297	G. hirsutum	2004	R.Y.Khan: Pakistan	2759
48	Cotton leaf curl Burewala virus	CLCuBuV	AM774302	G.hirsutum	2004	Muzzafargarh:Pakistan	2759
49	Cotton leaf curl Burewala virus	CLCuBuV	AM774300	G. hirsutum	2004	Lodhran:Pakistan	2759
50	Cotton leaf curl Burewala virus	CLCuBuV	AM774301	G. hirsutum	2004	Layyah:Pakistan	2759
51	Cotton leaf curl Burewala virus	CLCuBuV	AM774295	G. hirsutum	2004	Khanewal:Pakistan	2759
52	Cotton leaf curl Burewala virus	CLCuBuV	AM774305	G.hirsutum	2004	Faisalabad:Pakistan	2759
53	Cotton leaf curl Burewala virus	CLCuBuV	AM774303	G. hirsutum	2004	Arifwala:Pakistan	2758
54	Cotton leaf curl Burewala virus	CLCuBuV	AM774298	G. hirsutum	2004	D GKhan:Pakistan	2758
55	Cotton leaf curl Burewala virus	CLCuBuV	AM774299	G.hirsutum	2004	Faisalabad:Pakistan	2758
56	Cotton leaf curl Burewala virus	CLCuBuV	AM774304	G. hirsutum	2005	Faisalabad:Pakistan	2758

57	Cotton leaf curl Burewala virus	CLCuBuV	JF510459	G. hirsutum	2010	Dabwalil:India	2759
58	Cotton leaf curl Burewala virus	CLCuBuV	JF510458	G.hirsutum	2010	Bhatinda:India	2760
59	Cotton leaf curl Burewala virus	CLCuBuV	JF502373	G.hirsutum	2010	Hanumangarh:India	2759
60	Cotton leaf curl Burewala virus	CLCuBuV	JF502367	G.hirsutum	2010	SriGanganagar:India	2759
61	Cotton leaf curl Burewala virus	CLCuBuV	JF502370	G.hirsutum	2010	Hissar: India	2759
62	Cotton leaf curl Burewala virus	CLCuBuV	JF502360	G.hirsutum	2010	Abohar: India	2759
63	Cotton leaf curl Burewala virus	CLCuBuV	JF502366	G.hirsutum	2010	Fazilka2: India	2759
64	Cotton leaf curl Burewala virus	CLCuBuV	JF502365	G.hirsutum	2010	Fazilka1: India	2759
65	Cotton leaf curl Alabad virus	CLCuAV	AJ002452	G.hirsutum	1996	Alabad:Pakistan	2744
66	Cotton leaf curl Alabad virus	CLCuAV	AJ002455	G.hirsutum	1996	Alabad:Pakistan	2744
68	Sida leaf curl virus	SiLCuV	DQ641706	A. indicum	2006	Thanhhoa: Viet Nam	2760
69	Mesta yellow vein mosaic virus	MYVM	FN645922	A.esculentus	2007	Harayana:India	2742
70	Hollyhock leaf curl virus	HoLCuV	FR772082	A. rosea	2006	Faisalabad:Pakistan	2748
71	Okra yellow crinkle virus	OYCrV	FM164724	A.esculentus	2008	Njombe: Cameroon	2776
72	Cotton leaf curl Gezira virus	CLCuGV	AY036007	S.cordifolia	2001	Gezira:Sudan	2761
73	Bhendi yellow vein mosaic virus	BYVMV	GU112078	A.esculentus	2006	Dehli:India	2758
74	Cotton leaf curl Rajasthan virus	CLCuRV	FM202328	D.arvensis	2008	Rajasthan:Pakistan	2751
75	Cotton leaf curl Rajasthan virus	CLCuRV	JF502364	G.hirsutum	2010	Abohar:India	2751
76	Cotton leaf curl Rajasthan virus	CLCuRV	AY795606	G.hirsutum	2003	Abohar:India	2755
77	Cotton leaf curl Rajasthan virus	CLCuRV	AY795605	G. hirsutum	2003	New Dehli:India	2754
78	Cotton leaf curl Rajasthan virus	CLCuRV	AF363011	G.hirsutum	2001	Sriganganagar:India	2754
79	Cotton leaf curl Rajasthan virus	CLCuRV	NC_003199	G. hirsutum	2001	Sriganganagar:India	2754
80	Cotton leaf curl Rajasthan virus	CLCuRV	AY765254	G. hirsutum	2001	Sirsa:India	2739
81	Cotton leaf curl Kokhran virus	CLCuKoV	HM468427	G. stocksii	2008	Multan:Pakistan	2748
82	Cotton leaf curl Kokhran virus	CLCuKoV	HQ257374	G. hirsutum	2010	New Dehli: India	2746
83	Cotton leaf curl Kokhran virus	CLCuKoV	AJ002449	G. hirsutum	1996	Multan:Pakistan	2758
84	Cotton leaf curl Kokhran virus	CLCuKoV	AJ496286	G.hirsutum	2002	Faisalabad:Pakistan	2750
85	Cotton leaf curl Kokhran virus	CLCuKoV	FN552006	G.hirsutum	2005	Sarkand:Pakistan	2750

Abbreviation A: Ageratum, Alcea, Abelmoschus, Abutilon; S; Solanum, Sida, R: Rynchosia, G: Glycine; C: Capsicum annum; M: Malvastrum; G: Gossypium; D: Digeria.

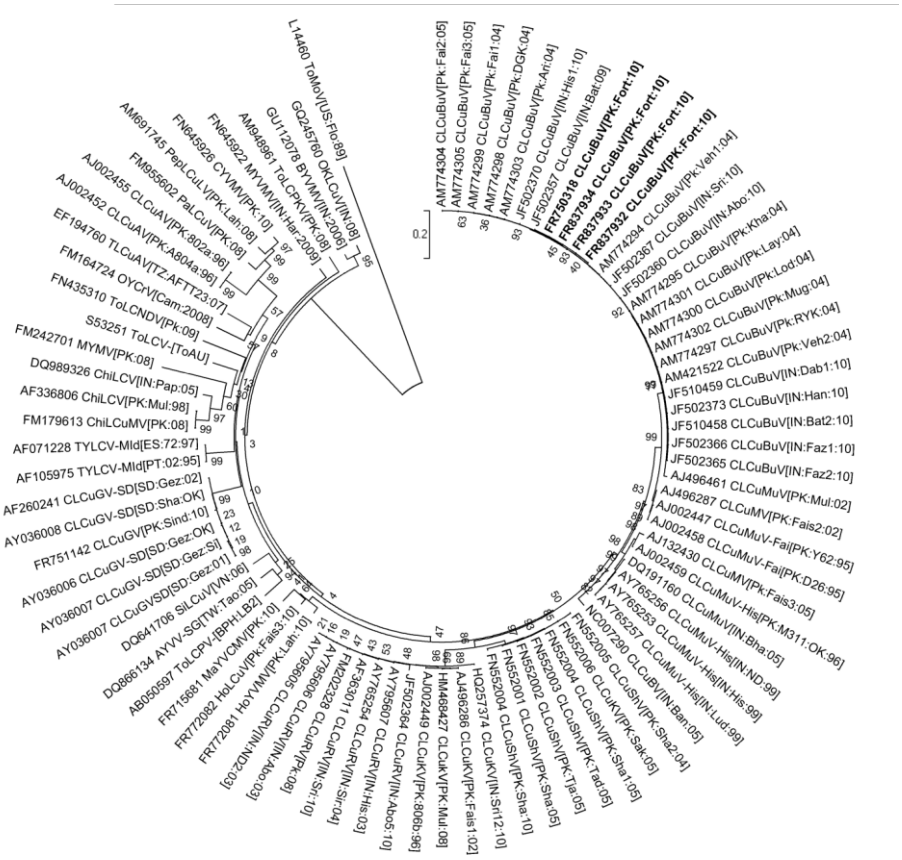


Fig 2. CLCuBuV AC1 phylogenetic tree. The phylogenetic methods and abbreviations were as described for figure 1.

showed close homology with CLCuKoV [IN:Sri12:10], CLCuKoV [PK:806b:96], CLCuBuV [IN:Abo:10], CLCuBuV [IN:Faz2:10], CLCuBuV [IN:Faz1:10], CLCuBuV [PK:RYK:04] , CLCuBuV [PK:Veh2:04] , CLCuBuV [PK:Mug:04], CLCuShV [Pk:Tja:05] and CLCuShV [Pk:Sha:05] as in(fig.3). The phylogenetic analysis of the LIR revealed that CLCuBuV [IN:Abo:10], CLCuBuV [IN:Faz2:10], CLCuBuV [IN:Faz1:10], CLCuBuV [IN:Bat2:10], CLCuBuV [PK:Mug:04], CLCuBuV [PK:Ari:04], CLCuKoV [IN:Sri12:10], CLCuShV [Pk:Sha:10], and CLCuKoV [IN:Sri12:10] are clustered together in the same group (Fig.4). Phylogenetic analysis predicted that there was the possibility of recombination events in the CLCuBuV isolates under consideration, and their variable genes (ORFs) and intergenic regions exhibited sequence homology with various begomoviruses.

Recombination analysis

The recombination detection hypothesis was evaluated by applying a computer-based program, RDP3. The study analyzed the first possible potential intermolecular recombination event with a high probability ($P=5.746 \times 10^{-24}$), which revealed the recombination origin of the CLCuBuV isolates in this study from nucleotide 1309-2495 within the complementary sense sequences when complete DNA-A was subjected to analysis by RDP3 with major parent CLCuMuV[PK:Fai:02](AJ496287) with highest probability ($P=3.646\times 10^{-15}$) and second potential recombination event that was observed within the virion-sense sequence from nucleotides 403-1090 with major parent CLCuKoV [PK:Mul:96](AJ002449). The third potential recombination event was observed within LIR with high probability ($P=1.067 \times 10^{-03}$) determined between nucleotide 172-325 and detected all of six methods implemented in RDP3 with major and minor parents AEV [Pk:Lah: 06] (AM698011) and CLCuBV[IN: Ban:04] (NC_007290) respectively. Break point analysis of CLCuBuV isolates under study (Table.5).

Discussion

RCA based amplification followed by sequencing suggested that leaf curl disease in South Punjab, Pakistan, is now predominantly caused by CLCuBuV. There is no difference in the symptoms of CLCuBuV with other cotton leaf curl viruses associated with CLCuD in Pakistan. The CLCuBuV exhibited characteristic CLCuD symptoms such as leaf curling, vein darkening, vein swelling and enation. The symptoms are similar with other CLCuVs associated with CLCuD reported from Pakistan such as CLCuBuV (Amrao et al., 2010b; Nawaz-ul-Rehman et al., 2012) CLCuShV (Amrao et al., 2010a) CLCuGV (Tahir et al., 2011) and CLCuBuV (Rajagopalan et al., 2012) CLCuRV (Kumar et al., 2010) reported from India .The samples were collected from Fortabbas which is very close to Burewala, an area where CLCuBuV was first reported. CLCuBuV is also reported from other cotton growing areas of Pakistan which are Muzzafarabad(Multan), Muzzafargarah, Lodhran, Layyah, Bahawalpur ,Faisalabad, Arifwala,Loddan, Dera Ghazi Khan, Khanewal, Rahim Yar Khan, Vehari (Amrao et al., 2010b) and also from India from Fazilka, Abohar,Bathinda,Hissar,Sirsa,Dabawali,Fatehabad ,Haryana,Hanumangarh. Three variants of CLCuBuV are prevalent in northwest India (Rajagopalan et al., 2012). The same variants of CLCuBuV were detected from Fortabbas, Vehari, Burewala(Pakistan) Abohar, Sirsa, Dabawali (India) which are about 200km apart, suggesting that dissemination of new variants of viruses associated with CLCuD can occur within few years. RCA is a common replication procedure of the geminiviridae family. Earlier studies suggested that the causative agent of the CLCuD epidemic that occurred in Pakistan and India has a very complex genetic organization with the association of different begomovirus species and strains (Chatterji et al., 1999; Mansoor et al., 2003b).The study demonstrated the characteristic monopartite genetic structure of the isolates under study. Because the structure of CLCuBuV has the 2.7 kb DNA-A component, which is a typical characteristic of monopartite begomoviruses, the structure lacks the DNA-B component, which is typical feature of bipartite begomoviruses. For sequence characterization and analysis of samples from different areas, samples were initially phi amplified. The full-length nucleotide sequence analysis revealed that the CLCuBuV isolates had the maximum percentage nucleotide identity with CLCuBuV, CLCuKoV, CLCuMuV and CLCuShV reported from Pakistan and India. This result may be due to a high rate of mutation, recombination in begomoviruses (Padidam et al., 1999) and insect vector migration from neighbor countries (Kumar et al., 2010) For further study, to determine the phylogenetic relationship with other geminiviruses and for the detection of recombinant events within DNA-A, the complete sequence includes the AC1 ORF, the AV1 ORF and the LIR. The DNA-A sequence analysis revealed that these isolates under study are variants of highly infectious CLCuBuV. They showed maximum sequence identity (98.6-99.9%) with CLCuBuV isolates reported from Pakistan and India. They showed 88.5-86.5% sequence similarity with CLCuShV and 88.5-86.5% with CLCuMuV that was reported from Pakistan. The nucleotide sequence comparison analysis is consistent with previously reported CLCuV isolates from Pakistan. The phylogenetic analysis of DNA-A component of CLCuBuV isolates under study exhibited maximum sequence homology with CLCuBuV, CLCuKoV, CLCuShV, CLCuBV, CLCuRV, MYVMV,CLCuMuV, CLCuAV and BYVMV sequences reported from Pakistan and India as shown in fig1.These findings are in accordance with previously reported begomoviruses associated with CLCuD in Pakistan. The phylogenetic analysis of AC1 ORF of DNA-A of CLCuBuV genome under consideration revealed that CLCuMuV AC1 ORF sequences are more closely related to CLCuBuV isolates reported from India and Pakistan as shown in Fig2 which supports the recombination analysis that CLCuMuV is a major parent of complementory sense ORFs (AC1, AC3, AC4) of CLCuBuV isolates under study as in table5.. The phylogenetic analysis of AV1 ORF under study, exhibited maximum sequence identity with CLCuKoV AV1 ORF as shown in fig.3 which is in accordance with the recombination analysis of DNA-A component that showed that virion sense ORFs(AV1 and AV2) are created from CLCuKoV(Table.5). The phylogenetic analysis of LIR revealed that CLCuShV and CLCuKoV sequences are more closely related to LIR of CLCuBuV isolates under study. Recombination analysis was performed by applying RDP (version 3.0). It was concluded that the DNA-A component of CLCuBuV isolates under study was created by a recombination event between the DNA-A of CLCuKoV and CLCuMuV. Recently, two new recombinant *cotton leaf curl viruses* were identified in Sindh province and were named cotton leaf curl Shahdadpur virus (Amrao et al., 2010a) and cotton leaf curl Gezira virus

Table 3. Genome features of CLCuBuV isolates associated with CLCuD from infected cotton plants.

ORF	Start codon (nucleotides coordinates)	Stop codon (nucleotides Coordinates)	Predicted size of ORFs (bp)	Predicted size of proteins (no. of amino acids)	Predicted coding capacity (KDal)
AV1(CP)*	292	1062	770	256	29.70
V2	132	488	356	118	13.70
Rep(C1)	2595	1534	1061	361	40.90
REn(C3)	1463	1059	404	134	15.50
C4	2681	2136	545	97	20.40

*Genes are indicated as Coat protein (CP), Replication-associated protein (Rep) and Replication Enhancer (REn). The products encoded by by ORFs V2 and C4 have yet not been named.

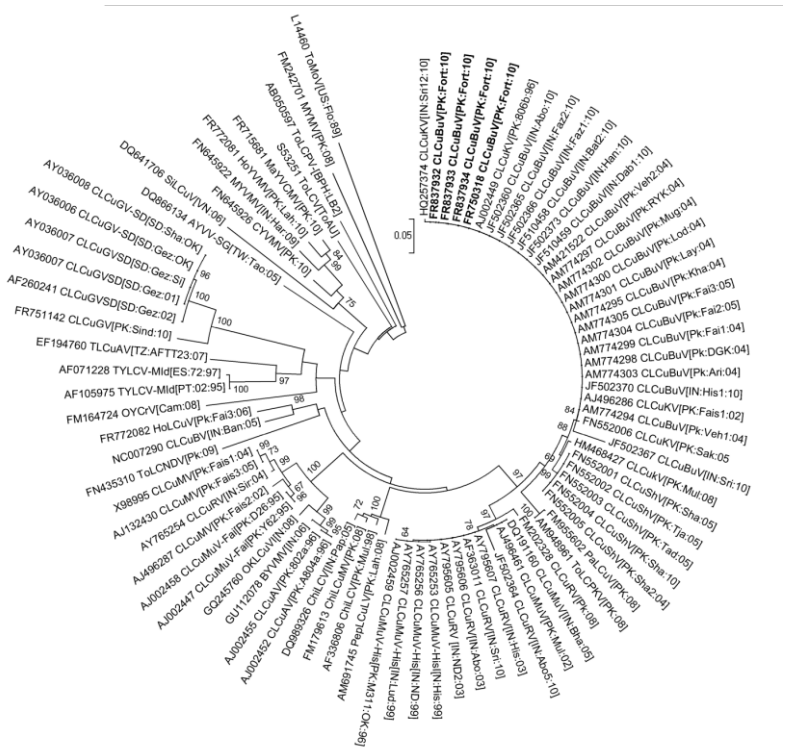


Fig 3. CLCuBuV AV1 phylogenetic tree. The phylogenetic methods and abbreviations were as described for Figure 1.

(Tahir et al., 2011). Many factors are involved in the progression of cotton leaf curl virus, among which, recombination plays a key role (Zhou et al., 1997; Zhou et al., 1998) Recombination is a major component of geminivirus evolution (Padidam et al., 1999). Mixed infection is a driving force for recombination. The South Punjab region of Pakistan is a major cotton growing area that is adjacent to the western states of India. It has been observed that *B. tabaci* that harbor begomoviruses, particularly CLCuV, cross the Pakistan-India boundary and transmit disease. This insect vector is a major source of the cross-border inoculation of begomoviruses within and outside the country. The recombination analysis found that the DNA-A component of CLCuBuV isolates under study was the product of a recent CLCuKoV- CLCuMuV recombination event. These results are in consistent with previous studies on DNA-A component of CLCuBuV reported (Amrao et al., 2010b;Nawaz-ul-Rehman et al., 2012; Rajagopalan et al., 2012) It can be concluded that continual recombination among geminiviruses may enrich the further development of CLCuD. A need exists to rule out other elements that are subject to the emergence of CLCuD in Pakistan. Geminiviruses cause losses worth millions to a wide range of crops, especially cotton, in Pakistan. The CLCuV has been reported on many crops other than cotton and is disseminated from one crop to another through whiteflies. Therefore, the need to control whiteflies is high because, otherwise, this disease may harm Pakistan's economy. Further development of infectious clones of CLCuBuV is in progress to determine the promoter activity of the C1 and V1 genes of LIR with GUS as reporter gene and transformation of RNAi construct of CLCuBuV in cotton.

Materials and methods

Sample collection

Cotton leaves that exhibited characteristic CLCuD symptoms such as greening (affected plants at early stages of infection appear darker green than unaffected plants), leaf curling, vein darkening, vein swelling, enation and production of cup-shaped leaf-like structures on the undersides of leaves were collected from a farm filed in Tehsil Fortabbas district Bahawalnagar, South Punjab, Pakistan, during the period of

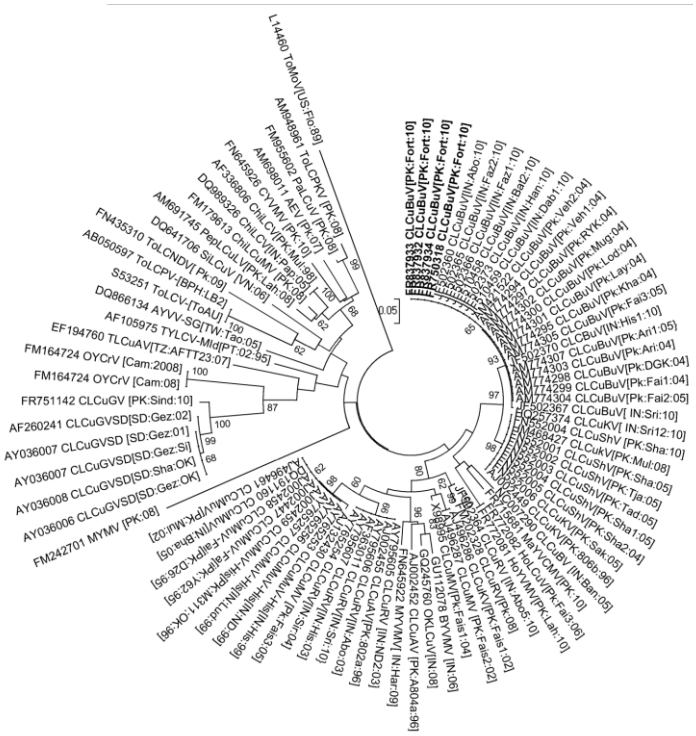
Table 4. Highest and lowest percentage nucleotides sequence identities for pairwise comparisons of the sequences of the CLCuBuV isolates with all available sequences of other cotton leaf curl disease-associated begomoviruses.

	CLCuBuV (20)[a]	CLCuMuV(10)[a]	CLCuKoV(5)[a]	CLCuBV(2)[a]	CLCuAV(2)[a]	CLCuRV(7)[a]	CLCuShV(5)[a]
CLCuBuV (4)[a]	99.9-98.6	85.5-76.2	86.7-83.1	78.8 -78.4	68.7-68.2	79.9-74.3	88.5-86.5
CLCuMuV(10)[a]		99.2-86.4	77.7-71.4	83.1-79.6	78.5-76.4	88.5-82.7	84.7-79.4
CLCuKoV(5)[a]			99.8-92.4	80.2-76.9	67.6-66.7	87.7-82.3	90.3-83.4
CLCuBV(2)[a]				99.9-99.6	73.4-73.2	80.5-80.3	67.9-67.1
CLCuAV(2)[a]					99.9-99.7	76.7-74.	77.5-77.3
CLCuRV(7)[a]						99.6-94.5	87.2-82.9
CLCuShV(5)[a]							99.9-99.7

a; The number of sequences available which were used in the comparisons.

Table5. Break point analysis of CLCuBuV isolates under study associated with CLCuD, with putative parental sequences.

Virus-DNA-A segments	Break point Begin-end	Major parent	RDP	GENECOV	Boot Scan	MaxChi	Chimera	Si Scan
Virion sense ORF	403-1090	CLCukoV	$3.646X10^{-15}$	$9.461X10^{-9}$	$2.21X10^{-24}$	$6.11X10^{-13}$	$6.67X10^{-8}$	$6.980X10^{-22}$
Complementary sense ORF	1309-2495	CLCuMuV	$5.746X10^{-24}$	$4.788X10^{-3}$	$2.046X10^{-5}$	$2.23X10^{-1}$	$1.85X10^{-3}$	$3.803X10^{-7}$
LIR	172-325	AEV	$1.067X10^{-3}$	$2.116X10^{-2}$	$1.494X10^{-2}$	$2.96X10^{-4}$	$2.36X10^{-7}$	$1.363X10^{-5}$



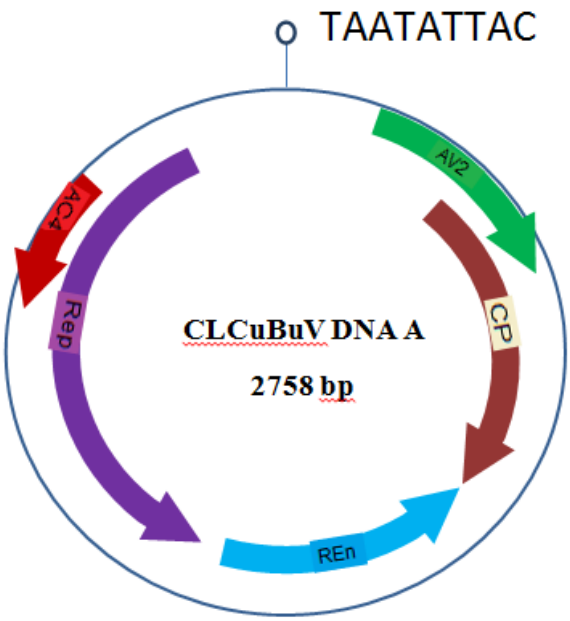


Fig 4. CLCuBuV LIR phylogenetic tree. The phylogenetic methods and abbreviations were as described for Figure 1.

Fig 5. Organization of DNA-A component of CLCuBuV genome showing five overlapping ORFs and LIR, including stem-loop structure.

August-September 2010. A total of twenty samples were collected at the rate of five leaves per sample per plant.

Full-length genome amplification

The total DNA was isolated from infected cotton leaves by the CTAB method (Doyle and Doyle, 1990). The full-length genome of CLCuBuV was amplified by rolling circle amplification (RCA) using a Temliphi^{TM} DNA amplification kit (GE healthcare, USA) following the manufacturer's protocol. Each RCA reaction mix was centrifuged and then incubated at 30°C for 16 to 18 h. After 18 h, the mixture was placed at 65°C for 10 min to denature the restriction enzyme and gradually cooled. The RCA method, which used φ29

DNA polymerase, was performed to amplify all circular DNA molecules in DNA samples (Haible et al., 2006; Inoue-Nagata et al., 2004)

Cloning and sequencing:

A Nano Drop machine was used to quantify the DNA. Subsequently, 2 µg of each RCA phi-amplified product was digested with 1 U of restriction enzyme. After 3 hrs of incubation, the digested products were loaded on a 1 % agarose gel. The expected 2.7 kb band was excised from the agarose gel and were purified using the DNA purification kit, (Fermentas, USA). The eluted amplified product was then ligated in pBluescript II KS (+) and transformed into *E. coli*

DH5a. The cloned viruses were sequenced in its entirety, initially by M-13 forward and M-13 reverse primers and subsequently by Primer walk

Phylogenetic and sequence analysis

The complete DNA-A nucleotide sequences that were initially found via a similarity search using NCBI BLASTn (Altschul et al., 1990) (http://www.ncbi.nlm.nih.gov/BLAST/) were selected based on high percentage identity. Selected sequences were then aligned with the CLCuBuV isolates using ClustalW2 (Larkin et al., 2007). The sequence data were assembled and analyzed with the aid of the Lasergene package of sequence analysis software (DNA Star Inc., Madison WI USA). Open reading frames (ORFs) were located using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html).Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011). The evolutionary history was inferred using the Neighbor-joining method with bootstrap values of 1000 replicates (Felsenstein, 1985; Saitou and Nei, 1987).The evolutionary distances were computed using the JTT matrix-based method (Jones et al., 1992) Accession numbers of selected DNA-A nucleotide sequences are given in table.

Recombination analysis

Recombination events were analyzed using the software Recombination Detection Program Version 3.0 (RDP3), which simultaneously uses a range of different recombination detection methods to both analyze and detect the recombination events that are logical in a sequence alignment after any previous user input of a non-recombinant set of reference sequences. For the recombination analysis, the DNA-A component of CLCuBuV isolates under study were taken into account. Identification of recombinant fragments, detection of acceptable paternal sequences, and localization of potential recombination breakpoints were performed using the RDP, GENECONV, BOOTSCAN, MAXIMUM CHI SQUARE, CHIMAERA and SISTER SCAN recombination detection methods as implemented in RDP3 (Martin and Rybicki, 2000) The analysis used the default levels for the altered recombination detection methods, and a Bonferroni fixed-up P-value of 0.05 was selected. Only recombinant events that were analyzed applying three of more were accepted.

Acknowledgements

We are extremely thankful to PARB (Pakistan Agriculture Research Board) and the Higher Education Commission (HEC) for providing funding.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. 215:403-410.

Amrao L, Akhter S, Tahir MN, Amin I, Briddon RW, Mansoor S (2010a) Cotton leaf curl disease in Sindh province of Pakistan is associated with recombinant begomovirus components. Virus Res. 153:161-165.

Amrao L, Amin I, Shahid MS, Briddon RW, Mansoor S (2010b) Cotton leaf curl disease in resistant cotton is associated with a single begomovirus that lacks an intact transcriptional activator protein. Virus Res. 152:153-163.

Asad S, Haris WAA, Bashir A, Zafar Y, Malik KA, Malik NN, Lichtenstein CP (2003) Transgenic tobacco expressing geminiviral RNAs are resistant to the serious viral pathogen causing cotton leaf curl disease. Arch Virol. 148:2341-2352.

Chatterji A, Padidam M, Beachy RN, Fauquet CM (1999) Identification of replication specificity determinants in two strains of tomato leaf curl virus from New Delhi. J Virol. 73:5481-5489.

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

Fauquet C, Briddon R, Brown J, Moriones E, Stanley J, Zerbini M, Zhou X (2008) Geminivirus strain demarcation and nomenclature. Arch Virol. 153:783-821.

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 39:783-791.

Haible D, Kober S, Jeske H (2006) Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. J Virol Methods. 135:9-16.

Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D (2000) Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. Crit Rev Biochem Mol Biol. 35:105-140.

Hussnain T, Ali M (1975) A review of cotton diseases of Pakistan. Pak Cott. 19:77-78.

Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T (2004) A simple method for cloning the complete begomovirus genome using the bacteriophage φ29 DNA polymerase. J Virol Methods. 116:209-211.

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Computer applications in the biosciences: CABIOS. 8:275-282.

Kumar A, Kumar JKhan JA (2010) Sequence characterization of cotton leaf curl virus from Rajasthan: phylogenetic relationship with other members of geminiviruses and detection of recombination. Virus Genes. 40:282-289.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics. 23:2947-2948.

Laufs J, Jupin I, David C, Schumacher S, Heyraud-Nitschke F, Gronenborn B (1995) Geminivirus replication: Genetic and biochemical characterization of rep protein function, a review. Biochimie. 77:765-773.

Lazarowitz SG, Shepherd RJ (1992) Geminiviruses: Genome structure and gene function. Crit Rev Plant Sci. 11:327-349.

Mansoor S, Amin I, Iram S, Hussain M, Zafar Y, Malik KA, Briddon RW (2003a) Breakdown of resistance in cotton to cotton leaf curl disease in Pakistan. Plant Pathol. 52:784-784.

Mansoor S, Briddon RW, 2007. Cotton leaf curl disease. In: Proceedings of the world cotton research conferences-4, Texas, pp. 10-14 september 2007 http://www.icac.org/meetings/wcrc/wcrc2004/presentations/start.htm.

Mansoor S, Briddon RW, Bull SE, Bedford ID, Bashir A, Hussain M, Saeed M, Zafar Y, Malik KA, Fauquet C, Markham PG (2003b) Cotton leaf curl disease is associated with multiple monopartite begomoviruses supported by single DNA ß. Arch Virol. 148:1969-1986.

Mansoor S, Briddon RW, Zafar Y, Stanley J (2003c) Geminivirus disease complexes: an emerging threat. Trends Plant Sci. 8:128-134.

Mansoor S, Zafar Y, Briddon RW (2006) Geminivirus disease complexes: the threat is spreading. Trends Plant Sci. 11:209-212.

Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics. 16:562-563.

Mubin M, Mansoor S, Hussain M, Zafar Y (2007) Silencing of the AV2 gene by antisense RNA protects transgenic plants against a bipartite begomovirus. Virol J. 4:10-10.

Nawaz-ul-Rehman MS, Briddon RW, Fauquet CM (2012) A melting pot of old world begomoviruses and their satellites infecting a collection of Gossypium species in Pakistan. PLoS One. 7:e40050.

Nawaz-ul-Rehman MS, Fauquet CM (2009) Evolution of geminiviruses and their satellites. FEBS Letters. 583:1825-1832.

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology. 265:218-225.

Rajagopalan P, Naik A, Katturi P, Kurulekar M, Kankanallu R, Anandalakshmi R (2012) Dominance of resistance-breaking cotton leaf curl Burewala virus (CLCuBuV) in northwestern India. Arch Virol. 157:855-868.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4:406-425.

Saunders K, Salim N, Mali VR, Malathi VG, Briddon R, Markham PG, Stanley J (2002) Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: Evidence for acquisition of a DNA B component by a monopartite begomovirus. Virology. 293:63-74.

Seal SE, vandenBosch F, Jeger MJ (2006) Factors influencing begomovirus evolution and their increasing global significance: Implications for sustainable control. Crit Rev Plant Sci. 25:23-46.

Stanley J BDM, Briddon RW, Brown JK, Fauquet CM (2005) Geminiviridae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA, Editors., Virus Taxonomy, VIIth Report of the ICTV. Elsevier/Academic Press, London, pp. 301-326.

Tahir MN, Amin I, Briddon RW, Mansoor S (2011) The merging of two dynasties—identification of an African cotton leaf curl disease-associated begomovirus with cotton in Pakistan. PLoS ONE. 6:e20366-e20366.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731-2739.

Zhou X, Liu Y, Calvert L, Munoz C, Otim-Nape GW, Robinson DJ, Harrison BD (1997) Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. J General Virol. 78:2101-2111.

Zhou X, Liu Y, Robinson DJ, Harrison BD (1998) Four DNA-A variants among Pakistani isolates of cotton leaf curl virus and their affinities to DNA-A of geminivirus isolates from okra. J General Virol. 79:915-923.