

Begomovirus and DNA satellites diversity in *Bemisia tabaci* from cotton growing areas of Pakistan

Zafar IQBAL^{1*}, Mariyam MASOOD², Sayyad Ali Raza BUKHARI³,
Imran AMIN⁴, Adil ALSHOAIBI¹, Rob W. BRIDDON⁴

¹King Faisal University, Central Laboratories, Al-Ahsa 31982, Saudi Arabia;
zafar@kfu.edu.sa (*corresponding author); adshoaibi@kfu.edu.sa

²University of Agriculture, Center for Advanced Studies in Agriculture and Food Security, Faisalabad,
Pakistan; mariyam.masood@uaf.edu.pk

³University of Sargodha, Department of Biotechnology, Main Campus, Sargodha, Pakistan; ali.bukhari@uos.edu.pk

⁴National Institute for Biotechnology and Genetic Engineering, Agricultural Biotechnology Division, Faisalabad,
Pakistan; imranamin1@yahoo.com; rob.briddon@gmail.com

Abstract

Cotton leaf curl disease (CLCuD), a major threat to cotton production in Pakistan and northwestern India, is caused by a complex of begomoviruses transmitted by the whitefly *Bemisia tabaci*. These begomoviruses are typically associated with Cotton leaf curl Multan betasatellite (CLCuMuB) and often alphasatellites. Historically, the Burewala strain of Cotton leaf curl Kokhran virus (CLCuKoV-Bur) with the Burewala strain of CLCuMuB (CLCuMuB^{Bur}) was the dominant CLCuD strain in Pakistan. This study investigates the diversity and distribution of begomoviruses and their associated satellites in *B. tabaci* populations across major cotton-growing district of Punjab (including Faisalabad, Burewala, Vehari, and Rahim Yar Khan) and Sindh (Tando Allah Yar, Dadu, Moro, and Sakarand) of Pakistan, providing critical insights into their role in CLCuD epidemiology. The full-length begomoviruses (n = 41), betasatellites (n = 20), and alphasatellites (n = 24) were amplified, cloned and sequenced from *B. tabaci* (n = 56) collected during 2012-2014. Phylogenetic analyses were performed for isolates classification, and recombination events were investigated using Recombination Detection Program (RDP) and SimPlot. Phylogenetic analyses revealed a complex landscape of viral diversity, with Cotton leaf curl Kokhran virus - Burewala strain (CLCuKoV-Bur) (28/41 isolates, 68%) and its cognate betasatellite, Cotton leaf curl Multan betasatellite - Burewala strain (CLCuMuB^{Bur}) (18/20 isolates, 90%), as the dominant species. Cotton leaf curl Multan alphasatellite (14/24 isolates, 58%) was the most prevalent satellite. The predominant *B. tabaci* cryptic species, Asia II 1 (or Asia-Pacific Islands-Australia, (ASIA) putative cryptic species), harbored a greater diversity of viruses and satellites compared to the less prevalent MEAM 1 (or North Africa-Middle East (NAFME) putative cryptic species). Recombination events were observed in 34% of begomovirus isolates and 40% of satellite sequences, underscoring the rapid evolutionary adaptation of these pathogens. The predominant *B. tabaci* cryptic species, Asia II 1 (or Asia-Pacific Islands-Australia, (ASIA) putative cryptic species), harbored a greater diversity of viruses and satellites compared to the less prevalent MEAM 1 (or North Africa-Middle East NAFME putative species). Unlike prior studies, this work is the first to identify full-length genome sequencing of begomoviruses and their associated satellites, isolated directly from *B. tabaci* populations, with cryptic species identification,

Received: 07 May 2025. Received in revised form: 11 Jun 2025. Accepted: 25 Jun 2025. Published online: 30 Jun 2025.

From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

providing a comprehensive analysis of virus-vector associations. These findings enhance our understanding of CLCuD's molecular epidemiology and inform targeted management strategies.

Keywords: alphasatellite; betasatellite; cryptic species; geminivirus; sequencing; whitefly

Introduction

Cotton is the major cash crop of Pakistan, contributing up to 60% of its foreign exchange earnings. Punjab province, the leading cotton producer with approximately 2.5 million hectares, cultivating three times the area of Sindh province. However, since the early 1990s, cotton leaf curl disease (CLCuD) has wreaked havoc on Punjab's yield (Briddon & Markham, 2000). CLCuD was first reported in the vicinity of Multan, Punjab province, in late 1967 and remained localized until 1986 (Hussain *et al.*, 1991). During the 1990s, however, the disease spread rapidly throughout the Punjab and into northern Sindh province and northwestern India. Estimated economic losses in Pakistan between 1992 and 1997 reached US\$5 billion (Briddon and Markham, 2000). By the late 1990s, conventionally bred CLCuD-resistant cotton varieties virtually eliminated these losses (Rahman *et al.*, 2005). Due to the sporadic nature of CLCuD in Sindh, Pakistan, and northwestern India, growers faced a trade-off between planting lower-yielding resistant varieties or higher-yielding susceptible ones, risking losses if the disease was prevalent. Consequently, the adoption of resistant varieties was inconsistent in these regions. This resistance proved short-lived, however, as by 2001, a resistance-breaking strain of CLCuD emerged and spread into northwestern India (Mansoor *et al.*, 2003a; Zaffalon *et al.*, 2012).

CLCuD is a devastating disease of cotton caused by begomoviruses in association with a specific betasatellite, Cotton leaf curl Multan betasatellite (CLCuMuB; (Briddon *et al.*, 2001)). During the first CLCuD epidemic in the 1990s, several begomovirus species were associated with the disease, Cotton leaf curl Multan virus (CLCuMuV) and Cotton leaf curl Kokhran virus (CLCuKoV) being the most prevalent (Mansoor *et al.*, 2003b). These viruses typically require the presence of CLCuMuB to induce severe CLCuD symptoms on *Gossypium hirsutum*. Following the breakdown of resistance in 2001, leading to the second epidemic, a recombinant virus, the "Burewala" strain of CLCuKoV (CLCuKoV-Bur), derived from CLCuKoV and CLCuMuV, emerged as the dominant causative agent of CLCuD in Punjab, Pakistan (Amrao *et al.*, 2010). While CLCuKoV-Bur became dominant in Punjab, greater begomovirus diversity persisted in Sindh and northwestern India, where CLCuKoV-Bur was present alongside other virus species not found in Punjab, likely due to the continued cultivation of susceptible cotton varieties (Amrao *et al.*, 2010; Rajagopalan *et al.*, 2012; Zaffalon *et al.*, 2012). Notably, CLCuKoV-Bur was frequently associated with a recombinant CLCuMuB, the "Burewala" strain (CLCuMuB^{Bur}), resulting from recombination with Tomato leaf curl betasatellite (Amin *et al.*, 2006). It has been reported that resurgence of multiple begomoviruses previously associated with the first epidemic in the 1990s, marking the predicted third CLCuD epidemic (Sattar *et al.*, 2017a). Notably, the bipartite begomoviruses Tomato leaf curl virus (ToLCV) and Tomato leaf curl New Delhi virus have also been reported in CLCuD-infected cotton in Pakistan (Zaidi *et al.*, 2017).

At the present time, there are Mac7 derived lines which showed resistance to CLCuD, are promising tolerant varieties (Zaidi *et al.*, 2020; Aslam *et al.*, 2022; Hussain *et al.*, 2023). The available evidence suggests a reversion to the begomovirus diversity observed in the 1990s, with multiple begomoviruses present in cotton across the affected regions (Zubair *et al.*, 2017; Mahmood *et al.*, 2024). However, CLCuKoV-Bur and CLCuMuB^{Bur} have been the predominant cause of CLCuD in Punjab, Pakistan, from 2001 to 2014.

Begomoviruses are transmitted by the whitefly *Bemisia tabaci* in a circulative, persistent (non-propagative) manner (Czosnek *et al.*, 2017). *B. tabaci* is now recognized as a species complex comprising numerous cryptic species (Judith K. Brown, 2010) that are morphologically indistinguishable but reproductively isolated being comprised of numerous morphologically indistinguishable, but partially

reproductively isolated species (Dinsdale *et al.*, 2010; Alemandri *et al.*, 2012; Firdaus *et al.*, 2013; Hadjistylli *et al.*, 2016; de Moya *et al.*, 2019; Paredes-Montero *et al.*, 2021). The diversity of *B. tabaci* in Pakistan has shown that the cryptic species Asia II 1 (or Asia–Pacific Islands–Australia, (ASIA) putative cryptic species) predominates in Punjab province whereas the exotic MEAM 1 (or North Africa–Middle East NAFME putative species) predominates in Sindh province (Masood *et al.*, 2017; Mahmood *et al.*, 2022). But recently It has been reported that Asia II 1 or (ASIA) dominates in the Sindh province of Pakistan (Mahmood *et al.*, 2022).

The objective of this study was to investigate the diversity and distribution of begomoviruses and their associated satellites within *B. tabaci* populations collected from cotton fields across major growing regions in Punjab and Sindh, Pakistan. By integrating full viral genome sequencing with cryptic species identification, this work provides novel insights into virus-vector associations and the regional molecular epidemiology of CLCuD. This work offers a novel insight into virus-vector associations and regional dynamics of CLCuD epidemiology. These findings can contribute valuable information towards the development of regions specific CLCuD management strategies.

Materials and Methods

Whitefly collection and amplification of mitochondrial (COI) genes

Adult whiteflies were collected from cotton plants using an aspirator between 2012 and 2014 and kept in ethanol (85%) and stored at -80 °C until used. Geographic coordinates for each sampling site were recorded using a GPS device (eTrex 10; Garmin, Schaffhausen, Switzerland). Total DNA (from a single whitefly) was extracted and subjected to amplification of a 780 bp fragment of the mtCOI-3' region, as previously described (Mariyam Masood *et al.*, 2017). The isolated DNA was stored at -30 °C for further use.

Amplification of begomoviruses and DNA satellites

To detect begomoviruses and their associated satellites, initially, rolling circle amplification (RCA) was performed to enrich any circular DNA molecules potentially harboring these viruses, as previously described (Nahid *et al.*, 2008). RCA products were then used as a template in PCR to amplify begomoviruses and satellites with the universal primer pairs BegomoF/BegomoR (Shahid *et al.*, 2007), DNA101/ DNA102 (Bull *et al.*, 2003) and beta01/beta02 (Bridson *et al.*, 2002), respectively. PCR amplicons were visualized on ethidium bromide agarose gels (1%) and products of ~2.8 kb (begomovirus) and ~1.4 kb (satellites), were excised for the subsequent cloning.

Cloning and sequencing of viruses and DNA satellites

The resultant purified PCR amplicons were ligated into the pTZ57R/T TA cloning vector using an InsTAclone PCR Cloning Kit (Thermo Fisher Scientific, Estonia) according to the manufacturer instructions. Ligate mixture (1 µL) was transformed into *E. coli* (DH5α), and positive clones were identified by restriction analysis. Plasmid DNA was isolated from selected clones using an AxyPrep Plasmid Miniprep Kit (Axygen, South Korea) following the manufacturer's protocol. Purified plasmids were sequenced at the Genomics and Bioinformatics Research Unit (USDA-ARS, Stoneville, MS, USA). Sequences were extended by primer walking using custom-designed oligonucleotide primers and assembled using SeqMan (Lasergene, DNASTar Inc., Madison, WI, USA).

Sequence and phylogenetic analysis

Virus and satellite sequences were initially identified using the nucleotide Basic Local Alignment Search Tool (BLASTn) (<http://www.ncbi.nlm.nih.gov/BLASTn>). Final confirmation of begomovirus and DNA satellites were identified using the Sequence Demarcation Tool (SDT ver1.3) according to the methods described by Brown *et al.* (Brown *et al.*, 2015) begomoviruses, by Briddon *et al.* (R. W. Briddon *et al.*, 2018) for alphasatellites, and by (Briddon *et al.*, 2003) for betasatellites. Open reading frames (ORFs) were identified using ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Multiple sequence alignments of each dataset were generated with ClustalW in MEGA 11 and phylogenetic dendrograms were constructed using the Neighbor-Joining algorithm also within MEGA 11 (Tamura *et al.*, 2021).

Detection of recombination

Recombination events in the genomes of isolated begomoviruses, alphasatellites, and betasatellites were identified using the Recombination Detection Program (RDP v5.56) (Martin *et al.*, 2021) and SimPlot v3.5.1 (Lole *et al.*, 1999). RDP simultaneously utilizes a range of non-parametric recombination detection methods: RDP, GENECONV, BootScan, MaxChi, Chimera, SiScan, and 3eq. To ensure reliable detection, only events identified by at least four of these programs with a statistically significant *p*-value (≤ 0.05) were considered credible recombination events. Subsequently, these putative recombination events were further confirmed using similarity plot and bootscan analyses implemented in SimPlot, with a window size of 3000 nucleotides (nt) with a step size of 200 nt. Additionally, Boot Scanning was performed in SimPlot to identify the evidence of sequence identity in the isolated sequences.

Results

Cloning and sequencing of begomoviruses and satellites

An analysis of DNA extracted from fifty-six whiteflies, collected between 2012 and 2014, revealed a distribution of cryptic species using the mitochondrial cytochrome c oxidase 1 gene (Masood *et al.*, 2017). Analysis revealed thirty-seven whiteflies belonged to Asia II 1 (or Asia–Pacific Islands–Australia, (ASIA) putative cryptic species), 16 to MEAM 1 (or North Africa–Middle East NAFME putative species), and 3 to Asia I (Figure 1).

Rolling circle amplification (RCA) was performed on all samples, and agarose gels analysis confirmed high molecular weight product, which was insufficient to digest with different unique cutters to yield 2.8 kb and 1.4 kb fragments. Therefore, RCA products, enriched with circular DNA molecules, were used as template in PCR reactions. This successfully amplified viral DNA in forty-one samples, yielding forty-one putative full-length begomovirus clones (approx. size 2.8), 20 betasatellite clones (~1.4 Kb size), and 24 alphasatellite clones (~1.4 kb size), with specific universal primers (as mentioned previously in MM section). The origin of each clone is detailed in Table 1, and their sequences are available in the nucleotide sequence databases under the accession numbers listed therein. Most insects from which begomovirus, betasatellite, and/or alphasatellite clones were obtained were collected from cotton, with three collected from eggplant (Table 1).

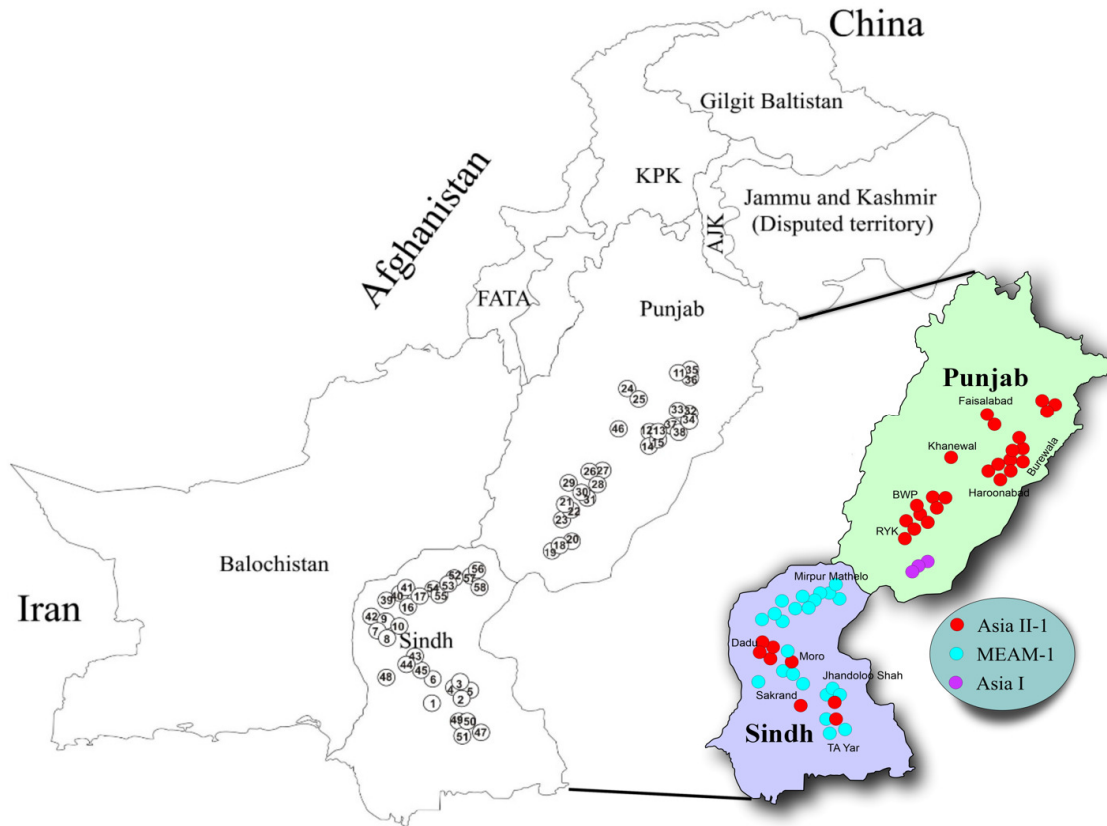


Figure 1. Whitefly sampling sites and species distribution in Pakistan's cotton-growing regions

Table 1. Origins of the begomovirus, betasatellite and alphasatellite clones

Insect no.	Biotype [®]	Host	Location (city/province)	COI accession [®]	Virus identified [*]	Virus (clone no./ accession no.)	Betasatellite identified [*]	Betasatellite (clone no./ accession no.)	Alphasatellite Identified [*]	Alphasatellite (clone no./ accession no.)
1	Asia II 1	Gh	Rashidabad Tayar/Sindh	HG918197	HoLCV	P-566/MH538339	-	-	-	-
2	MEAM 1	Gh	Tando Allahyar/Sindh	HG918201	ChiLCV	P-622/MH538340	ChLCB	P-620/MH411247	AEA	P-577/MH510291
3	Asia II 1	Gh	Sakrand/Sindh	LN832289	CLCuMV-Raj	P-593/MH555071	-	-	-	-
4	Asia II 1	Gh	Sakrand/Sindh	LN832303	CLCuMV-Raj	P-595/MH555070	CLCuMB ^{Sha}	P-629/MH427318	CLCuMuA	P-611/MH517042
5	Asia II 1	Gh	Sakrand/Sindh	LN832304	CLCuMV-Raj	P-596/MH555072	CLCuMB ^{Sha}	P-630/MH427319	-	-
6	Asia II 1	Gh	Sakrand/Sindh	LN832305	CLCuMV-Raj	P-598/MH560503	-	-	-	-
7	Asia II 1	Gh	Faisalabad/Punjab	HF934975	OELCuV	P-107/MH538338	-	-	AYVSGA	P-101/MH517043
8	Asia II 1	Gh	Haroonabad/Punjab	HG915744	AIYVV	P-663/MH538342	-	-	-	-
9	Asia II 1	Gh	Haroonabad/Punjab	HG915745	AIYVV	P-664/MH538343	-	-	-	-
10	Asia II 1	Gh	Haroonabad/Punjab	HG915746	AIYVV	P-665/MH538344	-	-	GDarSLA	P-720/MH517039
11	Asia II 1	Gh	Haroonabad/Punjab	HF935011	AIYVV	P-666/MH538345	-	-	-	-

12	Asia II 1	Gh	Qaziabad/Sindh	LN832306	CLCuKoV-Sha	P-627/ MH538341	-	-	AEA	P-631/ MH517040
13	Asia II 1	Gh	Moro/Sindh	LN832307	CLCuKoV-Sha	P-643/ MH781030	-	-	AEA	P-633/ MH517041
14	Asia II 1	Gh	Rahim Yar Khan/Punjab	LT222297	CLCuKoV-Bu	P-651/ MH766871	CLCuMuB ^{Bur}	P-693/ MH427326	CLCuMuA	P-676/ MH450227
15	Asia II 1	Gh	Rahim Yar Khan/Punjab	HG315647	CLCuKoV-Bu	P-652/ MH766872	CLCuMuB ^{Bur}	P-694/ MH427327	CLCuMuA	P-677/ MH464254
16	Asia II 1	Gh	Rahim Yar Khan/Punjab	HG315648	CLCuKoV-Bu	P-653/ MH766873	-	-	CLCuMuA	P-678/ MH464255
17	Asia II 1	Gh	Kot Samba/Punjab	HF934993	CLCuKoV-Bu	P-667/ MH766876	CLCuMuB ^{Bur}	P-701	GDarSLA	P-609/ MH817849
18	Asia II 1	Gh	Kot Samba/Punjab	HG315657	CLCuKoV-Bu	P-668/ MH766877	-	-	CLCuMuA	P-679/ MH464256
19	Asia II 1	Gh	Kot Samba/Punjab	HG315658	CLCuKoV-Bu	P-669/ MH766878	-	-	-	-
20	Asia II 1	Gh	Khanewal/Punjab	HF934984	CLCuKoV-Bu	P-646/ MH756640	CLCuMuB ^{Bur}	P-658/ MH427323	CLCuMuA	P-644/ MH510284
21	Asia II 1	Gh	Kacha Khuh/Punjab	HF934982	CLCuKoV-Bu	P-661/ MH781029	GuLCB	P-692/ MH411248	CLCuMuA	P-634/ MH510283
22	Asia 1	Sm	Wahi Hussain/Punjab	HF934987	CLCuKoV-Bu	P-649/ MH756643	CLCuMuB ^{Bur}	P-696/ MH427328	CLCuMuA	P-656/ MH510285
23	Asia 1	Sm	Wahi Hussain/Punjab	HF934997	CLCuKoV-Bu	P-650/ MH756644	CLCuMuB ^{Bur}	P-639/ MH427322	CLCuMuA	P-659/ MH510286
24	Asia 1	Sm	Wahi Hussain/Punjab	HG315652	CLCuKoV-Bu	P-647/ MH756641	-	-	-	-
25	Asia II 1	Gh	Firoza/Punjab	HF934992	CLCuKoV-Bu	P-648/ MH756642	CLCuMuB ^{Bur}	P-673/ MH427325	CLCuMuA	P-674/ MH510287
26	Asia II 1	Gh	Firoza/Punjab	HF935005	CLCuKoV-Bu	P-654/ MH766874	CLCuMuB ^{Bur}	P-704/ MH450221	-	-
27	Asia II 1	Gh	Firoza/Punjab	HG315649	CLCuKoV-Bu	P-655/ MH766875	-	-	-	-
28	Asia II 1	Gh	Burewala/Punjab	HG918181	CLCuKoV-Bu	P-709/ MH766879	CLCuMuB ^{Bur}	P-721/ MH450223	CLCuMuA	P-680/ MH510279
29	Asia II 1	Gh	Burewala/Punjab	HG315640	CLCuKoV-Bu	P-710/ MH766880	CLCuMuB ^{Bur}	P-722/ MH450224	-	-
30	Asia II 1	Gh	Burewala/Punjab	LN897417	CLCuKoV-Bu	P-711/ MH766881	-	-	GDarSLA	P-699/ MH517037
31	Asia II 1	Gh	Faisalabad/Punjab	HF934976	CLCuKoV-Bu	P-625/ MH756638	CLCuMuB ^{Bur}	P-637/ MH427321	CLCuMA	P-681/ MH510280
32	Asia II 1	Gh	Faisalabad/Punjab	HF934977	CLCuKoV-Bu	P-626/ MH756639	-	-	GDarSLA	P-108/ LN874303
33	Asia II 1	Gh	Vehari/Punjab	HF935013	CLCuKoV-Bu	P-712/ MH766882	CLCuMuB ^{Bur}	P-706/ MH450222	CLCuMuA	P-690/ MH510281
34	Asia II 1	Gh	Vehari/Punjab	HG315642	CLCuKoV-Bu	P-687/ MH766884	-	-	GDarSLA	P-700/ MH517038
35	Asia II 1	Gh	Dadu/Sindh	LN832312	CLCuKoV-Bu	P-682/ MH766885	CLCuMuB ^{Sha}	P-638/ MH427320	CLCuMuA	P-691/ MH510282
36	Asia II 1	Gh	Dadu, Sindh	LN832309	CLCuKoV-Bu	P-685/ MH766886	CLCuMuB ^{Bur}	P-672/ MH427324	-	-
37	Asia II 1	Gh	Dadu, Sindh	LN832310	CLCuKoV-Bu	P-686/ MH766887	-	-	AEA	P-610/ MK167471
38	Asia II 1	Gh	Sakrand/Sindh	LN832299	CLCuKoV-Bu	P-671/ MH766883	CLCuMuB ^{Bur}	P-723/ MH450225	-	-
39	Asia II 1	Gh	Jhandoloo Shah/Sindh	LN832292	CLCuKoV-Bu	P-702/ MH766888	CLCuMuB ^{Bur}	P-724/ MH450226	-	-
40	Asia II 1	Gh	Jhandoloo Shah/Sindh	LN832293	CLCuKoV-Bu	P-703/ MH766889	-	-	-	-
41	Asia II 1	Gh	Jhandoloo Shah/Sindh	LN832294	CLCuKoV-Bu	P-662/ MH766890	-	-	-	-

*Viruses are denoted as Cotton leaf curl Khokran virus-Burewala strain (CLCuKoV-Bur), Cotton leaf curl Multan virus-Rajasthan strain (CLCuMuV-Raj), Cotton leaf curl Kokhran virus – Shadadpur strain (CLCuKoV-Sha), Alternanthera yellow vein virus (ALYVV), Okra enation leaf curl virus (OELCuV), Chilli leaf curl virus (ChLCV),

Hollyhock leaf curl virus (HoLCV). Betasatellites are denoted as Chili leaf curl betasatellite (ChLCB), Cotton leaf curl Multan betasatellite strain Burewala (CLCuMuBBur) and strain Shahdadpur (CLCuMuBSha). Alphasatellites are denoted as Ageratum enation alphasatellite (AEA), Ageratum yellow vein Singapore alphasatellite (AYVSGA), Cotton leaf curl Multan alphasatellite (CLCuMuA) and *Gossypium darwini* symptomless alphasatellite (GDaSLA); \$ Host plant species from which whiteflies were collected are denoted as *Gossypium hirsutum* (Gh) and *Solanum melongena* (Sm; eggplant); @ As determined by Masood *et al.* (2017)

Characterization of begomoviruses isolated from whiteflies

SDT analysis of the forty-one virus sequences grouped them into five distinct species (Figure 2), based on a species demarcation threshold of 91% nucleotide sequence identity for begomoviruses (Judith K Brown *et al.*, 2015). The groups were consisted as follows:

Group 1 - P-663, P-664, P-665, P-666 (4 isolates)

Group 2 - P-622 (1 isolate)

Group 3 - P-566 (1 isolate)

Group 4 - P-107 (1 isolate)

Group 5 - P-593, P-595, P596, P-598 (4 isolates).

Group 6 - P-668, P-625, P-626, P-646, P-647, P-648, P-649, P-650, P-651, P652, P653, P-654, P-655, P-661, P-667, P-669, P-671, P-682, P-685, P-686, P-687, P-702, P-703, P-662, P-709, P-710, P-711, P-712, P-627, P-643 (30 isolates).

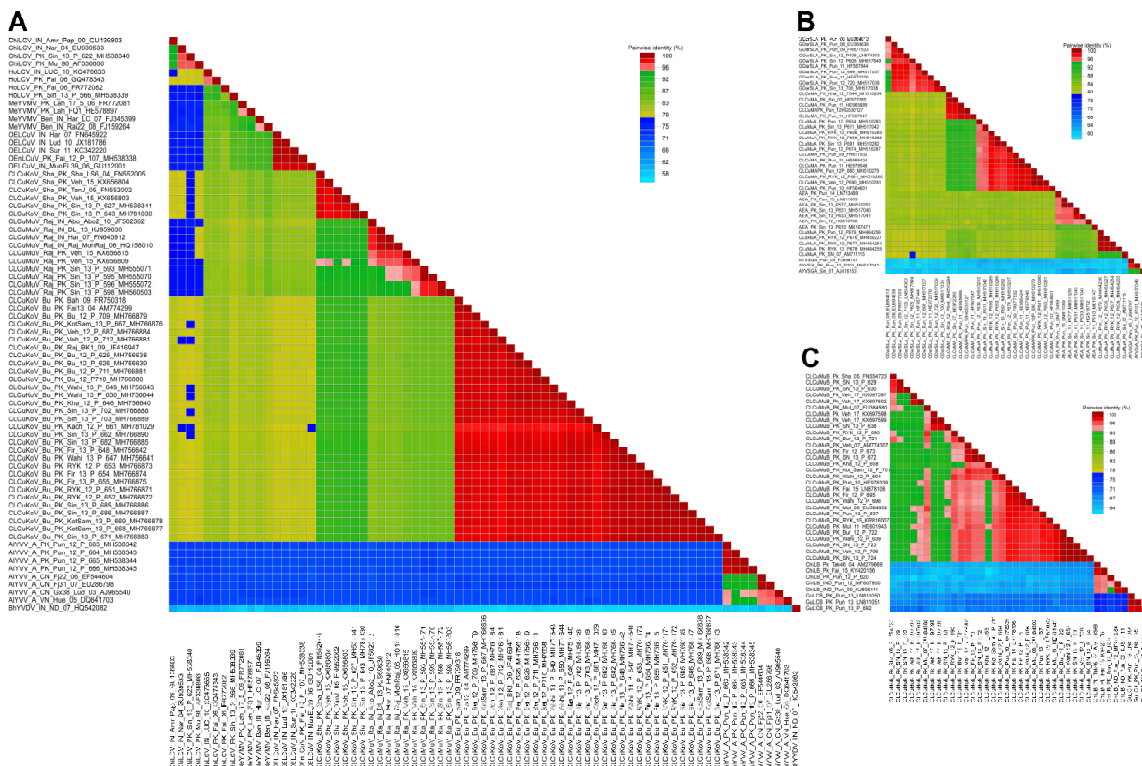


Figure 2. SDT analysis of the isolated begomoviruses (A), betasatellites (B), and alphasatellites (C)

Comparison of Group 1 sequences with database entries using SDT revealed high identity to Alternanthera yellow vein virus (AIYVV), with the highest (95.8–96.5%) to an *Eclipta prostrata* isolate from India (EU188920), confirming that P-663, P-664, P-665 and P-666 sequences as AIYVV isolates.

SDT analysis of P-622 (Group 2) showed high sequence identity to Chili leaf curl virus (ChiLCV) sequences in the databases. Identity was higher to the “Pakistan” strain of ChiLCV than to the other strains,

with highest identity to two ChiLCV-PK isolates from *Capsicum annuum* in India (EU939533, (Senanayake *et al.*, 2007)) and cotton in Pakistan (KY420138), at 96.4 and 96% identity, respectively, identifying P-622 as a ChiLCV-PK.

P-566 (Group 3) showed relatively high identity to Mesta yellow vein mosaic virus (MeYVMV; KT948076; 90.6%) and Hollyhock leaf curl virus (HoLCV; GQ478343; 90.8%). Based on the criteria of Brown *et al.* (Judith K Brown *et al.*, 2015), P-566 is classified as HoLCV due to its higher identity. However, it is likely that, in future revision of begomovirus taxonomy, the species MeYVMV and HoLCV will need to be merged.

Sequence of P-107 (Group 4) revealed high sequence identity to numerous (>100) okra enation leaf curl virus (OEnLCV) sequences available in the databases. Notably, P-107 shared 99.9% sequence identity with two Indian OEnLCV isolates (KC019308, KC019309), strongly indicating its classification as an OEnLCV isolate.

Isolates in Group 5 exhibited high sequence identity to Cotton leaf curl Multan virus (CLCuMuV), with the highest (90.5 to 94.5%) observed with the isolates of “Rajasthan” strain (CLCuMuV-Raj). This confirms that P-593, P-595, P-596, P-598 are CLCuMuV-Raj.

SDT analysis of Group 6 isolates revealed the highest identity to Cotton leaf curl Kokhran virus (CLCuKoV) isolates. Specifically, two isolates (P-627 and P-643) exhibited overall higher identity to the “Shahdadpur” strain of CLCuKoV (CLCuKoV-Sha) than to other CLCuKoV strains, confirming their identification as CLCuKoV-Sha. Conversely, the remaining twenty-eight isolates showed higher sequence identity to the “Burewala” strain of CLCuKoV (CLCuKoV-Bur) indicating that they are isolates of CLCuKoV-Bur strain. These isolates constitute the biggest group of isolates.

Species identification was corroborated by phylogenetic analysis (Figure 3A), which demonstrated that each virus sequence obtained from whiteflies clustered with corresponding reference sequences from public databases.

The obtained virus sequences were analyzed for the predicted amino acid sequence using the ORF Finder tool. All sequences except CLCuKoV-Bur exhibited the typical Old World (OW) begomovirus genome (or DNA A component) organization, with six predicted genes encoding proteins of the expected size (Table S1). In contrast, ORF Finder analysis of CLCuKoV-Bur isolates revealed a significantly truncated TrAP gene, encoding a predicted 35-amino-acid TrAP instead of the typical 134 amino acids. This truncated Transcriptional Activator Protein (TrAP) is characteristic of CLCuKoV-Bur isolates associated with resistance breakdown in cotton (L. Amrao *et al.*, 2010).

Consistent with previously characterized CLCuKoV-Sha isolates (Luqman Amrao *et al.*, 2010), the TrAP gene of the two isolates obtained in this study is predicted to encode a 150-amino-acid protein (Table S1), rather than the more common 134-amino-acid TrAP found in OW begomoviruses.

Characterization of satellites isolated from whiteflies

A total of forty-four cloned fragments, ranging from 1.3 to 1.4 kb, were amplified from whiteflies. ORF Finder analysis revealed a single large, virion-sense ORF of ~950 nucleotides in twenty-four of these sequences, consistent with other alphasatellites genome organization (Table S2). The remaining twenty clones contained numerous small ORFs but also a single, conserved, complementary-sense ORF encoding at least 118 amino acids, characteristic of betasatellites (Table S2).

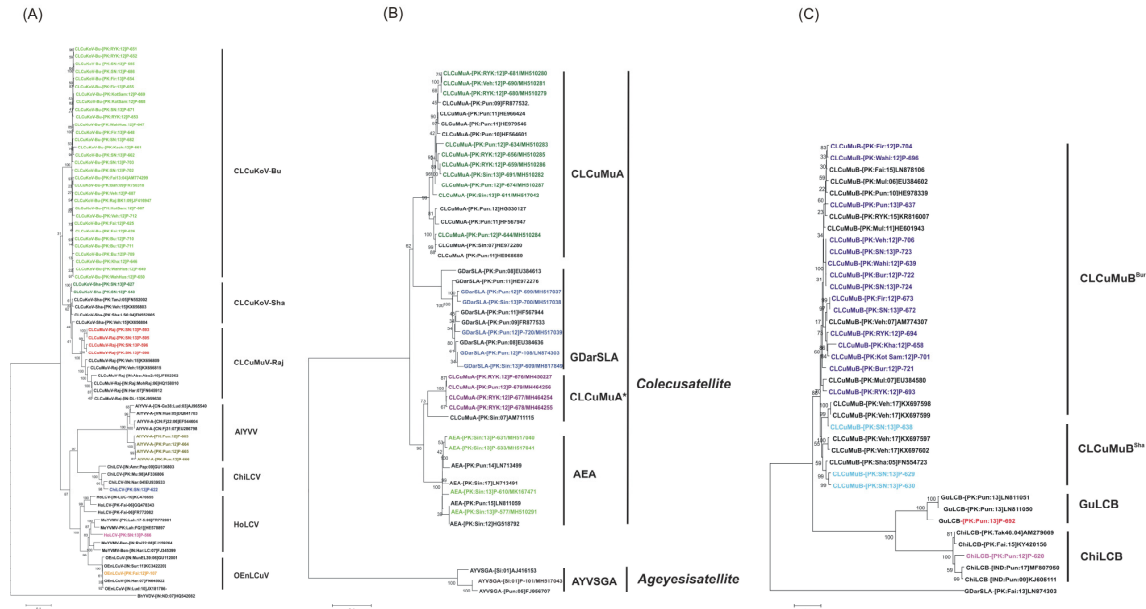


Figure 3. Phylogenetic analysis of full-length begomovirus (A), alphasatellite (B), and betasatellite (C), sequences obtained from whiteflies

The viruses are indicated as Alternanthera yellow vein virus (AIVVV), chili leaf curl virus (ChiLCV), cotton leaf curl Kokhran virus strain burewala (CLCuKoV-Bur), cotton leaf curl Kokhran virus strain Shahdadpur (CLCuKoV-Sha), cotton leaf curl Multan virus strain Hisar (CLCuMuV-His), cotton leaf curl Multan virus strain Pakistan (CLCuMuV-PK), cotton leaf curl Multan virus strain Rajasthan (CLCuMuV-Raj), hollyhock leaf curl virus (HoLCV), mesta yellow vein mosaic virus (MeYVMV) and okra enation leaf curl virus (OEnLCV). The alphasatellites are given as Ageratum enation alphasatellite (AEA), Ageratum yellow vein Singapore alphasatellite (AYVSGA), cotton leaf curl Multan alphasatellite (CLCuMuA), *Gossypium darwinii* symptomless alphasatellite (GDarSLA). The betasatellites are given as Cotton leaf curl Multan betasatellite (CLCuMuB), chili leaf curl betasatellite (ChiLCB) and the proposed species guar leaf curl betasatellite (GuLCB). The strains of CLCuMuB are given as “Burewala” (CLCuMuB^{Bur}) and “Shahdadpur” (CLCuMuB^{Sha})

Phylogenetic trees were rooted using (a) bhendi yellow vein Delhi virus DNA B, (b) AYVSGA-[PK:Pun:05], and (c) an isolate of GDarSLA as outgroups. Branch lengths are proportional to the calculated evolutionary distance, while vertical distances are arbitrary. Numbers at nodes represent bootstrap support values based on 1000 replicates. Sequences generated in this study are color-coded, while sequences obtained from databases are shown in black

SDT analysis of the twenty-four alphasatellite sequences grouped them into five distinct clusters, based on the established species demarcation threshold of 88% for alphasatellites (Briddon *et al.*, 2018).

- Group 1** - P-609, P-699, P-700, P-720, P-108 (5 isolates)
- Group 2** - P-577, P-631, P-633, P-610 (defective molecule) (4 isolates)
- Group 3** - P-676, P-677, P-678, P-679 (4 isolates)
- Group 4** - P-644, P-611, P-634, P-674, P-680, P-681, P-659, P-690, P-691, P-656 (10 isolates)
- Group 5** - P-101 (1 isolate)

The five sequences comprising Group 1 shared >90% sequence identity, indicating they represent isolates of a single alphasatellite species, consistent with established guidelines of Briddon *et al.* 2018 (R. W. Briddon *et al.*, 2018). BLASTn analysis against public databases revealed high sequence identity to isolates of *Gossypium darwinii* symptomless alphasatellite (GDarSLA), confirming their classification as GDarSLA. Phylogenetic analysis (Figure 3B) further supported this identification, demonstrating that Group 1 sequences clustered with known GDarSLA isolates.

The four sequences comprising Group 2 exhibited 95.2-100% identity, classifying them as isolates of a single alphasatellite species. BLASTn analysis against database sequences revealed high sequence identity (>95% identity) to isolates of *Ageratum enation* alphasatellite (AEA). Following the recommendations of Briddon *et al.* (R. W. Briddon *et al.*, 2018), this high identity confirms that Group 2 sequences are AEA isolates. Phylogenetic analysis (Figure 3B) further supported this conclusion, demonstrating that the Group 2 sequences clustered with selected isolates of AEA. One AEA sequence (P-610) was shorter than typical alphasatellite sequences, likely representing a defective deletion mutant and was therefore excluded from the phylogenetic analyses.

SDT analysis of all Cotton leaf curl Multan alphasatellite (CLCuMuA) sequences available in public databases, including those designated as Group 3 and Group 4, revealed greater than 88% sequence identity between each isolate and at least one previously characterized CLCuMuA sequence. These findings confirm the classification of both Group 3 and Group 4 sequences as CLCuMuA isolates. The distinct grouping of the Group 3 isolates observed in the SDT analysis was also supported by phylogenetic analysis (Figure 3B), which placed these four sequences with Cotton leaf curl Shahdampur alphasatellite (CLCuMuA; AM711115) (Luqman Amrao *et al.*, 2010), exhibiting >99% identity to this isolate. This unusual phylogenetic clustering of the Group 3 sequences likely results from recombination. Recombination Detection Program (RDP) analysis of all available CLCuMuA sequences implicated a recombination event in the Group 3 sequences, with a significant fragment derived from AM711115, although the parental sequences and direction of recombination could not be definitively determined.

The P-101 sequence (Group 5) exhibited high sequence identity (>91%) to only three *ageratum* yellow vein Singapore alphasatellite (AYVSGA) sequences available in the databases. The highest identity (97.7%) was with an isolate from Oman isolated from tomato (FJ956707). All other alphasatellite sequences in the databases showed less than 78.7% identity to P-101. This strong sequence similarity and phylogenetic analysis grouping P-101 with known AYVSGA isolates (Figure 3B) strongly suggests that P-101 is an AYVSGA isolate (Figure 3B).

Analysis of the putative betasatellite sequences using SDT revealed three distinct groups (Figure 3C). This aligns with the proposed species demarcation threshold of 91% nucleotide sequence identity for betasatellites (<https://ictv.global/proposals-16/2016.021a-kP.A.v2.Tolecusatellitidae>).

Group 1 - P-620 (1 isolate)

Group 2 - P-692 (1 isolate)

Group 3 - P-629, P-630, P-637, P-638, P-639, P-658, P-672, P-673, P-693, P-694, P-696, P-701, P-704, P-706, P-721, P-722, P-723, P-724 (18 isolates)

Sequence analysis using SDT revealed that P-620 (Group 1) shared high nucleotide sequence identity with Chili leaf curl betasatellite (ChiLCB) isolates, exhibiting 95.8% identity with two isolates (AM279662 and AM279663) from chili (*Capsicum annuum*) in Pakistan. This confirms that P-620 is a ChiLCB isolate.

SDT analysis revealed that P-692 (Group 2) clustered closely with two unclassified betasatellite isolates from cluster bean (*Cyamopsis tetragonoloba*) in Pakistan (LN811050 and LN811051), exhibiting high nucleotide sequence identity (98.9% and 99.1%, respectively). Sequence identity levels with all other betasatellites were below 86%, suggesting that LN811050, LN811051, and P-692 represent a distinct, unclassified species, which we propose to name guar leaf curl betasatellite (GuLCB). Phylogenetic analysis further supported this classification, with P-692 clustering with LN811050 and LN811051, forming a distinct clade separate from all other sequences (Figure 3C).

The eighteen sequences in Group 3 shared 90.3-100% sequence identity, confirming their classification as isolates of a single species based on the 91% species demarcation threshold. These sequences exhibited high sequence identity to CLCuMuB available in the databases, indicating they are CLCuMuB isolates. Alignment analysis with representative isolates of the "Multan" (CLCuMuB^{Mul}, AJ298903) and "Burewala" (CLCuMuB^{Bur}, FN554719) strains of CLCuMuB revealed that most isolates possessed the tomato leaf curl

betasatellite (ToLCB) recombinant fragment characteristic of CLCuMuB^{Bur} (Figure S1) (Akhtar *et al.*, 2014; Amin *et al.*, 2006). However, three isolates (P-627, P-643, and P-693) had a reduced ToLCB fragment, typical of the "Shahdadpur" strain of CLCuMuB (CLCuMuB^{Sha}, Figure 3C) (Akhtar *et al.*, 2014; Luqman Amrao *et al.*, 2010). Phylogenetic analysis further supported the classification of Group 3 isolates as CLCuMuB, clustering the majority with previously characterized CLCuMuB^{Bur} isolates and the three atypical isolates (P-627, P-643, and P-693) with CLCuMuB^{Sha}.

Interestingly, three isolates (P-627, P-643, and P-693) possessed a reduced ToLCB fragment typical of the "Shahdadpur" strain. Phylogenetic placement consistently grouped all Group 3 isolates with CLCuMuB, further supporting their classification, with the three exceptional isolates clustering with the "Shahdadpur" lineage.

Recombination analysis of virus isolates

Recombination analysis of all virus isolates was performed using RDP and SimPlot. Among AIYVV isolates, two (P663 and P664) exhibited a single putative recombination event within the 1335-1615 nucleotide region, while P665 showed a recombination event at positions 2733-18. In these cases, CLCuKoV-Bu was identified as the major parent and MeYVMV as the minor parent (Table 2). SimPlot analysis corroborated these findings, identifying breakpoint regions at 200-400, 1200-1400, and 2600-3000 with similarity to CLCuKoV-Bu, and at 500-1000 and 1400-1800 with similarity to MeYVMV. Notably, the SimPlot analysis consistently indicated a high overall similarity of over 80% between CLCuKoV-Bu, AIYVV, and the specific region (Figure 4).

The Bootscan plots revealed highly fragmented genomes, indicative of frequent recombination events. Additionally, SimPlot analysis demonstrated sequence similarity between the query sequence and other sequences. Nucleotide positions in the plots refer to the alignment, excluding gaps and repeat regions.

Recombination analysis revealed CLCuMuV isolated from Punjab and Sindh to be highly recombinant. RDP analysis identified four putative recombination breakpoints across the entire genome in both populations (Table 2). SimPlot corroborated these findings, indicating several recombination events, some coinciding with RDP-identified breakpoints and others occurring at distinct positions. Notably, CLCuMuV-Raj isolates exhibited the highest nucleotide sequence identity to CLCuKoV-Bu and the lowest to AIYVV (Figure 4). Among CLCuKoV-Bu isolates, three distinct recombination patterns were identified. The first group exhibited two recombination events between nucleotide positions 192-427 and 1246-1515, with CLCuMuV-Raj as major parent and HoLCV as minor parent. The second group also displayed two recombination events, located at nucleotide positions 1255-1500 and 3073-224, involving the same parental combination. The third group displayed a more complex recombination profile with three recombination events. Two of these events involved CLCuMuV-Raj and HoLCV as major parents, with breakpoints at positions 192-427 and 139-1500. The third event was unique, incorporating sequences from both MeYVMV and CLCuMuV-Raj, with breakpoints at positions 2014-2687 (Table 2). SimPlot analysis further supported the presence of multiple recombination events, with some overlapping and others occurring independently in different regions (Figure 4).

Table 2. Multiple recombination events detected by RDP in the isolated begomoviruses

Virus	Accession number	No. of Recomb. events	Breakpoints		Parents		Detection Methods ^a	p-value ^b
			Begin	End	Major	Minor		
AIYVV	MH538343/44/45	2	1335	1615	CLCuKoV-Bu (MH766888)	MeYVMV (MH781029)	RGMS	1.03×10 ⁻⁷⁰
			2733	18	CLCuKoV-Bu (MH766887)	AIYVV (MH538344)	RGMS	1.03×10 ⁻⁷⁰
ChiLCV	MH538340	3	1978	2644	MeYVMV (FR72081)	AIYVV (AJ965540)	RMCS3	3.47×10 ⁻⁰⁴
			2027	2435	AIYVV (MH538342)	CLCuKoV (MH766881)	RGCS	9.03×10 ⁻⁰⁸
			149	1327	MeYVMV (FJ345399)	CLCuKoV (MH766879)	RGBMCS3	1.08×10 ⁻²⁵
HoLCV		3	196	427	CLCuMuV-Raj (MH555072)	OELCuV (GU112001)	MCS3	7.33×10 ⁻⁰⁶
			598	1112	HoLCV (KC476655)	Unknown	RGBMCS3	5.91×10 ⁻³⁰
			1827	1	MeYVMV (HE578897)	HoLCV (KC476655)	RGBMCS3	2.51×10 ⁻³⁹
OEnLCuV		2	1804	2247	CLCuMuV-Raj (JF502362)	CLCuKoV-Bu (MH766881)	RMCS3	6.94×10 ⁻⁰⁷
			2251	2399	ChiLCV (MH538340)	MeYVMV (FR72081)	RGBMCS3	105×10 ⁻⁰⁹
CLCuMuV-Raj (Punjab)		4	482	1062	HoLCV (MH538339)	CLCuKoV-Bu (FR750318)	RGBMCS3	8.63×10 ⁻³²
			1045	2005	CLCuMuV-Raj (JF502362)	CLCuKoV (MH766884)	RGBMCS3	3.66×10 ⁻²⁵
			1207	1442	CLCuKoV-Bu (FR750318)	OELCuV (GU112001)	RGBMCS3	7.48×10 ⁻¹²
			1492	1870	Unknown	CLCuKoV-Bu (MH766888)	RGMS	1.48×10 ⁻²²
CLCuMuV-Raj (Sindh)		4	376	1252	OELCuV (GU112001)	CLCuKoV-Bu (FR750318)	RGBMCS3	1.93×10 ⁻¹²
			1492	1871	AIYVV (MH538345)	CLCuKoV-Bu (MH766888)	RGBMCS3	1.47×10 ⁻²²
			1874	2476	CLCuKoV-Bu (FR750318)	OELCuV (GU112001)	BGMCS3	7.48×10 ⁻¹²
			2184	2493	Unknown	HoLCV (FR72082)	RGBMCS3	1.05×10 ⁻⁰⁹
CLCuKoV-Bu (G1)		2	192	427	CLCuMuV-Raj (KX656809)	Unknown	RMCS3	4.88×10 ⁻⁰⁴
			1246	1515	CLCuMuV-Raj (JF503282)	HoLCV (FR72082)	RBMCS3	2.47×10 ⁻⁰⁴
CLCuKoV-Bu (G2)		2	1255	1500	HoLCV (FR72082)	CLCuMuV-Raj (JF502362)	RBMCS3	9.40×10 ⁻⁰⁴
			3073	224	HoLCV (FR72082)	CLCuMuV-Raj (JF502362)	RMCS3	4.88×10 ⁻⁰⁴
CLCuKoV-Bu (G3)		3	192	427	CLCuMuV-Raj (KX656809)	MeYVMV (FH345399)	RBMCS3	9.40×10 ⁻⁰⁴
			1239	1500	HoLCV (FR72082)	CLCuMuV-Raj (JF502362)	RMCS3	4.89×10 ⁻⁰⁴
			2014	2687	AIYVV (DQ641703)	CLCuMuV-Raj (KX656815)	GMCS	1.98×10 ⁻¹⁷

^a B, Bootscan; C, Chimaera; G, GeneConv; L, LARD; M, MaxChi; P, Phylpro; R, RDP; S, SisScan; 3, 3SEQ.

^b The lowest p-value corresponds to the recombination program (bold) is mentioned

CLCuKoV-Bu (G1) = MH766881 (P711); MH756638 (P625); MH756639 (P626); MH756642 (P648); MH781029 (P661); MH766873 (P653); MH766875 (P656); MH766877 (P668); MH781029; MH766873; MH766888 (P702); MH766889 (P703); MH766884 (P687); and MH756641 (P649).

CLCuKoV-Bu (G2) = MH756640 (P646); MH766876 (P668); MH766879 (P709); MH766880 (P710); MH766885 (P682); MH766887 (P686); MH766881 (P712); MH756642 (P650); MH766871 (P651); MH766872 (P652)

CLCuKoV-Bu (G3) = MH766873 (P653); MH766874 (P654); MH766878 (P669); KX656803; KX656804; MH766883 (P671); MH766886 (P685)

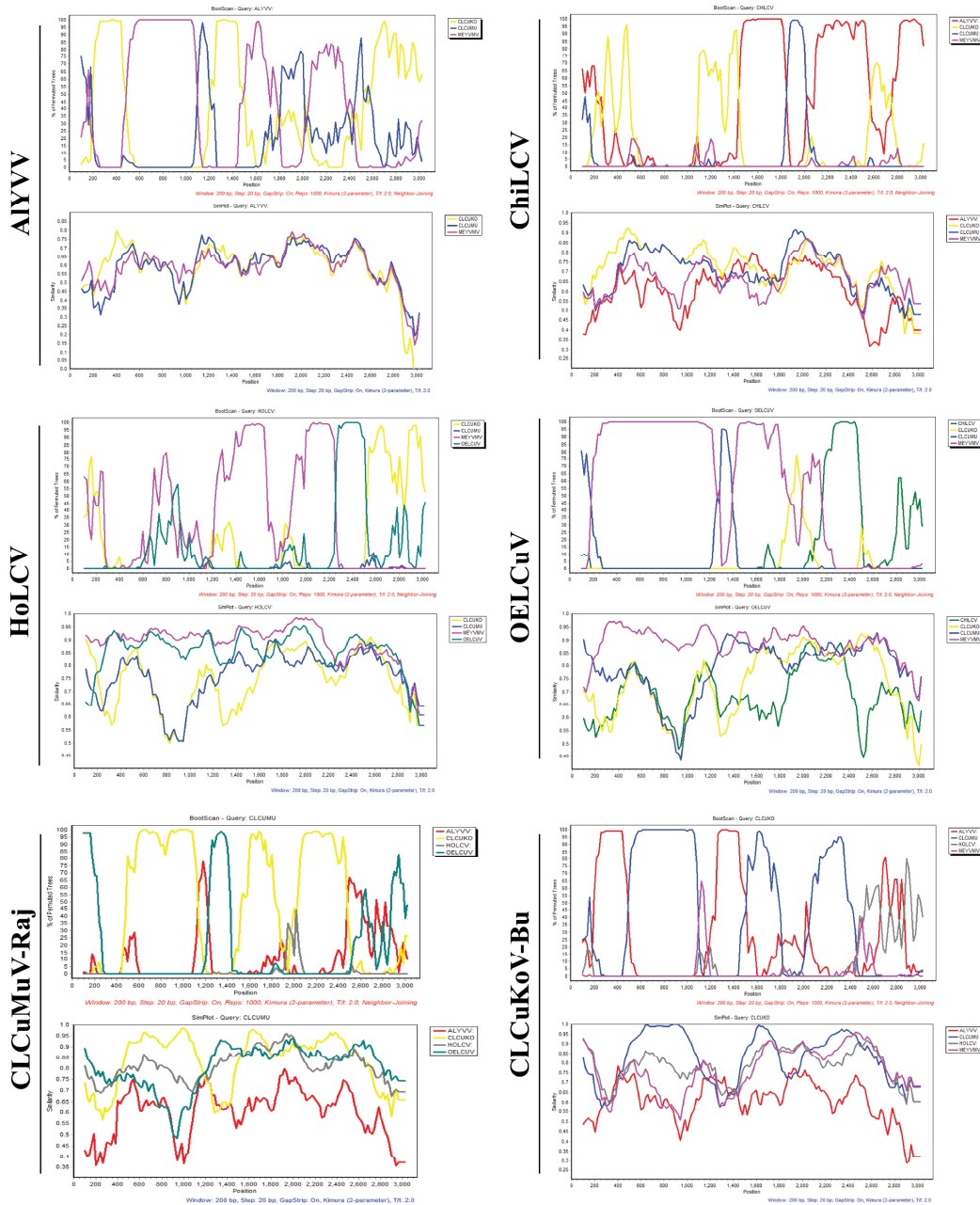


Figure 4. Recombination analysis using Bootscan and SimPlot of full-length genome sequence of isolated begomoviruses, including AIYVV, ChiLCV, HoLCV, OELCuV, CLCuMuV-Raj, and CLCuKoV-Bu

Recombination analysis of DNA satellites

Alphasatellite isolates exhibited evidence of putative recombination events as determined by both RDP and SimPlot analyses (Table 3, Figure S2). GDarSLA isolates shared a putative recombination event at the 700-900 nucleotide position, with CLCuMuA and GDarSLA as potential parents, as identified by both methods. Interestingly, while RDP detected a single putative recombinant event in AYSVGA at the 877-129 nucleotide position, SimPlot did not identify any recombination in this isolate. In contrast, all AEA isolates displayed two recombination events at distinct genomic locations (Table 3).

Table 3. Multiple recombination events detected by RDP in the isolated alphasatellites

Satellite	Accession number	No. of Recomb. events	Breakpoints		Parents		Detection Methods ^a	<i>p</i> -value ^b
			Begin	End	Major	Minor		
GDarSLA	MH517038, MH517039	1	725	836	GDarSLA (FR877533)	CLCuMuA (MH510281)	RGBMC3	1.23×10 ⁻⁰⁸
AEA	MH510291	2	777	1288	CLCuMuA (MH450227)	GDarSLA (MH817849)	RGBMCS3	1.47×10 ⁻¹²
			1313	178	CLCuMuA (MH510285)	CLCuMuA (MH464254)	BMCS3	3.28×10 ⁻⁰⁴
	MK167471	2	1205	1278	CLCuMuA (MH464255)	GDarSLA (HF567944)	RGMCS3	9.28×10 ⁻⁰⁴
			1313	178	CLCuMuA (MH510285)	CLCuMuA (MH464254)	BMCS3	3.28×10 ⁻⁰⁴
	MH517040, MH517041	2	229	326	CLCuMuA (MH510285)	CLCuMuA (MH464254)	BMCS3	3.28×10 ⁻⁰⁴
			940	1300	CLCuMuA (MH450227)	GDarSLA (MH817849)	RGBMCS3	1.47×10 ⁻¹²
CLCuMuA	MH510285	1	509	736	CLCuMuA (MH510287)	AEA (MH517040)	RBMC3	1.28×10 ⁻⁰⁷
	MH464256, MH450227	2	275	572	CLCuMuA (MH464256)	GDarSLA (EU384613)	RGMCS3	2.88×10 ⁻²⁰
			1190	1258	CLCuMuA (MH464255)	CLCuMuA (AM711115)	MCS3	2.56×10 ⁻⁰³
MH464254, MH464255	1	264	572	CLCuMuA (HF564601)	GDarSLA (EU384613)	RGCS	2.88×10 ⁻²⁰	
AYVSGA		1	877	129	Unknown	GDarSLA (MH517037)	RMCS3	2.54×10 ⁻⁰⁷

Although SimPlot indicated a potential recombinant region at the 600-800 nucleotide position, it did not fully corroborate the RDP findings. Five out of the CLCuMuA isolates showed potential recombination events at multiple nucleotide positions (509-736, 275-572, 1190-1258, and 264-572) according to RDP analysis, with most of these events supported by SimPlot results.

Betasatellite isolates exhibited evidence of putative recombination events as determined by both RDP and SimPlot analyses (Table 4, Figure S2). The GuLCB isolate shared a putative recombination event spanning nucleotides 19 to 666, with CLCuMuB and ChiLCB as potential parents, although this event was not confirmed by SimPlot. Intriguingly, seven CLCuMuB isolates displayed evidence of intergenomic recombination. Isolate MH427323 exhibited three recombination events at positions 106-217, 857-1076, and 1168-1313. Five other isolates shared a putative recombination event at position 1155-1289, while isolate MH427325 showed a broader recombination event spanning nucleotide 834-1460. Most of these recombination events were supported by SimPlot analysis.

Co-occurrence of virus and satellites, and association with B. tabaci

The co-occurrence of the isolated viruses, betasatellites, and alphasatellites in *B. tabaci* (n=41) is presented (Table 5). CLCuKoV-Bur was the most prevalent virus (n=28), predominantly associated with CLCuMuB^{Bur} (n=15). Other viruses like AIYVV (n=4), ChiLCV (n=1), and OELCuV (n=1) often lack associated betasatellites. CLCuMuA was the most common alphasatellite and showed highest association (n=11) to CLCuKoV-Bur and CLCuMuB^{Bur}. Rare associations included CLCuMuB^{Sha}, GuLCB, AEA, and GDarSLA. Notably, a significant portion of insects harbored unassociated entities, suggesting potential

interactions with unknown or unrelated hosts, underscoring the intricate nature of these viral associations within the whitefly vector.

Table 4. Multiple recombination events detected by RDP in the isolated betasatellites.

Satellite	Accession number	No. of Recomb. events	Breakpoints		Parents		Detection Methods ^a	<i>p</i> -value ^b
			Begin	End	Major	Minor		
GuLCB	MH411248	2	19	666	CLCuMuB (MH427327)	ChiLCB (KJ60511)	RMCS ₃	6.71×10 ⁻³³
			1026	1257	CLCuMuB (MH427320)	ChiLCB (KY420156)	RMCS ₃	6.52×10 ⁻⁰⁹
CLCuMuB	MH427323	3	106	217	CLCuMuB (MH450220)	CLCuMuB (KX697598)	RGBMCS ₃	3.21×10 ⁻¹¹
			857	1076	CLCuMuB (MH450226)	CLCuMuB (MH427328)	GMCS ₃	2.54×10 ⁻⁰⁴
			1168	1313	CLCuMuB (KX697598)	CLCuMuB (MH427322)	RGBMCS ₃	3.58×10 ⁻¹²
	MH427321, MH427326, MH427327, MH427328, MH450225, MH427325	1	1155	1289	CLCuMuB (MH427322)	ChiLCB (MH411248)	RGBMCS ₃	5.1×10 ⁻¹⁵
			1	834	1460	CLCuMuB (MH450224)	CLCuMuB (MH427318)	MCS ₃

Table 5. Summary of the relationship of identified begomoviruses, alpha- and betasatellites.

Virus (no.)* [n = 41]	Betasatellite (no.)* [n = 20]	Alphasatellite (no.)* [n = 24]
AIYVV (4)	none (4)	none (3)
		GDarSLA (1)
ChiLCV (1)	ChiLCB (1)	AEA (1)
CLCuKoV-Bur (28)	CLCuMuB ^{Bur} (15)	CLCuMuA (09)
		none (5)
		GDarSLA (1)
		CLCuMuA (1)
	CLCuMuB ^{Sha} (1)	CLCuMuA (1)
	none (9)	none (3)
		GDarSLA (2)
		CLCuMuA (2)
GuLCB (1)	AEA (1)	
CLCuKoV-Sha (2)	none (2)	CLCuMuA (1)
		AEA (1)
CLCuMuV-Raj (4)	CLCuMuB ^{Sha} (2)	CLCuMuA (1)
		none (1)
	none (2)	none (2)
HoLCV (1)	none (1)	none (1)
OELCuV (1)	none (1)	AYVSGA (1)

*Number of insects harboring each virus/satellite.

Analysis of 41 *B. tabaci* individuals revealed a diverse distribution of viruses, betasatellites, and alphasatellites across cryptic species (Table 6).

Table 6. Summary of the virus, betasatellite and alphasatellites harboured by each *B. tabaci* cryptic species.

Cryptic species (no.) [n = 41]	Virus (no.) [n = 41]	Betasatellite (no.) [n = 20]	Alphasatellite (no.) [n = 24]
Asia II 1 (or ASIA) (37)	CLCuKoV-Bur (25)	CLCuMuB ^{Bur} (13)	CLCuMuA (8)
			GDarSLA (2)
			none (5)
		none (10)	CLCuMuB ^{Sha} (1)
			GuLCB (1)
			CLCuMuA (1)
	CLCuKoV-Raj (4)	CLCuMuB ^{Sha} (2)	GDarSLA (2)
			CLCuMuA (2)
		none (2)	none (6)
			CLCuMuA (1)
			none (1)
			none (2)
CLCuKoV-Sha (2)	none (2)	AEA (2)	
AIYVV (4)	none (4)	GDarSLA (1)	
HoLCV (1)	none (1)	none (3)	
OELCuV (1)	none (1)	none (1)	
Asia 1 (3)	CLCuKoV-Bur (3)	CLCuMuB ^{Bur} (2)	AYVSGA (1)
		none (1)	CLCuMuA (2)
MEAM 1 (NAFME) (1)	ChiLCV (1)	ChiLCB (1)	none (1)
			AEA (1)

The predominant Asia II 1 (or Asia–Pacific Islands–Australia, (ASIA) putative cryptic species) cryptic species (n = 37) exhibited high CLCuKoV-Bur prevalence (n=25), frequently associated with CLCuMuB^{Bur} (n = 13) and CLCuMuA (n = 8). Other viruses, such as CLCuKoV-Sha and CLCuKoV-Raj, were detected with less common satellite associations. Asia 1 (n=3) primarily harbored CLCuKoV-Bur with strong CLCuMuB^{Bur} and CLCuMuA associations. The MEAM 1 (or North Africa–Middle East NAFME putative species) cryptic species (n = 1) uniquely harbored ChiLCV associated with ChiLCB and AEA. Notably, unassociated entities were observed across cryptic species, highlighting the complex epidemiology of these plant pathogens.

Discussion

This study involved the isolation and full-length sequencing of begomoviruses, betasatellites, and alphasatellites from *B. tabaci* (cryptic species) collected across major cotton-growing regions of Punjab and Sindh provinces in Pakistan. Whiteflies were primarily collected from cotton (*G. hirsutum*) plants in CLCuD-affected areas. This work complements a previous analysis of *B. tabaci* cryptic species diversity in Pakistan (Mariyam Masood *et al.*, 2017) and a broader study of begomovirus and satellite diversity across the country (manuscript in preparation). In addition to these, recombination analysis of the isolates was performed to assess the pattern of recombination, which is a common evolutionary process in these viruses, often leading to the emergence of new variants with altered properties (Lima *et al.*, 2017; Xavier *et al.*, 2021; Iqbal *et al.*, 2023).

The identification of begomovirus in a *B. tabaci* individual does not confirm infection of the host plant from which it was collected, nor does it guarantee an association between the identified virus and any co-isolated betasatellite or alphasatellite. *B. tabaci* is highly polyphagous and can acquire begomoviruses, betasatellites and alphasatellites from various plants/plant species. Nevertheless, analyzing the begomoviruses, betasatellites, and alphasatellites carried by *B. tabaci* provides valuable insight into the range of these elements

present in the local plant community. In this study, complete sequences of 41 begomoviruses, 20 betasatellites, and 24 alphasatellites were determined from 41 individual insects. These virus sequences represented six species and three strains (Table 1). Alternanthera yellow vein virus (AIYVV) has been previously identified in several weed species: *Picrorhiza kurrooa* (India), *Sonchus arvensis* and *Eclipta prostrata* (Pakistan), *Alternanthera sessilis*, *Rumex nepalensis*, and *Synedrella nodiflora* (India), *Alternanthera philoxeroides*, *E. prostrata*, and *Ludwigia hyssopifolia* (China), and *E. prostrata* and *Zinnia elegans* (Vietnam) (Ha *et al.*, 2008). With the possible exception of *P. kurrooa*, a medicinal plant, these host plants are predominantly weeds. It is unlikely that the AIYVV isolates identified here from four Asia II 1 (or Asia–Pacific Islands–Australia, (ASIA) putative cryptic species) whiteflies collected on cotton originated from the cotton plants. Furthermore, previous studies examining plants where AIYVV was found rarely identified associated betasatellites, although some reported alphasatellites. This is consistent with our findings, where no betasatellite was found associated with AIYVV, except for one insect where an alphasatellite was identified (Table 1). ChiLCV, a frequently encountered monopartite begomovirus across the Indian subcontinent, has also spread into Oman. This virus poses a significant problem for chili pepper, tomato, papaya and various other crops and weeds. ChiLCV is commonly associated with distinct betasatellites, most notably Chili leaf curl betasatellite (ChLCB). The identification of ChiLCV with ChLCB in an insect collected in Sindh (Table 1) is consistent with previous reports. However, the insect was collected from cotton (*G. hirsutum*), a species not known to be infected by ChiLCV. While ChLCB has been identified once in *G. arboreum*, the significance of this finding remains unclear (Khan & Khan, 2017).

Hollyhock leaf curl virus (HoLCV) was detected in a single whitefly collected from Sindh; however, no alpha- or betasatellite was found in the same insect (Tables 1 and 7). While HoLCV typically infects *E. prostrata*, it has also been reported in *Althea rosea* (hollyhock), *Andrographis paniculata*, and *Malva parviflora* (Sattar *et al.*, 2017b). In *M. parviflora*, HoLCV was found in association with CLCuMuB (referred to as Kenaf leaf curl betasatellite in that study), which was also isolated from the whitefly in our study. Although HoLCV is known to infect plants of the *Malvaceae* family, it has not yet been shown to infect cotton (*Gossypium* spp.). Therefore, it is likely that the whitefly acquired the virus before colonizing the cotton plant.

Okra enation leaf curl virus (OELCuV) is commonly found infecting okra in India (Zaffalon *et al.*, 2012; Venkataravanappa *et al.*, 2015) and has recently been identified in okra in Sri Lanka, papaya in Iran, okra in Pakistan. While OELCuV has been identified in a single cotton plant exhibiting cotton leaf curl disease (CLCuD) symptoms, in association with Cotton leaf curl Multan virus (CLCuMuB), Saeed *et al.* demonstrated that OELCuV and CLCuMuB isolates from this cotton plant can infect *Nicotiana benthamiana* and induce CLCuD-like symptoms (Saeed *et al.*, 2018), this complex appears to be a minor contributor to CLCuD in the field. This is supported by the infrequent identification of this complex in cotton (only one plant to date) and the detection of OELCuV in a single insect from Faisalabad (Table 1).

The “Rajasthan” strain of CLCuMuV (formerly designated as Cotton leaf curl Rajasthan virus) was initially identified in northwestern India from cotton during 1994 (acc. no. AF363011; (Judith K Brown *et al.*, 2015)) but was not detected in cotton in Pakistan until recently (Zubair *et al.*, 2017; Mahmood *et al.*, 2024), despite being found infecting the weed *Digera arvensis* and tomato vegetable in Pakistan in 2005 (Shahid *et al.*, 2007). CLCuMuV-Raj was also identified (2006–2008) in exotic, presumably non-resistant *Gossypium* species maintained in Multan. CLCuMuV-Raj is not considered resistance-breaking and this likely explains its limited spread into Pakistani cotton when only resistant varieties were cultivated in Punjab, despite its presence within the country. More recently, with the breakdown of resistance in cotton, the selective pressure to maintain resistance (to CLCuKoV-Bur/CLCuMuB) has diminished, and the cultivation of susceptible varieties has likely facilitated the re-emergence of pre-resistance viruses and the establishment of non-resistance-breaking viruses such as CLCuMuV-Raj. The four isolates of *B. tabaci* (Asia II 1 or (ASIA) putative cryptic species) identified in this study were originated from whiteflies collected on cotton in Sindh (Tables 1, 6, and 7), representing the first report of CLCuMuV-Raj in this province. As CLCuD has only occurred sporadically in

Sindh, resistant cotton varieties were never widely adopted there. So, the appearance of CLCuMuV-Raj in Sindh likely resulted from the spread of *B. tabaci* Asia II 1 or (ASIA) putative cryptic species from Punjab (Masood *et al.*, 2017), with the virus itself likely originating in neighboring Rajasthan, India, assuming efficient transmission by this cryptic species (Pan *et al.*, 2018) used the "Faisalabad" strain of CLCuMuV in their transmission studies).

Table 7. Summary of viruses, alpha- and betasatellites harboured by each *B. tabaci* cryptic species in Punjab and Sindh provinces

Province (no.) [n = 41]	Species (no.)	Virus (no.)	Betasatellite (no.)	Alphasatellite (no.)
Punjab (26)	-	Total = 26	Total = 13	Total = 18
	Asia II 1 (or ASIA) (23)	CLCuKoV-Bur (18)	CLCuMuB ^{Bur} (10)	CLCuMuA (7)
				GDarSLA (1)
				none (2)
			GuLCB (1)	CLCuMuA (1)
			none (7)	CLCuMuA (2)
				GDarSLA (2)
				none (5)
	AIYVV (4)	none (4)	GDarSLA (1)	
	OELCuV (1)	none (1)	none (3)	
Asia I (3)	CLCuKoV-Bur (3)	CLCuMuB ^{Bur} (2)	AYVSGA (1)	
		none (1)	CLCuMuA (2)	
Sindh (15)	-	Total = 13	Total = 7	Total = 6
	Asia II 1 (or ASIA) (14)	CLCuKoV-Bur (7)	CLCuMuB ^{Bur} (3)	none (3)
			CLCuMuB ^{Sha} (1)	CLCuMuA (1)
			none (3)	AEA (1)
		CLCuMuV-Raj (4)	CLCuMuB ^{Sha} (2)	none (2)
			none (2)	CLCuMuA (1)
			none (2)	none (1)
	CLCuKoV-Sha (2)	none (2)	AEA (2)	
	HoLCV (1)	none (1)	none (1)	
	MEAM 1 (NAFME) (1)	ChiLCV (1)	ChiLCB (1)	AEA (1)

The Cotton leaf curl Kokhran virus Shahdadpur strain (CLCuKoV-Sha), formerly referred to as Cotton leaf curl Shahdadpur virus, was first identified in association with CLCuD in Sindh during 2004 and 2005 (Luqman Amrao *et al.*, 2010). This strain was subsequently identified in Punjab in samples from 2015 (Zubair *et al.*, 2017). While it is unclear whether CLCuKoV-Sha is resistance-breaking, its initial identification in Sindh and later detection in Punjab, well after the resistance breakdown of 2001, suggests it is not. Our results, which identify CLCuKoV-Sha in insects collected from Sindh (Tables 1 and 7), confirm the continued circulation of this virus strain in the region.

Most of the virus isolates identified in this study were the "Burewala" strain of CLCuKoV (formerly referred to as Cotton leaf curl Burewala virus; Table 1 and 5). This strain, first identified by Amrao *et al.* (2010b), was associated with CLCuD-resistant cotton varieties. Although the molecular basis of resistance breakdown remains unclear, evidence suggests that CLCuKoV-Bur overcame resistance through a mutated TrAP gene (Amrao *et al.*, 2010) implying that TrAP is the avirulence determinant recognized by the resistance

gene product(s) in resistant cotton (Amrao *et al.*, 2010; Rob W. Briddon *et al.*, 2014; Akbar *et al.*, 2016). All CLCuKoV-Bur isolates identified here possess TrAP-encoding gene mutations that truncate the protein to 35 amino acids (aa) (Table S1), a characteristic of the resistance-breaking strain. While the studies have identified CLCuKoV-Bur isolates with TrAP proteins longer than 35 aa or with full-length TrAP, these have not been associated with resistant cotton. These isolates likely represent reversions due to the decreased prevalence of resistant cotton in the field.

Alphasatellites, members of the family *Alphasatellitidae*, are self-replicating single-stranded, small, circular, DNA molecules (approximately 1000-1400 nt) that are found associated with geminiviruses and nanoviruses (family *Nanoviridae*; (Briddon *et al.*, 2018)). These satellite-like molecules depend on a helper virus for encapsidation, movement within plants, and transmission between plants. While alphasatellites are not essential for the infectivity or symptom induction of their helper viruses, they encode a highly conserved Rep protein (rolling-circle replication initiator protein). Recent studies have shown that Rep proteins from geminivirus-associated alphasatellites (subfamily *Geminialphasatellitinae*) can suppress transcriptional gene silencing in plants, suggesting a potential role in evading host defense mechanisms. With the exception of AYVSGA, all alphasatellites identified in this study belong to the subfamily *Geminialphasatellitinae* and the genus *Colecusatellite*, which are widespread in the Old World (Briddon *et al.*, 2018). The most frequently identified alphasatellite in whiteflies in this study was CLCuMuA, the most common alphasatellite associated with CLCuD-affected cotton.

A notably divergent alphasatellite was identified in this study is AYVSGV (sub-family *Geminialphasatellitinae*, genus *Ageyesisatellite*; (Briddon *et al.*, 2018)). AYVSGV was first identified in *Ageratum conyzoides* from Singapore, associated with Ageratum yellow vein virus and Ageratum yellow vein betasatellite (Saunders *et al.*, 2002), and subsequently in tomato from Oman, associated with Tomato yellow leaf curl virus and Tomato leaf curl betasatellite (Idris *et al.*, 2011). The Omani isolate was shown experimentally to reduce disease symptom severity in tomato by decreasing betasatellite levels (Idris *et al.*, 2011). The present identification of AYVSGV represents the first report of this alphasatellite species outside of Singapore and Oman. Its presence in Punjab province, Pakistan, and within an insect also harboring OELCuV, is therefore unexpected.

The study identified a total of 20 betasatellites sequences, of which eighteen were identified as CLCuMuB. CLCuMuB is known to associate various monopartite begomoviruses and cause CLCuD (Briddon *et al.*, 2001; Mansoor, Briddon, *et al.*, 2003). Given the predominance of cotton cultivation during the summer months in the study regions and the focus on insects collected from cotton fields, the identification of CLCuMuB was expected. Among the 18 CLCuMuB isolates, fifteen carried a distinctive recombinant fragment derived from ToLCB, characteristics of the “Burewala” strain of CLCuMuB. First identified by Amin *et al.* (Amin *et al.*, 2006), CLCuMuB^{Bur} emerged following a resistance breaking in cotton 2001 across Pakistan (Mansoor *et al.*, 2003a). Subsequently, with the widespread cultivation of resistant cotton (*G. hirsutum*) across Pakistan (except Sindh), CLCuMuB^{Bur} became the predominant betasatellite associated with CLCuD in symptomatic plants. The current findings align with this observation: in instances where CLCuMuB^{Bur} was detected in insects, the associated virus was consistently identified as CLCuKoV-Bur.

Three CLCuMuB isolates identified in whiteflies have a short ToLCB fragment characteristics of the “Shahdadpur” strain. Initially identified by Amrao *et al.* (Amrao *et al.*, 2010), CLCuMuB^{Sha} has since been reported in the Punjab (Zubair *et al.*, 2017). Zubair *et al.* (Zubair *et al.*, 2017) designated these isolates as the “Vehari” strain due to minor variations, this nomenclature is considered inappropriate as they do not form a distinct phylogenetic clade from CLCuMuB^{Sha} (see KX697597 and KX697602 in Figure 3C). The phylogenetic analysis in Zubair *et al.* (Zubair *et al.*, 2017) erroneously suggested a distinct lineage due to the inclusion of three partial CLCuMuB sequences from India in the alignment. These findings confirm the continued presence of CLCuMuB^{Sha} in Sindh, where it was first identified, highlighting its ongoing threat to

cotton production in the region. Notably, two of the CLCuMuB^{Sha} isolates were detected in whiteflies also harboring CLCuMuV-Raj, a begomovirus not previously reported to associate with CLCuMuB^{Sha}.

An unusual betasatellite, exhibiting high sequence similarity to two previously characterized guar-associated betasatellites, was identified in an Asia II 1 (or (ASIA) putative cryptic species) collected from cotton fields in Punjab. For this reason, the name GuLCB is proposed for this newly discovered betasatellite species. This insect also carried CLCuKoV-Bur and CLCuMuA. The presence of GuLCB, which is typically associated with guar plants, suggests that the insect may have previously fed on guar or other related plant species before transmitting viruses to cotton. Guar (*Cyamopsis tetragonoloba*; family *Fabaceae*), a legume widely cultivated across the Indian subcontinent, serves as a vegetable, fodder, green manure, and a source of valuable beans.

This study examined whiteflies mostly from *G. hirsutum* (cotton) and three of the whiteflies from *Solanum melongena* (eggplant), limiting the ability to draw conclusions about differences in insect-virus associations between the two hosts. Nevertheless, a significant finding was the identification of three *B. tabaci* (insect nos. 22, 23 and 24; Table 1) harboring CLCuKoV-Bur. Two of these whiteflies also harbored CLCuMuB, one CLCuMuA, is consistent with the earlier identification of CLCuKoV-Bur and CLCuMuB in *S. melongena*. While Ullah *et al.* (Ullah *et al.*, 2014) reported yellow mosaic symptoms on *S. melongena*, which are not typically associated with CLCuKoV-Bur and CLCuMuB infections, these viruses were shown to induce CLCuD symptoms in resistant *G. hirsutum* through back transmission using *B. tabaci*. This demonstrates the potential of *S. melongena* to serve as a reservoir host for this virus complex.

Pan *et al.* demonstrated that among four *B. tabaci* cryptic species, only Asia II 1 (or Asia–Pacific Islands–Australia, (ASIA) putative cryptic species) efficiently transmits CLCuMuV/CLCuMuB to cotton (Pan *et al.*, 2018). However, while transmitting to tobacco, Mediterranean (MED) species exhibited low transmission efficiency (~10%), resulting in symptomatic plants, compared to ~70% for Asia II 1 or (ASIA) putative cryptic species. Interestingly, all four cryptic species (Asia II 1, MED, MEAM 1, and Asia 1) transmitted CLCuMuV to plants (10-20%) that remained asymptomatic but harboured the virus at extremely low titers, particularly MED, MEAM 1 and Asia 1. Unfortunately, the authors did not conduct further investigations on these plants, which may have been infected with the virus alone, as the study solely relied on diagnostics for the virus and not the associated betasatellite. The presence of CLCuD complex components (CLCuKoV-Bur and CLCuMuB) in three Asia 1 leafhoppers (MEAM 1 and Asia 1) in Pakistan raises the question of their potential role in transmitting the disease between alternate hosts, even if they do not transmit to cotton (Table 1 and 6). Notably, while MEAM 1 individuals from Sindh, Pakistan, were found to harbor ChiLCV, this cryptic species appears to play a minor or negligible role in CLCuD spread, as suggested by Pan *et al.* (Pan *et al.*, 2018). This aligns with the findings of Ahmed *et al.*, demonstrated that Asia II 1 or Asia–Pacific Islands–Australia, (ASIA) putative cryptic species exhibits a preference for cotton and better performance on it compared to MEAM 1 (Ahmed *et al.*, 2014). These observations suggest that while Asia 1 may contribute to the CLCuD disease cycle within the broader ecosystem, their impact on cotton may be limited.

While we examined sixteen MEAM 1 whiteflies, only one whitefly tested positive for a single virus, contrasting with the findings for Asia II 1 or Asia–Pacific Islands–Australia, (ASIA) putative cryptic species, where all insects (n=37) were infected (Table 1 and 6). This discrepancy may indicate that virus and betasatellite DNA levels in MEAM 1 (or North Africa–Middle East NAFME putative species) whitefly are very low. Furthermore, it is possible that the exotic MEAM 1 might exhibit limited or no transmission of the begomoviruses prevalent in Pakistan, as observed with CLCuMuV. It is crucial to acknowledge that even cryptic *B. tabaci* species incapable of transmitting a specific virus can still ingest it. However, in such cases, the virus likely remains confined to the digestive tract, failing to be acquired (transferred into the bloodstream) and instead being either digested or excreted (Czosnek *et al.*, 2017).

Several studies have indicated a shift in the composition of CLCuD complex (Hassan *et al.*, 2017; Sattar *et al.*, 2017; Zubair *et al.*, 2017; Ahmed *et al.*, 2021). Zubair *et al.* reported the re-emergence of begomoviruses like CLCuMuV, CLCuKoV (strains other than "Burewala"), and Cotton leaf curl Alabad virus in cotton, a

phenomenon not observed since the introduction of CLCuD resistance in the 1990s (Zubair *et al.*, 2017). This finding was based on an analysis of six cotton plants collected in 2015. In contrast, Datta *et al.* (2017) observed the sole presence of CLCuMuV in cotton samples collected from Indian Punjab during the same year. Furthermore, Zaidi *et al.* 2017 examined thirty-one symptomatic cotton plants and detected bipartite begomovirus Tomato leaf curl New Delhi virus, a virus not usually associated with CLCuD, in 20 of them (Zaidi *et al.*, 2017). However, CLCuKoV-Bur was also consistently detected in all plants. Notably, the analysis of *B. tabaci* populations in this study did not reveal any corresponding shifts in their composition. These findings largely align with the CLCuD complex dynamics observed between 2005 and 2010 in these regions, with the potential exception of CLCuKoV-Bur's incursion into Sindh (Amrao *et al.*, 2010). A possible explanation for the discrepancies between this study and others might lie in the sampling period. Most of the samples in this study were collected earlier (2012-2014) compared to those in other studies (2013-2015), suggesting that the observed shifts in the CLCuD complex might have occurred after the collection of the *B. tabaci* samples analyzed here.

The results presented herein offer compelling evidence for the prevalence of recombination events among the AIYVV, ChiLCV, HoLCV, OELCuV, CLCuMuV, CLCuKoV-Bu, betasatellites, and alphasatellites. Multiple recombination events, involving diverse breakpoints, varying combinations of parental sequences, and occurrence of both inter- and intraspecies recombination events. These findings underscore the dynamic nature of begomovirus evolution, highlight the critical role of recombination in generating genetic diversity, facilitating adaptation to changing environmental conditions, and a high degree of genetic exchange among different begomovirus lineages. Additionally, recombination was detected in the isolated satellites, emphasizing the importance of this process in the evolution of these satellite molecules. In the three decades, begomoviruses and their associated DNA-satellites have proliferated substantially, primarily driven by factors such as mutation, recombination, pseudo-recombination, genetic drift, component capture, the increased population of whitefly vectors, and the expansion of host plant ranges (Nigam, 2021). Among these genetic drivers, recombination stands out as a particularly potent force for swift evolution, as evidenced by the spread of several geminivirus species in Africa (Maize streak virus and cassava mosaic virus), Europe (tomato yellow leaf curl virus), and Asia (cotton leaf curl virus) (Sanz *et al.*, 2000; Pita *et al.*, 2001; García-Andrés *et al.*, 2007; Lima *et al.*, 2017; Sattar *et al.*, 2022; Iqbal *et al.*, 2023).

This study investigated the begomoviruses, betasatellites, and alphasatellites carried by *Bemisia tabaci* collected primarily from cotton in Punjab and Sindh, Pakistan. Previous research demonstrated distinct differences between *B. tabaci* populations in these provinces, likely reflecting variations in viral diversity and the presence of different cryptic species (particularly Asia II 1 or ASIA and MEAM 1 or NAFME). These cryptic species may exhibit distinct transmission specificities, with MEAM 1 (or NAFME) potentially less efficient at transmitting native begomoviruses. This information is crucial for developing effective begomovirus control strategies and predicting the impact of shifts in *B. tabaci* cryptic species prevalence on virus distribution. For instance, the increasing prevalence of MEAM 1 (or NAFME) in Punjab and Asia II 1 (or ASIA) putative cryptic species in Sindh (Masood *et al.*, 2017; Iqbal *et al.*, 2024) might lead to a decline in CLCuD-associated viruses in Punjab and an increase in Sindh (Masood and Briddon, 2018). Conversely, the rise of MEAM 1 (or NAFME) could exacerbate problems caused by Tomato yellow leaf curl virus, efficiently transmitted by MEAM 1 and recently detected in Pakistan. This study underscores the need to assess the begomovirus transmission capabilities of other prevalent *B. tabaci* cryptic species in Pakistan, particularly those associated with CLCuD. Furthermore, investigating the transmissibility of different begomoviruses, especially CLCuD-associated ones, by various *B. tabaci* cryptic species is essential (Masood and Briddon, 2018).

The outcome of the study demonstrated that future research should focus on real-time surveillance of begomovirus diversity in *B. tabaci*, developing novel vector control methods targeting Asia II 1, breeding resistant cotton varieties against recombinant strains, creating regional epidemiological models, and exploring molecular virus-vector interactions to enhance sustainable cotton production.

Conclusions

CLCuD is a serious constraint to cotton production across Pakistan and northwestern India. There is the fear that the disease could spread to other cotton-growing regions with environmental conditions favorable for both the virus and the vector; the whitefly *B. tabaci*. This fear was recently realized with the spread of the disease into parts of China (Gao *et al.*, 2025). The results here strongly support the recent demonstration that at least one of the viruses and associated satellite causing CLCuD is transmitted by Asia II 1 (or (ASIA) putative cryptic species), one of the cryptic species making-up *B. tabaci*, and not by other cryptic species, such as MEAM 1 (or NAFME), which has a near global distribution and has been responsible for epidemics of other related viruses. Additionally, the work shows distinct differences between the viruses/satellites harbored by *B. tabaci* in Punjab and Sindh provinces of Pakistan. Additionally, this study showed the highly recombinant nature of the isolated begomoviruses and their associated DNA satellites. In contrast to the Punjab, CLCuD in Sindh has been sporadic, likely due to prevalence there of MEAM 1 (or NAFME). The results support the contention that, with respect to CLCuD, only Asia II 1 (or (ASIA) putative cryptic species) is a problem and that any spread of the disease will necessitate a transfer of both the virus complex and the vector, Asia II 1 (or ASIA) putative cryptic species. Moving forward, these results emphasize the need for targeted surveillance and management strategies focusing on Asia II 1 to prevent further CLCuD spread. Future research should prioritize developing resistant cotton varieties, exploring novel vector control methods, and leveraging genomic tools to monitor viral recombination and vector dynamics, ensuring sustainable cotton production in at-risk regions.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the Gen bank INSDC member repository. The accession numbers are provided in Table 1.

Accessions for the Begomoviruses: MH538338- MH538345, MH555070- MH555073, MH560503, MH781030, MH766871- MH766890, MH756638- MH756644, MH781029, MH781030

Accessions for the Alphasatellites: MH510291- MH510295, MH517039- MH517043, MH450227, MH464254- MH464256, MH817849, MH510279- MH510287, MH517037, MH517038, LN874303, MK167471

Accessions of Betasatellites: MH411247, MH427318- MH427328, MH427326- MH427327, MH450220- MH450226, MH411248

Authors' contribution

Conceptualization, Supervision, Project Administration, R.W.B.; and Z.I., Writing, review, editing, R.W.B., A.I., and Z.I., Investigation, M.M.; Z.I., and S.A.R.B.; Final draft editing, Z.I., and R.W.B. Funding, Z.I.; and A.S.

All authors read and approved the final manuscript.

Funding

This scientific paper is derived from a research grant funded by the Research, Development, and Innovation Authority (RDIA) - Kingdom of Saudi Arabia - with grant number (12877-KFU-2023-KFU-R-2-

1-SE-), and publication of this research was supported by the Deanship of Scientific Research (DSR), King Faisal University, Kingdom of Saudi Arabia (KFU252398).

Acknowledgements

This scientific paper is derived from a research grant funded by the Research, Development, and Innovation Authority (RDIA) - Kingdom of Saudi Arabia - with grant number (12877-KFU-2023-KFU-R-2-1-SE-), and publication of this research was supported by the Deanship of Scientific Research (DSR), King Faisal University, Kingdom of Saudi Arabia (KFU).

Note

This research constitutes a portion of MM's doctoral dissertation submitted to Higher Education Commission (HEC), Pakistan. It is the only medium in which this content has appeared. As this work has not been previously published in any other format, its dissemination aligns with the MM's university regulations. The authors express their gratitude to the staff of the National Institute for Biotechnology and Genetic Engineering for their valuable support throughout this study.

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

- Ahmed MZ, Naveed M, Noor ul Ane M, Ren SX, De Barro P, Qiu B L (2014). Host suitability comparison between the MEAM1 and Asia II 1 cryptic species of *Bemisia tabaci* in cotton-growing zones of Pakistan. *Pest Management Science* 70:1531-1537. <https://doi.org/10.1002/ps.3716>
- Ahmed N, Amin I, Zaidi SS-e-A, Rahman SU, Farooq M, Fauquet CM, Mansoor S (2021). Circular DNA enrichment sequencing reveals the viral/satellites genetic diversity associated with the third epidemic of cotton leaf curl disease. *Biology Methods and Protocols* 6(1):bpab005. <https://doi.org/10.1093/biomethods/bpab005>
- Akbar F, Iqbal Z, Briddon R W, Vazquez F, Saeed M (2016). The 35-amino acid C2 protein of Cotton leaf curl Kokhran virus, Burewala, implicated in resistance breaking in cotton, retains some activities of the full-length protein. *Virus Genes* 52:688-697. <http://dx.doi.org/10.1007/s11262-016-1357-3>
- Akhtar S, Tahir MN, Baloch GR, Javaid S, Khan A Q, Amin I, ... Mansoor S (2014). Regional changes in the sequence of cotton leaf curl Multan betasatellite. *Viruses* 6:2186-2203. <https://doi.org/10.3390/v6052186>
- Alemndri V, De Barro P, Bejerman N, Caro EBA, Dumón AD, Mattio MF, ... Truol G (2012). Species within the *Bemisia tabaci* (Hemiptera: *Aleyrodidae*) complex in soybean and bean crops in Argentina. *Journal of Economic Entomology* 105:48-53. <https://doi.org/10.1603/EC11161>
- Amin I, Mansoor S, Amrao L, Hussain M, Irum S, Zafar Y, ... Briddon RW (2006). Mobilisation into cotton and spread of a recombinant cotton leaf curl disease satellite. *Archives of Virology* 151:2055-2065. <https://doi.org/10.1007/s00705-006-0773-4>
- Amrao L, Akhter S, Tahir MN, Amin I, Briddon RW, Mansoor S (2010). Cotton leaf curl disease in Sindh province of Pakistan is associated with recombinant begomovirus components. *Virus Research* 153:161-165. <https://doi.org/10.1016/j.virusres.2010.07.003>

- Aslam MQ, Naqvi RZ, Zaidi SS-e-A, Asif M, Akhter KP, Scheffler BE, ... Mansoor S (2022). Analysis of a tetraploid cotton line Mac7 transcriptome reveals mechanisms underlying resistance against the whitefly *Bemisia tabaci*. *Gene* 820:146200. <https://doi.org/10.1016/j.gene.2022.146200>
- Briddon R W, Akbar F, Iqbal Z, Amrao L, Amin I, Saeed M, Mansoor S (2014). Effects of genetic changes to the begomovirus/betasatellite complex causing cotton leaf curl disease in South Asia post-resistance breaking. *Virus Research* 186:114-119. <http://www.sciencedirect.com/science/article/pii/S0168170213004371>
- Briddon RW, Bull SE, Amin I, Idris AM, Mansoor S, Bedford ID, ... Markham P G (2003). Diversity of DNA β , a satellite molecule associated with some monopartite begomoviruses. *Virology* 312:106-121. [https://doi.org/10.1016/S0042-6822\(03\)00200-9](https://doi.org/10.1016/S0042-6822(03)00200-9)
- Briddon RW, Bull SE, Mansoor S, Amin I, Markham PG (2002). Universal primers for the PCR-mediated amplification of DNA β : a molecule associated with some monopartite begomoviruses. *Molecular Biotechnology* 20:315-318. <https://doi.org/10.1385/MB:20:3:315>
- Briddon RW, Mansoor S, Bedford ID, Pinner MS, Saunders K, Stanley J, ... Markham PG (2001). Identification of DNA components required for induction of cotton leaf curl disease. *Virology* 285:234-243. <https://doi.org/10.1006/viro.2001.0949>
- Briddon RW, Markham PG (2000). Cotton leaf curl virus disease. *Virus Research* 71:151-159. [https://doi.org/10.1016/S0168-1702\(00\)00195-7](https://doi.org/10.1016/S0168-1702(00)00195-7)
- Briddon RW, Martin DP, Roumagnac P, Navas-Castillo J, Fiallo-Olive E, Moriones E, ... Varsani A (2018). *Alphasatellitidae*: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Archives of Virology* 163:2587-2600. 2 <https://doi.org/10.1007/s00705-018-3854-2>
- Brown JK. (2010). Phylogenetic biology of the *Bemisia tabaci* sibling species group. In Stansly PA and Naranjo S.E (Eds.), *Bemisia: Bionomics and Management of a Global Pest*. Dordrecht: Springer Netherlands. pp 31-67.
- Brown JK, Zerbini FM, Navas-Castillo J, Moriones E, Ramos-Sobrinho R, Silva JF, ... Varsani A (2015). Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Archives of Virology* 160:1593-1619. <http://dx.doi.org/10.1007/s00705-015-2398-y>
- Bull SE, Briddon RW, Markham PG (2003). Universal primers for the PCR-mediated amplification of DNA 1: a satellite-like molecule associated with begomovirus-DNA β complexes. *Molecular Biotechnology* 23:83-86. <https://doi.org/10.1385/MB:23:1:83>
- Czosnek H, Hariton-Shalev A, Sobol I, Gorovits R, Ghanim M (2017). The incredible journey of begomoviruses in their whitefly vector. *Viruses* 9:273.
- Datta S, Budhauya R, Das B, Gopalakrishnan R, Sharma S, Chatterjee S, ... Veer V (2017). Rebound of Cotton leaf curl Multan virus and its exclusive detection in cotton leaf curl disease outbreak, Punjab (India), 2015. *Scientific Reports* 7:17361. <https://doi.org/10.1038/s41598-017-17680-9>
- de Moya RS, Brown JK, Sweet AD, Walden KKO, Paredes-Montero JR, Waterhouse RM, Johnson KP (2019). Nuclear orthologs derived from whole genome sequencing indicate cryptic diversity in the *Bemisia tabaci* (Insecta: *Aleyrodidae*) complex of whiteflies. *Diversity* 11(9):151. <https://doi.org/10.3390/d11090151>
- Dinsdale A, Cook L, Riginos C, Buckley YM, Barro PJ (2010). Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: *Aleyrodoidea*: *Aleyrodidae*) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America* 103:196-208. <https://doi.org/10.1603/AN09061>
- Firdaus S, van Heusden AW, Hidayati N, Supena EDJ, Mumm R, de Vos RCH, ... Vosman B (2013). Identification and QTL mapping of whitefly resistance components in *Solanum galapagense*. *Theoretical and Applied Genetics* 126:1487-1501. <https://doi.org/10.1007/s00122-013-2067-z>
- Gao G, Zhang X, Du Y (2025). The first report of cotton leaf curl Multan virus of spinach in China. *Plant Disease* 109:725.
- García-Andrés S, Accotto G P, Navas-Castillo J, Moriones E (2007). Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. *Virology* 359:302-312. <https://doi.org/10.1016/j.virol.2006.09.030>
- Ha C, Coombs S, Revill P, Harding R, VuM, Dale J (2008). Molecular characterization of begomoviruses and DNA satellites from Vietnam: additional evidence that the New World geminiviruses were present in the Old World prior to continental separation. *Journal of General Virology* 89:312-326.

- Hadjistylli M, Roderick GK, Brown JK (2016). Global population structure of a worldwide pest and virus vector: genetic diversity and population history of the *Bemisia tabaci* sibling species group. *PLoS ONE* 11:e0165105. <https://doi.org/10.1371/journal.pone.0165105>
- Hassan I, Amin I, Mansoor S, Briddon R W (2017). Further changes in the cotton leaf curl disease complex: an indication of things to come? *Virus Genes* 53:759-761. <https://doi.org/10.1007/s11262-017-1496-1>
- Hussain A, Farooq M, Naqvi RZ, Aslam MQ, Siddiqui HA, Amin I, ... Mansoor S (2023). Whole-genome resequencing deciphers new insight into genetic diversity and signatures of resistance in cultivated cotton *Gossypium hirsutum*. *Molecular Biotechnology* 65:34-51. <https://doi.org/10.1007/s12033-022-00527-8>
- Hussain T, Tahir M, Mahmood T (1991). Cotton leaf curl virus. *Pakistan Journal of Phytopathology* 3:57-61.
- Idris AM, Shahid MS, Briddon RW, Khan AJ, Zhu JK, Brown JK (2011). An unusual alphavirus associated with monopartite begomoviruses attenuates symptoms and reduces betasatellite accumulation. *Journal of General Virology* 92:706-717.
- Iqbal Z, Masood M, Shafiq M, Briddon R W (2024). Temporal changes in the levels of virus and betasatellite DNA in *B. tabaci* feeding on CLCuD affected cotton during the growing season. *Frontiers in Microbiology* 15:1410568. <https://doi.org/10.3389/fmicb.2024.1410568>
- Iqbal Z, Shafiq M, Sattar MN, Ali I, Khurshid M, Farooq U, Munir M (2023). Genetic diversity, evolutionary dynamics, and ongoing spread of pedilanthus leaf curl virus. *Viruses* 15:2358. <https://doi.org/10.3390/v15122358>
- Khan Z A, Khan J A (2017). Characterization of a new begomovirus and betasatellite associated with chilli leaf curl disease in India. *Archives of Virology* 162:561-565. <http://dx.doi.org/10.1007/s00705-016-3096-0>
- Lima ATM, Silva JCF, Silva F N, Castillo-Urquiza GP, Silva FF, Seah YM, ... Zerbini FM (2017). The diversification of begomovirus populations is predominantly driven by mutational dynamics. *Virus Evolution* 3:vex005. <https://doi.org/10.1093/ve/vex005>
- Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, ... Ray SC (1999). Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. *Journal of Virology* 73:152-160. <https://doi.org/10.1128/jvi.73.1.152-160.1999>
- Mahmood MA, Ahmed N, Hussain A, Naqvi RZ, Amin I, Mansoor S (2024). Dominance of Cotton leaf curl Multan virus-Rajasthan strain associated with third epidemic of cotton leaf curl disease in Pakistan. *Scientific Reports* 14:13532. <https://doi.org/10.1038/s41598-024-63211-8>
- Mahmood MA, Ahmed N, Hussain S, Muntaha ST, Amin I, Mansoor S (2022). Dominance of Asia II 1 species of *Bemisia tabaci* in Pakistan and beyond. *Scientific Reports* 12:1528. <https://doi.org/10.1038/s41598-022-05612-1>
- 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 Pathology* 52:784. <https://doi.org/10.1111/j.1365-3059.2003.00893.x>
- Mansoor S, Briddon RW, Bull SE, Bedford ID, Bashir A, Hussain M, ... Markham PG (2003b). Cotton leaf curl disease is associated with multiple monopartite begomoviruses supported by single DNA β . *Archives of Virology* 148:1969-1986. <https://doi.org/10.1007/s00705-003-0149-y>
- Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, ... Murrell B (2021). RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evolution* 7:veaa087. <https://doi.org/10.1093/ve/veaa087>
- Masood M, Amin I, Hassan I, Mansoor S, Brown J K, Briddon R W (2017). Diversity and distribution of cryptic species of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) complex in Pakistan. *Journal of Economic Entomology* 110:2295-2300. <http://dx.doi.org/10.1093/jee/tox221>
- Masood M, Briddon R W (2018). Transmission of cotton leaf curl disease: answer to a long-standing question. *Virus Genes* 54:743-745. <https://doi.org/10.1007/s11262-018-1605-9>
- Nahid N, Amin I, Mansoor S, Rybicki EP, van der Walt E, Briddon RW (2008). Two dicot-infecting mastreviruses (family *Geminiviridae*) occur in Pakistan. *Archives of Virology* 153:1441-1451. <https://doi.org/10.1007/s00705-008-0133-7>
- Nigam D (2021). Genomic variation and diversification in begomovirus genome in implication to host and vector adaptation. *Plants* 10:1706. <https://doi.org/10.3390/plants10081706>
- Pan LL, Cui XY, Chen QF, Wang XW, Liu SS (2018). Cotton Leaf Curl Disease: Which Whitefly Is the Vector? *Phytopathology* 108:1172-1183. <https://doi.org/10.1094/PHYTO-01-18-0015-R>

- Paredes-Montero JR, Haq QMI, Mohamed AA, Brown JK (2021). Phylogeographic and SNPs analyses of *Bemisia tabaci* B mitotype populations reveal only two of eight haplotypes are invasive. *Biology* 10:1048. <https://www.mdpi.com/2079-7737/10/10/1048>
- Pita JS, Fondong VN, Sangre A, Otim-Nape GW, Ogwal S, Fauquet C M (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *Journal of General Virology* 82:655-665. <https://doi.org/10.1099/0022-1317-82-3-655>
- Rahman M, Hussain D, Malik TA, Zafar Y (2005). Genetics of resistance against cotton leaf curl disease in *Gossypium hirsutum*. *Plant Pathology* 54:764-772. <https://doi.org/10.1111/j.1365-3059.2005.01280.x>
- Rajagopalan PA, Naik A, Katturi P, Kurulekar M, Kankanallu RS, Anandalakshmi R (2012). Dominance of resistance-breaking cotton leaf curl Burewala virus (CLCuBuV) in northwestern India. *Archives of Virology* 157:855-868. <http://dx.doi.org/10.1007/s00705-012-1225-y>
- Saeed F, Sattar MN, Hameed U, Ilyas M, Haider MS, Hamza M, ... Amin I (2018). Infectivity of okra enation leaf curl virus and the role of its V2 protein in pathogenicity. *Virus Research* 255:90-94. <https://doi.org/10.1016/j.virusres.2018.07.007>
- Sanz AI, Fraile A, García-Arenal F, Zhou X, Robinson DJ, Khalid S, ... Harrison BD (2000). Multiple infection, recombination and genome relationships among begomovirus isolates found in cotton and other plants in Pakistan. *Journal of General Virology* 81:1839-1849. <https://doi.org/10.1099/0022-1317-81-7-1839>
- Sattar MN, Iqbal Z, Tahir MN, Ullah S (2017a). The prediction of a new CLCuD epidemic in the Old World. *Frontiers in Microbiology* 8:631. <http://journal.frontiersin.org/article/10.3389/fmicb.2017.00631>
- Sattar MN, Khurshid M, El-Beltagi HS, Iqbal Z (2022). Identification and estimation of sequence variation dynamics of Tomato leaf curl Palampur virus and betasatellite complex infecting a new weed host. *Biotechnology & Biotechnological Equipment* 36:609-619. <https://doi.org/10.1080/13102818.2022.2112911>
- Sattar MN, Qurashi F, Iqbal Z, Haider MS (2017b). Molecular characterization of Hollyhock leaf curl virus and associated DNA-satellites infecting *Malva parviflora* in Pakistan. *Canadian Journal of Plant Pathology* 39:229-234. <http://dx.doi.org/10.1080/07060661.2017.1336490>
- Saunders K, Bedford I D, Stanley J (2002). Adaptation from whitefly to leafhopper transmission of an autonomously-replicating nanovirus-like DNA component associated with ageratum yellow vein disease. *Journal of General Virology* 83:909-915. <https://doi.org/10.1099/0022-1317-83-4-907>
- Senanayake DMJB, Mandal B, Lodha S, Varma A (2007). First report of *Chilli leaf curl virus* affecting chilli in India. *Plant Pathology* 56:343. <https://doi.org/10.1111/j.1365-3059.2007.01513.x>
- Shahid MS, Mansoor S, Briddon RW (2007). Complete nucleotide sequences of cotton leaf curl Rajasthan virus and its associated DNA b molecule infecting tomato. *Archives of Virology* 152(11):2131 <https://doi.org/10.1007/s00705-007-1043-9>
- Tamura K, Stecher G, Kumar S (2021). MEGA11: Molecular evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* 38:3022-3027. <https://doi.org/10.1093/molbev/msab120>
- Ullah R, Akhtar K P, Hassan I, Saeed M, Sarwar N, Mansoor S (2014). Evidence of cotton leaf curl Burewala virus variant and its associate betasatellite causing yellow mosaic of eggplant (*Solanum melongena*) in Pakistan. *Journal of Phytopathology* 163 233-237. <https://doi.org/10.1111/jph.12268>
- Venkataravanappa V, Reddy CNL, Jalali S, Briddon R, Reddy MK (2015). Molecular identification and biological characterisation of a begomovirus associated with okra enation leaf curl disease in India. *European Journal of Plant Pathology* 141:217-235. <http://dx.doi.org/10.1007/s10658-014-0463-0>
- Xavier C A D, Godinho MT, Mar TB, Ferro CG, Sande OFL, Silva JC, ... Murilo Zerbini F (2021). Evolutionary dynamics of bipartite begomoviruses revealed by complete genome analysis. *Molecular Ecology* 30:3747-3767. <https://doi.org/10.1111/mec.15997>
- Zaffalon V, Mukherjee S, Reddy V, Thompson J, Tepfer M (2012). A survey of geminiviruses and associated satellite DNAs in the cotton-growing areas of northwestern India. *Archives of Virology* 157:483-495. <https://doi.org/10.1007/s00705-011-1201-y>
- Zaidi SSEA, Martin DP, Amin I, Farooq M, Mansoor S (2017). Tomato leaf curl New Delhi virus: a widespread bipartite begomovirus in the territory of monopartite begomoviruses. *Molecular Plant Pathology* 18:901-911. <https://doi.org/10.1111/mpp.12481>

- Zaidi SS, Naqvi RZ, Asif M, Strickler S, Shakir S, Shafiq M, ... Mansoor S (2020). Molecular insight into cotton leaf curl geminivirus disease resistance in cultivated cotton (*Gossypium hirsutum*). Plant Biotechnology Journal 18:691-706. <https://doi.org/10.1111/pbi.13236>
- Zaidi SS, Shakir S, Farooq M, Amin I, Mansoor S (2017). First report of a novel strain of tomato yellow leaf curl virus causing yellow leaf curl disease on cluster bean in Pakistan. Plant Disease 101:1071-1071. <https://doi.org/10.1094/PDIS-12-16-1859-PDN>
- Zubair M, Zaidi S, Shakir S, Amin I, Mansoor S (2017). An insight into Cotton leaf curl Multan betasatellite, the most important component of cotton leaf curl disease complex. Viruses 9:280. <https://doi.org/10.3390/v9100280>
- Zubair M, Zaidi SSA, Shakir S, Farooq M, Amin I, Scheffler JA, ... Mansoor S (2017). Multiple begomoviruses found associated with cotton leaf curl disease in Pakistan in early 1990 are back in cultivated cotton. Scientific Reports 7:680. <https://doi.org/10.1038/s41598-017-00727-2>



The journal offers free, immediate, and unrestricted access to peer-reviewed research and scholarly work. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author.



License - Articles published in *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* are Open-Access, distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) License.

© Articles by the authors; Licensee UASVM and SHST, Cluj-Napoca, Romania. The journal allows the author(s) to hold the copyright/to retain publishing rights without restriction.

Notes:

- **Material disclaimer:** The authors are fully responsible for their work and they hold sole responsibility for the articles published in the journal.
- **Maps and affiliations:** The publisher stay neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- **Responsibilities:** The editors, editorial board and publisher do not assume any responsibility for the article's contents and for the authors' views expressed in their contributions. The statements and opinions published represent the views of the authors or persons to whom they are credited. Publication of research information does not constitute a recommendation or endorsement of products involved.

Supplementary Data

Begomovirus and DNA satellites diversity in *Bemisia tabaci* from cotton growing areas of Pakistan

Zafar IQBAL^{1*}, Mariyam MASOOD², Sayyad Ali Raza BUKHARI³, Imran AMIN⁴, Adil ALSHOAIBI¹, Rob W. BRIDDON⁴

¹King Faisal University, Central Laboratories, Al-Ahsa 31982, Saudi Arabia;

zafar@kfu.edu.sa (*corresponding author); adshoaibi@kfu.edu.sa

²University of Agriculture, Center for Advanced Studies in Agriculture and Food Security, Faisalabad,

Pakistan; mariyam.masood@uaf.edu.pk

³University of Sargodha, Department of Biotechnology, Main Campus, Sargodha, Pakistan; ali.bukhari@uos.edu.pk

⁴National Institute for Biotechnology and Genetic Engineering, Agricultural Biotechnology Division, Faisalabad,

Pakistan; imranamin1@yahoo.com; rob.briddon@gmail.com

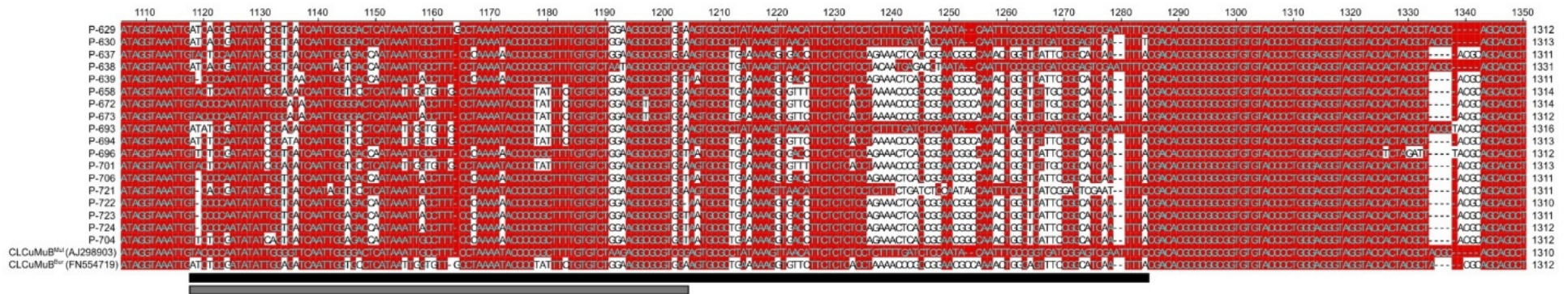
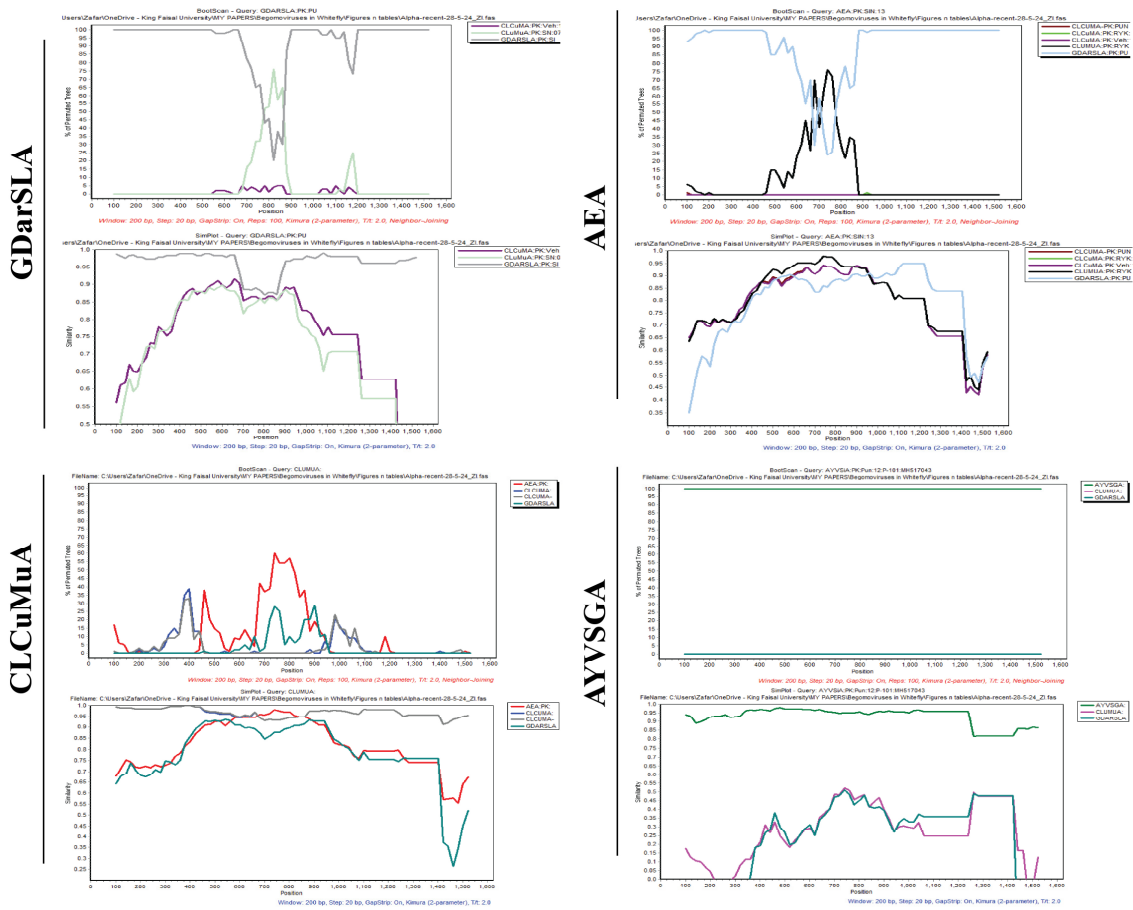


Figure S1. Alignment of the sequences of CLCuMuB isolates obtained from *B. tabaci* with those of a typical CLCuMuB^{Mul} (AJ298903) and CLCuMuB^{Burr} (FN554719) across the sequence in the satellite conserved region which has recombined between ToLCB and CLCuMuB (Amin *et al.*, 2006). The region of the recombination is indicated by the black bar. The grey bar indicates the typical recombinant sequences of CLCuMuB^{Burr}. Sequences differing from CLCuMB-[PK:Fa1:96] (AJ298903; a CLCuMB^{Mul} isolate) are highlighted as black text on a white background. The numbers above the alignment refer to positions in the alignment and the numbers along the right refer to the nucleotide coordinates of the respective sequences

Alphasatellites



Betasatellites

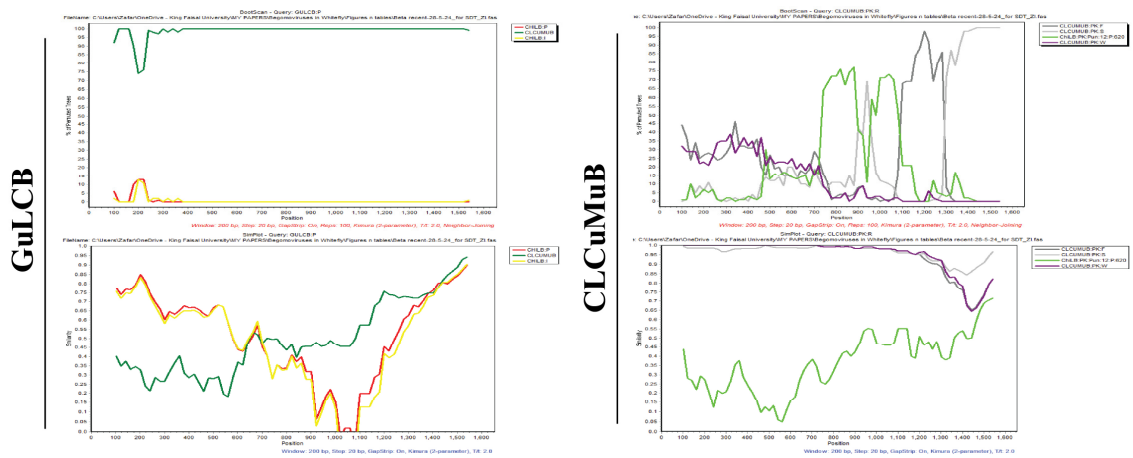


Figure S2. Recombination analysis using BootsCan and SimPlot of full-length genome sequence of isolated alpha- and betasatellites. The BootsCan plots revealed highly fragmented genomes, indicative of frequent recombination events. Additionally, SimPlot analysis demonstrated sequence similarity between the query sequence and other sequences. Nucleotide positions in the plots refer to the alignment, excluding gaps and repeat regions

Table S1. Details and features of the begomovirus clones isolated in the study

Virus*	N	Location/ Province	Plant species [†]	Biotype	Clone nos./ accession nos.	Size (bp)	Coding capacity (no. of amino acids) and position [nucleotide coordinates of start/stop codons]					
							V1	V2	C1	C2 [‡]	C3	C4
CLCuKoV-Bur	02	Faisalabad/ Punjab	<i>G. hirsutum</i>	Asia II 1	P-625/MH756638 P-626/MH756639	2759	256(292-1062)	118(132-488)	363(1505-2596)	1S	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	03	Burewala/ Punjab	<i>G. hirsutum</i>	Asia II 1	P-709/MH766879 P-710/MH766880 P-711/MH766881	2759	256(292-1062)	118(132-488)	363(1505-2596)	FS	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	02	Vehari/ Punjab	<i>G. hirsutum</i>	Asia II 1	P-712/MH766882 P-687/MH766884	2759	256(292-1062)	118(132-488)	363(1505-2596)	FS	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	01	Khanewal/ Punjab	<i>G. hirsutum</i>	Asia II 1	P-646/MH756640	2759	256(292-1062)	118(132-488)	363(1505-2596)	FS	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	03	Rahim Yar Khan/Punjab	<i>G. hirsutum</i>	Asia II 1	P-651/MH766871 P-652/MH766872 P-653/MH766873	2759	256(292-1062)	118(132-488)	363(1505-2596)	FS	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	01	Kaccha Khuh/ Punjab	<i>G. hirsutum</i>	Asia II 1	P-661/MH781029	2759	256(292-1062)	118(132-488)	363(1505-2596)	FS	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	03	Firoza/Punjab	<i>G. hirsutum</i>	Asia II 1	P-648/MH756642 P-654/MH766874 P-655/MH766875	2759	256(292-1062)	118(132-488)	363(1505-2596)	FS	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	03	Wahi Hussain/ Punjab	<i>S. melongena</i>	Asia I	P-647/MH756641 P-649/MH756643 P-650/MH756644	2759	256(292-1062)	118(132-488)	363(1505-2596)	1S	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	03	Kot Samba/ Punjab	<i>G. hirsutum</i>	Asia II 1	P-667/MH766876 P-668/MH766877 P-669/MH766878	2759	256(292-1062)	118(132-488)	363(1505-2596)	1S	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	01	Sakrand/Sindh	<i>G. hirsutum</i>	Asia II 1	P-671/MH766883	2759	256(292-1062)	118(132-488)	363(1505-2596)	FS	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	03	Dadu/Sindh	<i>G. hirsutum</i>	Asia II 1	P-682/MH766885 P-685/MH766886 P-686/MH766887	2759	256(292-1062)	118(132-488)	363(1505-2596)	FS	134(1059-1463)	146(2242-2682)
CLCuKoV-Bur	03	Jhandolo Shah/ Sindh	<i>G. hirsutum</i>	Asia II 1	P-662/MH766890 P-702/MH766888 P-703/MH766889	2759	256(292-1062)	118(132-488)	363(1505-2596)	1S	134(1059-1463)	146(2242-2682)
CLCuKoV-Sha	01	Moro/Sindh	<i>G. hirsutum</i>	Asia II 1	P-643/MH781030	2748	256(292-1062)	118(132-488)	362(1505-2593)	150(1156-1608)	134(1059-1463)	100(2137-2439)
CLCuKoV-Sha	01	Qaziabad/ Sindh	<i>G. hirsutum</i>	Asia II 1	P-627/MH538341	2748	256(292-1062)	118(132-488)	362(1505-2593)	150(1156-1608)	134(1059-1463)	100(2137-2439)
CLCuMuV-Raj	04	Sarkand/ Sindh	<i>G. hirsutum</i>	Asia II 1	P-593/MH555071 P-595/MH555070 P-596/MH555072 P-598/MH560503	2738	256(276-1046)	118(116-472)	362(1495-2583)	150(1156-1598)	134(1049-1453)	100(2127-2429)
ALYVV	04	Haroonabad/ Punjab	<i>G. hirsutum</i>	Asia II 1	P-666/MH538345 P-667/MH766876 P-668/MH766877 P-669/MH766878	2746	256(313-1083)	115(153-500)	361(1532-2617)	134(1125-1629)	134(1080-1484)	96(2170-2460)
OELCuV	1	Faisalabad/ Punjab	<i>G. hirsutum</i>	Asia II 1	P-107/MH538338	2741	256(281-1051)	115(121-468)	362(1500-2588)	150(1151-1603)	134(1054-1458)	102(2132-2440)

HoLCV	1	Rashidabad/ Sindh	<i>G. hirsutum</i>	Asia II 1	P-566/MH538339	2742	256(292-1062)	115(132-479)	361(1511-2598)	150(1162-1614)	134(1065-1465)	100(2145-2447)
ChiLCV-PK	1	Tando Allahyar/ Sindh	<i>G. hirsutum</i>	MEAM-1	P-622/MH538340	2756	256(308-1078)	118(148-504)	361(1527-2612)	134(1220-1624)	134(1079-1479)	97(2162-2455)

*Viruses are denoted as *Cotton leaf curl Khokran virus*-Burewala strain (CLCuKoV-Bur), *Cotton leaf curl Multan virus*-Rajasthan strain (CLCuMuV-Raj), *Cotton leaf curl Kokhran virus* – Shadadpur strain (CLCuKoV-Sha), *Alternanthera yellow vein virus* (AYVV), *Okra enation leaf curl virus* (OELCuV), *Chilli leaf curl virus* (ChiLCV), *Hollyhock leaf curl virus* (HoLCV)

[§]Plant species from which the insect was collected

£ The mutations of the TrAP-encoding gene of CLCuKoV-Bur are identified as first identified by Amrao et al. (2010). These mutations are indicated as either having one in-frame stop codon (1S; leaving a potential product of 35 amino acids ahead of the stop codon and an in-frame methionine initiation codon 76 amino acids downstream of the stop codon, only 26 amino acids upstream of the usual *trap* termination codon), two in-frame stop codons (2S; the first equivalent to that for the 1S mutation and the other just 3 amino acids further downstream, leaving the same 35 amino acids ahead of the stop codon and an in-frame methionine initiation codon 26 amino acids upstream of the usual *trap* termination codon) or a frame-shift mutation (truncating the product to 31 amino acids of which 23 are derived from the *trap* gene)

Table S2. Features of the alphasatellite and betasatellite clones isolated in the study

Satellite*	N	Location/ Province	Clone number/ waccession no.	Size (bp)	Coding sequence (nucleotide coordinates of start/stop codons/no. of amino acids)	
					Rep	βC1
CLCuMuA	01	Faisalabad/ Punjab	P-681/MH510280	1368	77-1024/315	-
CLCuMuA	01	Dadu/Sindh	P-691/MH510282	1365	77-1024/315	-
CLCuMuA	01	Burewala/ Punjab	P-680/MH510279	1375	77-1024/315	-
CLCuMuA	01	Vehari/ Punjab	P-690/MH510281	1375	77-1024/315	-
CLCuMuA	03	Rahim Yar Khan/ Punjab	P-676/MH450227 P-677/MH464254 P-678/MH464255	1374	77-1024/315	-
CLCuMuA	01	Kot Samba/ Punjab	P-679/MH464256	1374	77-1024/315	-
CLCuMuA	01	Firoza/ Punjab	P-674/MH510287	1374	77-1024/315	-
CLCuMuA	02	Wahi Hussain/ Punjab	P-656/MH510285 P-659/MH510286	1374	77-1024/315	-
CLCuMuA	01	Khanewal/ Punjab	P-644/MH510284	1374	77-1024/315	-
CLCuMuA	01	Kacha khuh/ Punjab	P-634/MH510283	1374	77-1024/315	-
CLCuMuA	01	Sarkand/ Sindh	P-611/MH517042	1374	77-1024/315	-
AEA	01	Tando Allah yar/Sindh	P-577/MH510291	1365	82-1029/315	-
AEA	01	Qaziabad/Sindh	P-631/MH517040	1362	82-1029/315	-
AEA	01	Moro/ Sindh	P-633/MH517041	1362	82-1029/315	-
AEA	01	Dadu/ Sindh	P-610/ MK167471	1165	defective	-
GDarSLA	01	Faisalabad/ Punjab	P-108/LN874303	1373	70-1017/315	-
GDarSLA	01	Burewala/ Punjab	P-699/MH517037	1373	70-1017/315	-
GDarSLA	01	Kot Samba/Punjab	P-609/MH817849	1373	70-1017/315	-
GDarSLA	01	Vehari/Punjab	P-700/MH517038	1373	70-1017/315	-
GDarSLA	01	Haroonabad/ Punjab	P-720/MH517039	1373	70-1017/315	-
AYVSGA	01	Faisalabad/ Punjab	P-101/MH517043	1379	57-944/295	-
CLCuMuB	02	Sarkand/Sindh	P-629/MH427318 P-630/MH427319	1353	-	195-551/118
CLCuMuB	01	Sarkand/ Sindh	P-723/MH450225	1353	-	195-551/118
CLCuMuB	01	Jhandoloo shah/ Sindh	P-724/MH450226	1353	-	195-551/118
CLCuMuB	02	Dadu/Sindh	P-638/MH427320 P-672/MH427324	1353	-	195-551/118
CLCuMuB	01	Kot Samba/Punjab	P-701/MH450220	1351	-	195-551/118

CLCuMuB	01	Vehari/Punjab	P-706/MH450222	1350	-	195-551/118
CLCuMuB	02	Wahi Hussain/ Punjab	P-696/MH427328 P-639/MH427322	1351	-	195-551/118
CLCuMuB	02	Firoza/ Punjab	P-673/MH427325 P-704/MH450221	1351	-	195-551/118
CLCuMuB	01	Khanewal/ Punjab	P-658/MH427323	1350	-	195-551/118
CLCuMuB	02	Rahim Yar khan/ Punjab	P-693/MH427326 P-694/MH427327	1350	-	195-551/118
CLCuMuB	01	Faisalabad/ Punjab	P-637/MH427321	1350	-	195-551/118
CLCuMuB	02	Burewala/ Punjab	P-721/MH450223 P-722/MH450224	1350	-	195-551/118
ChiLCB	01	Tando Allah yar/Sindh	P-620/MH411247	1358	-	200-562/200
GuLCB	01	Kacha khuh/ Punjab	P-692/MH411248	1356	-	187-570/127