# DEEP SURVEILLANCE OF COTTON-INFECTING BEGOMOVIRUSES BY TARGET ENRICHMENT AND HIGH-THROUGHPUT SEQUENCING

**Cory Keith**
**School of Plant Sciences, University of Arizona**
**Tucson, AZ**
**Javed Iqbal**
**School of Plant Sciences, University of Arizona**
**Tucson, AZ**
**Cecilia Villegas**
**Texas A&M University**
**Weslaco, TX**
**Olufemi Alabi**
**Department of Plant Pathology & Microbiology, Texas A&M AgriLife Research Center**
**Weslaco, TX**
**Judith K. Brown**
**School of Plant Sciences, University of Arizona**
**Tucson, AZ**

## Abstract

Characterization of the intra-host community of geminiviruses infecting cotton in Pakistan can shed insights into the evolution of the CLCuD complex as it responds to widespread planting of genetically uniform resistant cotton genotypes. In this study, a targeted enrichment sequencing technique, Capture-Seq, was explored to determine if the approach would yield sufficient sequence data to accurately characterize the CLCuD-begomovirome in individual cotton plants. The analysis used a sentinel plot strategy and four cotton genotypes reported to be differentially resistant to at least two of the known 'core species' of CLCuD-begomoviruses, CLCuMV and CLCuKoV-Bur, the causal begomoviruses associated with epidemics in Pakistan and India. Analyses of virome enriched sequencing data to reconstruct the intra-host communities are expected to result in new insights into begomovirus populations shifts, temporally and spatially among their host plants. Preliminary evidence suggests that begomoviral populations in one location, compared to the other six, may be hyper-responsive to MAC07 selections, based on the detection of CLCuMV and CLCuM² reads there. Further, identification of species and strains, other than those already shown to be predominant in previous studies, has revealed a surprising abundance of mixed infections in CLCuD-affected cotton plants. Detailed population analysis of this 'enriched' sequence database will help unravel how different cotton host genotypes may drive the evolution of specific loci within a viral species, and/or select for species or strains over others. The CaptureSeq approach shows great promise as a robust, rapid surveillance early warning tool to alert growers and breeders about potential future CLCuD epidemics by capturing diversification patterns associated with impending resistance-breaking strains and species inadvertently triggered by the widespread planting of genetically uniform CLCuD-resistant genotypes.

## Introduction

In agroecological settings, plant virus populations are characterized by large sizes that undergo rapid evolution by way of mutation and recombination (García-Arenal et al. 2001; García-Arenal et al. 2003). These large populations contribute to a fitness landscape within a region, where viral variants can contribute to reproductive fitness differentials between plant genotypes (Fragata et al. 2019). The large size of viral populations within a single host (Malpica et al. 2002) and rapid evolution (Malpica et al. 2002; García-Arenal et al. 2003; Saleem et al. 2016) can lead to the occurrence of many viral variants within a single plant with the capacity to adapt rapidly to the host genotype (Yarwood 1979; García-Arenal et al. 2003), thereby contributing to a fitness landscape within the host. Further, mixed viral infections can have antagonistic or synergistic effects on the disease outcome in a host plant. In the scenario of closely-related viral species and strains, mixed infections can increase the likelihood of recombination between genome segments, creating new species, and potentially contributing to new biological properties of the virus (Harrison 2002; García-Arenal and McDonald 2003). The viral community in a plant host, hereafter referred to as the 'virome', can make disease management challenging because multiple factors can lead to resistance breaking.

In an agroecological context, widespread planting of crop genotypes give a uniform genetic background for viruses to quickly respond through increased adaptation to their host genotypes (Alexander et al. 2014). When a resistant cultivar

is released and widely grown, viral populations are often found to be capable of overcoming host-encoded resistance genes in a relatively short evolutionary time, a well-known scenario associated with the cotton leaf curl disease (CLCuD) complex (Mansoor et al. 2003a). The CLCuD complex comprises species and strains of a well-characterized 'core' group of begomoviruses (Family: *Geminiviridae*) found to be associated with cotton leaf curl epidemics in Pakistan and India since ~1990's. The first clues of the impending outbreaks were evident in the mid-1970's when high-yielding, whitefly-susceptible cotton cultivars were introduced into production areas of Pakistan.

Geminiviruses are small, single-stranded DNA (ssDNA) viruses that utilize a rolling circle mechanism for replication. The ssDNA genome of approximately 2.8 kilo base pairs (kbp) in size are encapsidated in twinned, icosahedral particle of 32 x 20 nm in size. Monopartite begomoviruses such as those comprising the core CLCuD complex are associated with one or more alpha- and beta-satellite molecules. Betasatellites encode a suppressor of plant host gene silencing, and early after their discovery but before functional studies had revealed their role in suppressing host plant innate immunity, betasatellites were recognized for harboring 'symptom determinants' (Briddon et al. 2001; Mansoor et al. 2003b; Zubair et al. 2017b). Characteristic symptoms of leaf curl disease are leaf curling, vein thickening, and the development of leaf-like enations on the underside of leaves, collectively, contributing to the overall decline in plant productivity, and resulting in reduced boll set and fiber quality (Singh et al. 2013). In Pakistan, CLCuD-begomoviruses are transmitted by the whitefly *Bemisia tabaci* (Genn.) Asia II-1 mitotype and other more narrowly distributed *B. tabaci* mitotypes (Paredes-Montero et al. 2019). The propensity of *B. tabaci* to develop resistance to pesticides has confounded disease management approaches that rely on chemical and/or biological control.

Since the early 1990's, Pakistan and India have experienced two major CLCuD epidemics (Datta et al. 2017; Hassan et al. 2017). The first serious CLCuD outbreak was reported in 1988 after the cotton variety S12 was widely-planted, which was highly susceptible to the whitefly vector and the endemic begomoviruses (Farooq et al. 2011; Paredes-Montero et al. 2019). Between 1992-1997 an estimated $5 billion in losses was attributed to the begomovirus species Cotton leaf curl Multan virus (CLCuMV) (Farooq et al. 2011). Shortly after, CLCuMV was identified in vegetable and/or ornamental hosts in China (Cai et al. 2010), the Philippines (Saleem et al. 2016), and Thailand (Farooq et al. 2021); thought to be imported in ornamental hibiscus plants (She et al. 2017).

Resistant germplasm was identified through breeding programs in Pakistan and new resistant varieties were developed by 1998 (Rahman et al. 2002). This accomplishment restored cotton yields to pre-epidemic levels, but in 2001 a previously unknown virus, discovered to be a recombinant between CLCuMV and Cotton leaf curl Kohkran virus (CLCuKoV), broke the resistance. Studies have suggested the resistance was conferred through the evolution of a truncated C2 protein that retained essential functions, nonetheless (Akbar et al. 2016). The recombinant strain, referred to as CLCuKoV-Burewala (CLCuKoV-Bur), was responsible for the second epidemic in Pakistan, and was identified in leaf curl-diseased cotton in India at about the same time (Mahmood et al. 2003; Mansoor et al. 2003a; Datta et al. 2017). In approximately 2017, the predominant leaf curl species was identified as a previously unknown strain of CLCuMV, which together with several species have been associated with leaf curl outbreaks in Pakistan and India (Sattar et al. 2013; Datta et al. 2017; Sattar et al. 2017; Zubair et al. 2017a; Farooq et al. 2021). The Indian subcontinent is considered to be the center of diversity of CLCuD-begomoviruses (Zubair et al. 2017a; Farooq et al. 2021), and of at least two cryptic species of *B. tabaci*, Asia I and Asia II (see references in Paredes et al., 2019).

The high rates of mutation intrinsic to begomoviruses (Saleem et al. 2016; Lima et al. 2017), their propensity for recombination (Lima et al. 2017), and the history of CLCuD in Pakistan has necessitated the development of a deep surveillance tool with a capacity to explore the intra-host 'virome' of the CLCuD complex and inform researchers and cotton breeders of viral population and species shifts through time, especially in the context of the continuous process of selecting/developing new resistant cotton varieties to combat shifting viral genotype compositions. To accomplish this goal, targeted enrichment high-throughput sequencing, or Capture-Seq was applied to evaluate the potential of this surveillance approach for the CLCuD in Pakistan. A pilot study was undertaken using a collection of 126 cotton samples from sentinel plots located in seven different study sites, representing four differentially susceptible/resistant cotton genotypes. The 'virome' of each plant was subjected to Capture-Seq analysis following genome assembly. Extensive coverage of reads was achieved for the two best known strains, CLCuMV and Cotton leaf curl Multan betasatellite (CLCuM$^2$ ). In addition, the approach revealed the presence in cotton plants of other minor strains of the CLCuD complex, and begomoviruses and other geminiviruses previously known and/or unidentified in cotton. The results have demonstrated that Capture-Seq offers an effective approach for obtaining large virome datasets (genomes and satellites) and thereby deep surveillance of species and variants existing within a single cotton plant. With the continued annual sampling of cotton, it will become possible to provide advanced temporal and spatial information

about the distributions of predominant and minor CLCuD-begomoviral species and strains, and potentially, early-detection of potential emergent cotton-infecting geminiviruses in the cotton crop and alternative host plant species.

## Methods

Leaf samples were collected from cotton plants during 2019 and 2020 from seven regions of Pakistan located primarily in the Punjab Province. The four varieties of cotton, CIM-496, MNH-886, S-12, and MAC-07, are reported to harbor differential susceptibility to CLCuMV or CLCuKoV-Bur, the two species predominant in the field during the first disease epidemics. Samples were randomly selected from each field, placed in a cold, sealed container, transported to the lab, and transferred to tubes containing 100% glycerol. The samples were shipped to The University of Arizona (current APHIS-PPQ permit). Samples were processed 24 at a time, after removing the glycerol. Total DNA was isolated from 80mg of leaf tissue using the CTAB method (Doyle and Doyle 1987). The DNA quality was assessed by 1% agarose gel electrophoresis (TAE buffer, pH 8.0). The DNA was quantified by Qubit™ Fluorometric Quantification (ThermoFisher, Waltham, MA, USA) according to the manufacturer's protocol for Qubit™ 1X dsDNA HS Assay Kits (Catalog: Q33230). Samples showing evidence of degradation were subjected to PCR amplification using RedTaq (Sigma Aldrich, St. Louis, MO, USA) and the cotton specific primers, SAD1 (Yang et al. 2005). Primers previously developed to PCR-amplify a broad range of CLCuD-begomovirus helper components and associated betasatellites were used to screen the samples for evidence of viral infection (Brown et al., 2017; lab protocols, unpublished). A subset of amplicons was cloned and subjected to Sanger sequencing to identify the predominant species present. All samples were prepared for Capture-Seq to evaluate samples exhibiting different quality and quantity of DNA. The CLCuD-begomovirus curated database from the Brown Lab at the University of Arizona was sent to Rapid Genomics to design biotinylated probes to specifically enrich for pull down begomoviral genome and associated betasatellite sequences from total DNA isolated from cotton leaf samples. The sequences included 536 full length helper satellites and 977 full length betasatellites, comprising 13 individual species. Quality control of the DNA isolation provided 126 samples that met the standards for High Throughput Sequencing (HTS). The sample DNA was transferred to a microtiter plate, with positive experimental controls consisting of a mixture of selected cloned, CLCuD-begomoviral genomes at ratios of 1:1:1:1 and 6:4:2:1, respectively, a virus-free cotton control, and a no-template water control. The 126 samples were sent to Rapid Genomics for Capture-Seq HTS. Results were provided as paired-end reads, trimmed and quality filtered. A subset of samples were selected to evaluate the quality of the raw data using FastQC (Andrews 2010). The viral sequence data sets were uploaded to the High Performance Computing (HPC) cluster (University of Arizona) and the raw reads from each sample were mapped to a cotton reference index, using bowtie2 (Langmead and Salzberg 2012). Unused reads were mapped to a reference index of the CLCuMV and CLCuM² satellite variants found to be prevalent through PCR diagnostics. The unmapped reads were used to *de novo* assemble viral contigs using SPAdes (Nurk et al. 2013). The *de novo* contigs were filtered by size (150 bp) and 50x coverage. The filtered contigs were subjected to BLASTn (Madden 2013) using the viral nucleotide database to assign species or strains of viruses that were not mapped to the CLCuMV or CLCuM² references.

## Results and Discussion

Field samples were subjected to PCR amplification using CLCuD-begomovirus-specific primers. The results indicated 109 of 126 samples were positive for CLCuD-begomovirus presence. All 126 samples were prepared for sequencing, based on the observed leaf curl-like symptoms in plants from which leaf samples were collected. Those samples found negative by PCR amplification were primarily MAC-07 selections undergoing selection in breeding programs in Pakistan for resistance to CLCuMV and CLCuKoV-Bur. A subset of the Sanger DNA sequenced helper virus amplicons shared the highest nucleotide identity with a CLCuMV field isolate collected in 2018, whereas isolates subjected BLASTn analysis against the betasatellite database shared the highest nucleotide identity with a CLCuM² variant. These sequences were downloaded from GenBank and were used to build indices subsequently used as references for read mapping and genome reconstruction.

The summary statistics on data and data usage are shown in Table 1. The average number of reads in the DNA samples were 2,806,424 with over 40% of reads being used for assembling genomes of the CLCuMV and CLCuM² variants within samples. Approximately 16% of the reads were used to assemble the genome of distinct variants or species found in each individual cotton plant. A low percentage of reads from the cotton host plant genome were present among reads, however, the number of reads was dramatically lower than the number recovered from HTS sequencing methods that do not use an initial enrichment step, e.g. characteristically >95% of reads are of the plant host (Idris et al. 2014; Rodríguez-Negrete et al. 2019). Thus, this pilot study has yielded deep sequencing coverage from individual

cotton plants that will facilitate detection and characterization of rare variants within population of CLCuMV and CLCuM² species, and other variants, collectively, revealing a complex scenario that involves the presence of far more geminiviral genomes infecting a single plant than expected.

**Table 1.** Descriptive statistics of sequencing depth and data usage for the 126 cotton samples subjected to Capture-Seq. Coverage refers to the depth of sequencing for each component listed. Unused Reads denote reads used for assembling all other species or variants.

| | Total Reads | Cotton % reads | CLCuMuV % reads | CLCuMu² % reads | Unused % reads | CLCuMuV Coverage | CLCuMu² Coverage |
|---|---|---|---|---|---|---|---|
| Min | 1130 | 1 | 0.00 | 0.00 | 4 | 2 | 9 |
| Max | 28770940 | 93 | 61 | 81 | 89 | 462059 | 2531165 |
| Median | 925443 | 21 | 9 | 38 | 11 | 2471 | 3467 |
| Average | 2806424 | 11 | 11 | 33 | 16 | 21771 | 155793 |

The number and sum of total reads per cotton plant samples collected in 2019 and 2020 are shown in Figure 1. Samples 1 – 53, which were collected during 2019 field season yielded a lower number of reads than samples 54-126 collected in 2020. The lower number of reads for the 2019 samples is attributable to poor leaf tissue quality resulting from extremely high air temperature at the time of sample collection. Despite this unanticipated impediment sufficient coverage using the CaptureSeq method permitted assembly of full-length helper virus genomes, and both betasatellites (² ) and alphasatellites (*data not shown*). Samples 12-16, 32-34, 43-46, 52-54, 64-67, 76-78, 88-90, 100-102, and 112-115 represent a MAC07 genotype selection, which has been shown to exhibit some extent of field resistance to CLCuMV and CLCuKoV-Burewala. The mechanism of resistance associated with MAC07 lines has been attributed to a reduction in betasatellite accumulation (Zaidi et al. 2020), making the lower number of overall reads for MAC07 plants compared to the other three cotton genotypes consistent with a resistance response based on reduced virus load. By comparison, the MAC07 samples 76-78 yielded a relatively larger number of reads than most other MAC07 collections e.g., > 1 million reads, with more than 60% of reads mapping to either the CLCuMV helper genome or its' associated CLCuM² satellite component. The latter samples, collected from MAC07 breeding lines in one of the seven study sites might suggest the potential for an impending population shift in this region, over the others. Interestingly, within this same region the number of reads belonging to the CLCuD complex associated with the other cotton varieties were overall lower than was observed in other study sites in the Punjab region.

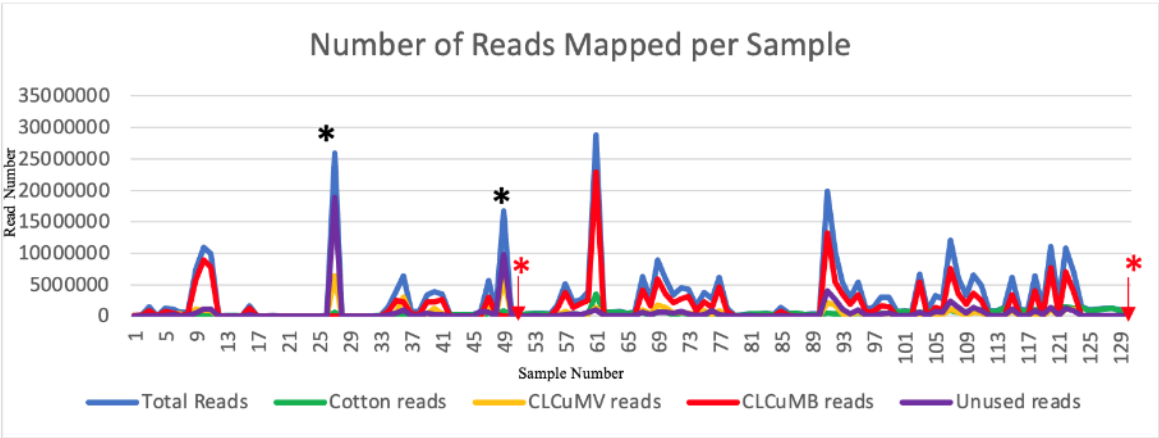


Figure 1. Number of sequencing reads mapped per sample. Total reads are shown in blue, cotton reads are shown in green, CLCuMV reads are in orange, CLCuM² red, and the unused reads after mapping are indicated by purple. The black asterisks indicate positive controls and red asterisks, negative. Sample 53 indicates the virus free control and sample 130, the no template control.

Full length genomes of the predominantly detected CLCuMV and CLCuM² satellite sets. The full-length helper viral genomes harbored varied numbers of polymorphic nucleotides (loci) at varying were assembled from samples

enriched using the CaptureSeq approach. Of the 126 samples, 92 full-length CLCuMV genomes and 114 full length CLCuM² satellite components were assembled from this pilot study sequence data sets. The full-length helper viral genomes harbored varied numbers of polymorphic nucleotides (loci) at varying frequencies among the genomic variants. Such signals are expected to contribute to an improved understanding of the evolutionary pathways of CLCuD-populations, and potentially, more closely estimate the true diversity of viral populations associated with an individual plant host compared to other previously deployed approaches.

The large number of genomes assembled (partial or full-length) belong to multiple genera of the Geminiviridae, providing evidence of the robustness of the CaptureSeq approach for detecting multiple target species and genera, as well as previously unknown species and strains that infect cotton in Pakistan. Sample 92 represents a suite of isolates for which the 'unused reads' (after mapping), indicated by the purple line, are present in greater abundance than targeted CLCuMV, indicated by the orange line. This trend was also observed for other cotton samples/cultivars analyzed here, indicating that cotton or other cultivated or uncultivated hosts in the selected study sites harbor mixed geminivirus infections. Several identified species belong to the 'core CLCuD species' group corroborates recent reports of the rebound of a number of other species of CLCuD in cotton (Zubair et al. 2017a). The results presented here are consistent with the hypothesis that although the 'core cotton leaf curl strains and species' shift in predominance over time, it appears they prevail in the environment even when they are not the predominant strain or species in an outbreak. This scenario is can likely lead to epidemics resulting from the emergence of resistance-breaking recombinant strains (Mansoor et al. 2003a) as the CLCuKoV-Bur strain was a recombinant presumably from mixed infection of CLCuMV and CLCuKoV. In many samples there was also evidence of geminiviruses not belonging to the 'core' CLCuD complex, and selected species that were detected frequently in cotton samples analyzed in this study are listed in Table 2, revealing numerous examples species that were detectable beyond those anticipated based on probe design. Also, pairwise nucleotide identities of the latter isolates, compared to their closest relatives, was as low as 71%, indicating that the probes can detect divergent geminivirus sequences, as well as the intentionally targeted strains and species.

**Table 2**. List of non-prevalent CLCuD-begomovirus species detected in cotton samples from the unused reads.

| Species Name | acronym | Genus |
|---|---|---|
| Chickpea chlorotic dwarf virus | CpClDV | Mastrevirus |
| Okra enation leaf curl virus | OELCuV | Begomovirus |
| Chili leaf curl virus | ChiLCV | Begomovirus |
| Pedilanthus leaf curl virus | PeLCV | Begomovirus |
| Hollyhock leaf curl virus | HoLCV | Begomovirus |
| Tomato leaf curl New Delhi virus | ToLCNDV-A | Begomovirus |
| Bhendi yellow vein mosaic virus | BYVMV | Begomovirus |

## References

Akbar F, Iqbal Z, Briddon RW, 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 . doi: 10.1007/s11262-016-1357-3

Alexander HM, Mauck KE, Whitfield AE, Garrett KA, Malmstrom CM (2014) Plant-virus interactions and the agro-ecological interface. Eur J Plant Pathol 138:529–547

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data

Briddon RW, Mansoor S, Bedford ID, Pinner MS, Saunders K, Stanley J, Zafar Y, Malik KA, Markham PG (2001) Identification of DNA Components Required for Induction of Cotton Leaf Curl Disease. Virology 285:234–243 . doi: 10.1006/VIRO.2001.0949

Cai JH, Xie K, Lin L, Qin BX, Chen BS, Meng JR, Liu YL (2010) Cotton leaf curl Multan virus newly reported to be associated with cotton leaf curl disease in China. Plant Pathol 59:794–795

Datta S, Budhauliya R, Das B, Gopalakrishnan R, Sharma S, Chatterjee S, Vanlalhmuaka, Raju PS, Veer V (2017) Rebound of Cotton leaf curl Multan virus and its exclusive detection in cotton leaf curl disease outbreak, Punjab (India), 2015. Sci Rep 7:17361 . doi: 10.1038/s41598-017-17680-9

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue

Farooq A, Farooq J, Mahmood A, Shakeel A, Rehman KA, Batool A, Riaz M, Shahid MTH, Mehboob S (2011) An overview of cotton leaf curl virus disease (CLCuD) a serious threat to cotton productivity. Aust J Crop Sci 5:1823–1831

Farooq T, Umar M, She X, Tang Y, He Z (2021) Molecular Phylogenetics and Evolutionary Analysis of a Highly Recombinant Begomovirus, Cotton Leaf Curl Multan Virus and Associated Satellites. Virus Evol

Fragata I, Blanckaert A, Dias Louro MA, Liberles DA, Bank C (2019) Evolution in the light of fitness landscape theory. Trends Ecol Evol 34:69–82 . doi: https://doi.org/10.1016/j.tree.2018.10.009

García-Arenal F, Fraile A, Malpica JM (2003) Variation and evolution of plant virus populations. Int Microbiol 6:225–232

García-Arenal F, Fraile A, Malpica JM (2001) Variability and genetic structure of plant virus populations. Annu Rev Phytopathol 39:157–186

García-Arenal F, McDonald BA (2003) An analysis of the durability of resistance to plant viruses. Phytopathology 93:941–952

Harrison BD (2002) Virus variation in relation to resistance-breaking in plants. Euphytica 124:181–192

Hassan I, Amin I, Mansoor S, Briddon RW (2017) Further changes in the cotton leaf curl disease complex: an indication of things to come? Virus Genes 53:759–761

Idris A, Al-Saleh M, Piatek MJ, Al-Shahwan I, Ali S, Brown JK (2014) Viral Metagenomics: Analysis of Begomoviruses by Illumina High-Throughput Sequencing. Viruses 6

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357

Lima ATM, Silva JCF, Silva FN, Castillo-Urquiza GP, Silva FF, Seah YM, Mizubuti ESG, Duffy S, Zerbini FM (2017) The diversification of begomovirus populations is predominantly driven by mutational dynamics. Virus Evol 3:vex005 . doi: 10.1093/ve/vex005

Madden T (2013) The BLAST sequence analysis tool. NCBI Handb 2:425–436

Mahmood T, Arshad M, Gill MI, Mahmood HT, Tahir M, Hussain S (2003) Burewala strain of cotton leaf curl virus: a threat to CLCuV cotton resistant varieties. Asian J Plant Sci

Malpica JM, Fraile A, Moreno I, Obies CI, Drake JW, García-Arenal F (2002) The rate and character of spontaneous mutation in an RNA virus. Genetics 162:1505–1511

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

Mansoor S, Briddon RW, Bull SE, Bedford ID, Bashir A, Hussain M, Saeed M, Zafar Y, Malik KA, Fauquet C (2003b) Cotton leaf curl disease is associated with multiple monopartite begomoviruses supported by single DNA $^{2}$. Arch Virol 148:1969–1986

Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, Prjibelsky A, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, McLean J, Lasken R, Clingenpeel SR, Woyke T, Tesler G, Alekseyev MA, Pevzner

PA (2013) Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads BT - Research in Computational Molecular Biology. In: Deng M, Jiang R, Sun F, Zhang X (eds). Springer Berlin Heidelberg, Berlin, Heidelberg, pp 158–170

Paredes-Montero JR, Hameed U, Zia-Ur-Rehman M, Rasool G, Haider MS, Herrmann H-W, Brown JK (2019) Demographic Expansion of the Predominant Bemisia tabaci (Gennadius) (Hemiptera: Aleyrodidae) Mitotypes Associated With the Cotton Leaf Curl Virus Epidemic in Pakistan. Ann Entomol Soc Am 112:265–280 . doi: 10.1093/aesa/saz002

Rahman M, Hussain D, Zafar Y (2002) Estimation of genetic divergence among elite cotton cultivars–genotypes by DNA fingerprinting technology. Crop Sci 42:2137–2144

Rodríguez-Negrete EA, Morales-Aguilar JJ, Domínguez-Duran G, Torres-Devora G, Camacho-Beltrán E, Leyva-López NE, Voloudakis AE, Bejarano ER, Méndez-Lozano J (2019) High-Throughput Sequencing Reveals Differential Begomovirus Species Diversity in Non-Cultivated Plants in Northern-Pacific Mexico. Viruses 11

Saleem H, Nahid N, Shakir S, Ijaz S, Murtaza G, Khan AA, Mubin M, Nawaz-Ul-Rehman MS (2016) Diversity, Mutation and Recombination Analysis of Cotton Leaf Curl Geminiviruses. PLoS One 11:e0151161–e0151161 . doi: 10.1371/journal.pone.0151161

Sattar MN, Iqbal Z, Tahir MN, Ullah S (2017) The prediction of a new CLCuD epidemic in the Old World. Front Microbiol 8:631

Sattar MN, Kvarnheden A, Saeed M, Briddon RW (2013) Cotton leaf curl disease–an emerging threat to cotton production worldwide. J Gen Virol 94:695–710

She XM, Tang YF, He ZF, Lan GB (2017) Molecular Characterization of Cotton leaf curl Multan virus and its associated betasatellite infecting Hibiscus rosa-sinensis in the Philippines. J Plant Pathol 765–768

Singh D, Gill JS, Gumber RK, Singh R, Singh S (2013) Yield and fibre quality associated with cotton leaf curl disease of Bt-cotton in Punjab. J Environ Biol 34:113–116

Yang L, Chen J, Huang C, Liu Y, Jia S, Pan L, Zhang D (2005) Validation of a cotton-specific gene, Sad1, used as an endogenous reference gene in qualitative and real-time quantitative PCR detection of transgenic cottons. Plant Cell Rep 24:237–245 . doi: 10.1007/s00299-005-0929-9

Yarwood CE (1979) Host passage effects with plant viruses. Adv Virus Res 25:169–190

Zaidi SS, Naqvi RZ, Asif M, Strickler S, Shakir S, Shafiq M, Khan AM, Amin I, Mishra B, Mukhtar MS (2020) Molecular insight into cotton leaf curl geminivirus disease resistance in cultivated cotton (Gossypium hirsutum). Plant Biotechnol J 18:691–706

Zubair M, Zaidi SS-A, Shakir S, Farooq M, Amin I, Scheffler JA, Scheffler BE, Mansoor S (2017a) Multiple begomoviruses found associated with cotton leaf curl disease in Pakistan in early 1990 are back in cultivated cotton. Sci Rep 7:1–11

Zubair M, Zaidi SS, Shakir S, Amin I, Mansoor S (2017b) An Insight into Cotton Leaf Curl Multan Betasatellite, the Most Important Component of Cotton Leaf Curl Disease Complex. Viruses 9