

## ISOLATION AND SEQUENCE ANALYSIS OF COTTON INFECTING BEGOMOVIRUSES

SARWAT HINA<sup>1</sup>, MUHAMMAD ARSHAD JAVED<sup>2</sup>, SALEEM HAIDER<sup>3</sup> AND MUHAMMAD SALEEM<sup>1</sup>

<sup>1</sup>Department of Botany, University of the Punjab, Lahore Pakistan.

<sup>2</sup>Fakulti Biosains Dan Biokejuruteraan, Universiti Teknologimaysia, Skudai 81310, Johor, Bahru, Malaysia

<sup>3</sup>School of Biological Sciences, University of the Punjab, Lahore Pakistan.

### Abstract

Cotton is an important cash crop of Pakistan which is severely suffered from a leaf curl disease called cotton leaf curl disease. Studies had shown that a number of begomoviruses collectively called cotton leaf curl viruses (CLCVs) are responsible for leaf curl disease. Some of the important viruses in this group are *Cotton leaf curl Multan virus*, *Cotton leaf curl Kohran virus*, *Cotton leaf curl Burewala virus*, *Cotton leaf curl Rajasthan virus*, and *Cotton leaf curl Allahabad virus*. In this study diseased cotton plants were screened for the presence of cotton-infecting begomoviruses. Amplified viruses were cloned and sequenced. Sequence analysis showed that Cotton leaf curl Burewala virus (CLCuBuV) is the causal organism of cotton leaf curl disease and this result is the affirmation of previous findings that, unlike first epidemic of the cotton leaf curl disease when a group of begomoviruses were affecting the cotton crop, there is a single virus CLCuBuV damaging the economically important crop of the country. Pairwise sequence comparison showed that sequences of isolates obtained in this study have maximum 98.9% to 99.55% homogeneity to published sequences of CLCuBuV. Surprisingly one of the isolates lacks an intact Rep, which is invariably present in begomoviruses.

### Introduction

Geminiviruses are a diverse group of minute insect vector-transmitted plant-infecting viruses with circular single stranded DNA (ssDNA) genome comprises of one or two components of 2700–3000 nt. (Moffat, 1999). Geminiviruses are responsible for a considerable amount of crop damage worldwide and their spread has increased because of a number of factors, such as the recombination of different geminiviruses in coinfection which enables virulent viruses to be developed, expansion of agriculture into new growing areas, transport of infected plant material to new locations and the expansion and migration of vectors that can spread the virus from one plant to another (Gray & Benerjee, 1999). On the basis of genome structure, host range and insect vector geminiviruses are divided into four genera- *Mastrevirus*, *Curtovirus*, *Begomovirus* and *Topocuvirus* (Fauquet & Stanley, 2003; Rybicki *et al.*, 2000 & Stanley *et al.*, 2005). These four genera infect a wide range of plants and lead to extensive crop losses worldwide (Mansoor *et al.*, 2003b; Morales & Anderson, 2001). The economically most substantial, geographically most widely distributed and numerous geminiviruses are in the genus *Begomovirus*. Begomoviruses have gained much attention as a peculiar group of plant viruses due to their unique ssDNA genome and circular topology, an aptitude to infect dicotyledonous plants and the ease with which the genomes can be manipulated employing molecular approaches. In this respect, much has been learned about the specific function(s) of viral-encoded genes and the replication strategy employed (Brown, 2000).

On the basis of their genome organization, Begomoviruses can be divided into two groups. Most have their genome partitioned between two DNA molecules of approximately 2600 nt. each, known as DNA-A and DNA-B such type of begomoviruses are termed as bipartite, while others have a single DNA-A-like genome of about 2800 nt and termed as monopartite (Fauquet *et al.*, 2008). Monopartite begomoviruses are associated with additional circular ssDNA molecules, known as betasatellite or alphasatellite that are about half the size of DNA-A. Betasatellites have been involved in

pathogenicity but alphasatellites have no known function and are certainly not involved in symptom induction (Mansoor *et al.*, 1999 & Saunders *et al.*, 2008). Alphasatellites have only been shown to be present in plants infected with monopartite begomoviruses in association with betasatellites (Mubin *et al.*, 2009). The component designated DNA A encodes all virus functions required for DNA replication, control of gene expression and insect transmission whereas the DNA B component encodes two genes involved in virus movement within plants (Rybicki *et al.*, 2000; Stanley, 1983 & Noueir *et al.*, 1994) and their products are symptom determinants for the bipartite begomoviruses (Arnim & Stanley, 1992; Klinkenberg & Stanley, 1990). However, a small number of monopartite begomoviruses have been determined that lack the second component. For such kind of viruses, all viral products required for replication, gene expression, whitefly transmission, and systemic infection are encoded on a single component which is a homolog of the DNA A component of the bipartite begomoviruses (Navot *et al.*, 1991 & Rojas, *et al.*, 2001). DNA A has the potential to code for six gene products, i.e., AV1, AV2, AC1, AC2, AC3, and AC4. On virion-sense strand, AV1 and AV2 are encoded whereas AC1, AC2, AC3 and AC4 are coded by virion complimentary-sense strand. AV1 or the coat protein (CP) is the single protein that is required for virus assembly. AV2 is considered to be important for cell to cell movement of the virus (Padidam *et al.*, 1996). AC1 and AC3 take part in viral replication. AC2 is essential for transcriptional activation of viral genes whereas AC4 is involved in symptom development (Lazarowitz, 1992). Recently, these proteins have been found to function as suppressors of gene silencing (Bisaro, 2006 & Gopal *et al.*, 2007). DNA B encodes two gene products i.e., BV1 and BC1. On the sense strand BV1 is encoded whereas on the complimentary-sense strand BC1 is encoded and both are essentially involved in viral movement (Lazarowitz, 1992; Stanley & Townsend, 1985).

Monopartite viral *cis*-acting elements which take part in replication are located in the non-coding intergenic region (200 nt, analogous to CR sequences in bipartite genomes) and certain gene products like Rep are similar in function to the analogous ORFs of their bipartite

relatives. However, monopartite viruses differ from their bipartite counterparts in certain aspects, including ORFs involved in nuclear transport and long distance movement functions which are accredited to CP and other viral-encoded proteins whose functions are not still not understood completely (Bisaro, 1996 & Lazarowitz *et al.*, 1992). For the true monopartite begomoviruses, the coat protein V1 and C4 proteins seems to be functionally replace the missing DNA B products of the bipartite viruses (Rojas *et al.*, 2001). One potential effect of the absence of the movement functions encoded on DNA B is phloem limitation. Monopartite begomoviruses appear to be phloem limited, whereas some bipartite viruses infect other tissues as well those banded together with the phloem (Morra & Petty, 2000; Rojas *et al.*, 2001).

Available evidence represents that geminivirus disease complexes are swiftly expanding in terms of their geographical distribution and host range. For example, cotton leaf curl disease was basically a greater problem in central Pakistan but is now causing inclusive damage in India. In the same region, new diseases are emerging in crops such as chilli, papaya, tomato, and tobacco. The presence of such a distinct population of begomoviruses in a single region, coupled with the ability of these viruses to exchange genetic material by recombination (Padidam *et al.*, 1999; Roberts & Stanley, 1994; Saunders *et al.*, 2001 & Saunders *et al.*, 2002), raises the possibility of new virus diseases emerging to cause epidemics in previously unaffected crops, a problem that will be compounded by the selective use of exclusive cultivars, the widespread introduction of the whitefly vector and the movement of infected material. In view of the enduring growth in international trade and travel, it might be only a matter of time before whitefly-transmissible disease complexes reach the New World as currently happened with *Tomato yellow leaf curl virus* (TYLCV) with such serious consequences (Polston *et al.*, 1999).

In Pakistan, Cotton Leaf Curl Virus (CLCV) was detected near Multan in 1967 (Hussain & Ali, 1975). Affected cotton (*Gossypium hirsutum*) plants have begomovirus DNA-A like molecules (Nadeem, 1995 & Zhou *et al.*, 1998), which named subsequently as DNA-A, but DNA-B-like component has not been found (Liu *et al.*, 1998a & Mansoor *et al.*, 1999). Initially the disease did not get much attention because of its casual occurrence and less economic importance. Since 1987 it has become a major threat to Pakistan's cotton crop (Hussain & Mahmood, 1988). In 1992, Cotton leaf curl (CLCuV) disease was happened to be well established in the districts of Vehari, Sahiwal, Bahawalpur, Khanewal, and Multan, which ranged from 30-80% with *Bemisia tabaci* (Gennadius) in abundance (Hameed *et al.*, 1994). Its dominant symptoms include leaf curling, vein swelling, darkened veins, and enations in which leaf like structures are frequently developed on underside of the leaves (Mansoor *et al.*, 1997 & Harrison *et al.*, 1997). Cotton production in Pakistan has declined considerably due to many reasons, including unprecedented infestations by cotton leaf curl *Begomovirus* and its vector the whitefly. The advancement of moderate to very strong resistance in *Bemisia tabaci* to all of the frequently used insecticides encouraged numerous claims that these chemicals are depleting their effectiveness under field conditions. These findings boost the necessity of identifying and achieving sustainable pest management

practices for the cotton cropping system (Cahill *et al.*, 1994). So continuous screening of available cotton germplasm for CLCuV is compulsory to identify the resistance sources for CLCuV in order to obtain successful management of the disease.

## Materials and Methods

**Sample collection:** Cotton fields in various areas of Punjab province were visited and young leaves of infected plants were collected.

**DNA extraction from plant tissues:** Plant genomic DNA from cotton leaf tissue was extracted by using Cetyl Trimethyl Ammonium Bromide (CTAB) method depicted by Doyle and Doyle (1990) with some modifications.

**Quantification of DNA:** The DNA yield per gram of leaf tissue was measured at 260nm wave length. An Optical Density of 1 at 260nm wave length was taken as 50µg/mL of DNA.

**Amplification of DNA:** For DNA amplification of Begomoviruses by Polymerase chain reaction (PCR) primers used were Forward primers and reverse primers having the sequence:

Begomo F   ACGCGTGCCGTGCTGCTGCCCCATTGTCC  
Begomo R   ACGCGTATGGGCTGYCGAAGTTSAGAC

## Cloning of PCR product

**Ligation:** Amplified DNA by PCR was cloned using an insTA clone PCR cloning kit (fermentas) following the instructions given by the manufacturer.

**Transformation of *E. coli* competent cells:** Transformation of competent *E. coli* cells was carried out following the procedure described by Sambrook *et al.*, (1989).

**Preparation of competent *Escherichia coli* cells:** Competent cells were prepared essentially as described by Cohen *et al.* (1972). The strain of *Escherichia coli* used for vector transformation was DH5α for plasmid vector pTZ57R/T.

**Plasmid isolation:** Plasmids were also isolated with a QIAprep spin Miniprep Kit (Fermentas) for DNA sequencing.

**Digestion of plasmid DNA:** Digestion of plasmid preparations and PCR products was done the enzymes and their corresponding buffers following the instructions as given by the suppliers (Fermentas).

**Gel extraction and PCR product purification:** DNA from the gel was isolated by using a PCR Clean Up system (Nucleospin kit) following the instructions given by manufacturers.

**Sequencing and sequence analysis:** Selected plasmid clones were purified using a GeneJET Plasmid Miniprep Kit (Fermentas) and sent to Macrogen (South Korea) for sequencing with universal primers (M13F [-20] and

M13R [-20]). To extend the sequence specific primers were designed (primer walking). The sequence data were assembled and analysed with the aid of the Lasergene package of sequence analysis software (DNASTar Inc., Madison, WI, USA). Sequence similarity searches (Blast) were performed by comparing the sequence to other begomovirus/betasatellite sequences in the database (<http://www.ncbi.nlm.nih.gov/BLAST/>). Open reading frames (ORFs) were located using ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Final sequences were submitted to the EMBL database (<http://www.ebi.ac.uk/embl>).



Fig. 1. Foliar symptoms of cotton leaf curl disease i.e. leaf curling, vein thickening, and enations.

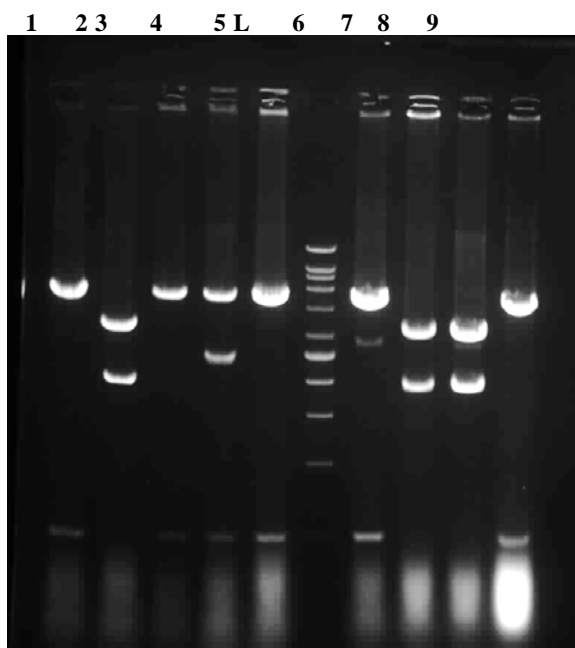


Fig. 3. Restriction analysis for confirmation of cloning of 2.8kb DNA molecules amplified from cotton samples. Samples were digested with *Hind*III endonuclease and sample no. 1, 2, 3, 5, 6, 7, 8 and 9 are desired clones. DNA fragments were compared with 1kb DNA ladder (L).

## Results

Leaves of a cotton plant showing typical symptoms of cotton leaf curl viruses which were vein thickening, leaf curling and enations (Fig. 1) were collected from a cotton field in Vehari district in the month of July 2010. Total genomic DNA of the sample was extracted and used as template in PCR with universal primers for begomoviruses (Begomo F and Begomo R). PCR amplified DNA of approx. 2.8kb size (Fig. 2) was cloned in a cloning vector pTZ57R/T and confirmed by restriction analysis (Fig. 3). Both clones (MV2A and MV2B) were sent to Macrogen, Korea for sequencing.

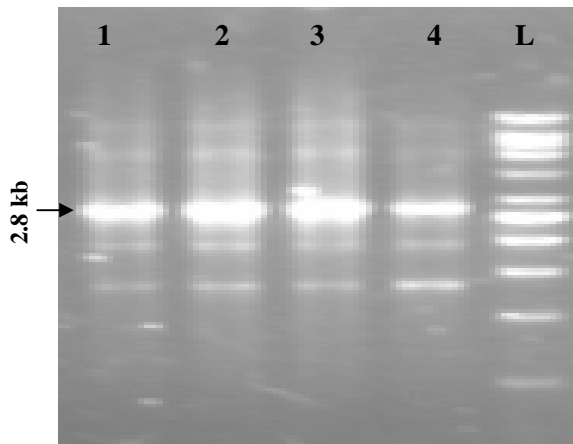


Fig. 2. PCR results showing approx. 2.8kb DNA amplified from cotton sample showing cotton leaf curl disease. Sizes of amplified products were compared with 1kb DNA ladder (L).

To determine whether the clones had some identity to known viral sequences, they were initially analyzed by BLAST available at the National Center for Biotechnology Information ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) and later on compared with twelve selected sequences, which were very close in percent identity, obtained from NCBI database on the basis of sequence alignment (Megalign, DNA STAR package software). Results showed that both clones had maximum sequence identity to CLCuBV as clone MV2A had 98.9% nucleotide sequence identity (Table 1) to CLCuBuV isolated from Indian Punjab (accession no. FN64932) and CLCuBuV isolated from Layya, Pakistan (accession no. AM774301). Clone MV2B had 99.5% sequence identity (Table 1) to CLCuBuV isolated from Indian Punjab (accession no. FN64932), CLCuBuV isolated from Layya, Pakistan (accession no. AM774301), CLCuBuV isolated from Lodhran (accession no. AM774300) and CLCuBuV isolated from Muzaffargarh (accession no. AM774302). Based on the presently applicable species demarcation threshold of 89% nucleotide sequence identity, these findings indicate that the virus identified in cotton is CLCuBuV.

Phylogenetic analysis also supports these results as both clones (MV2A and MV2B) are associated with CLCuBuV isolated from Pakistan and India (Fig. 4).

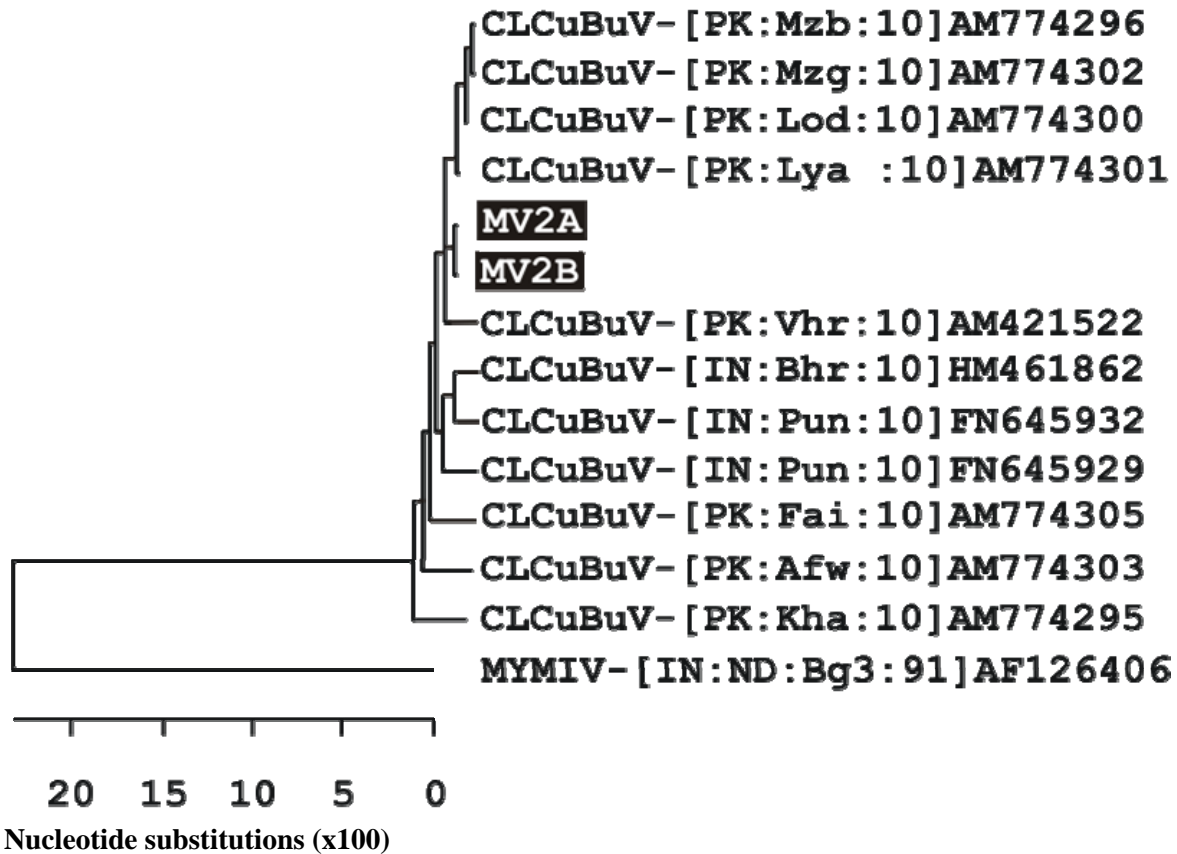


Fig. 4. Phylogenetic tree based on alignment of sequences of MV2A and MV2B clones produced in this study and some selected sequences of *Cotton leaf curl Burewala virus* (CLCuBuV) extracted from NCBI database. The tree was rooted on the sequence of *Mungbean yellow mosaic India virus* (MYMIV).

The sequencing results of both clones are as follows:

#### **MV2A**

ACCGGATGGCCGCGGATTTTTTTGTGGGCCCTACCATTA ACTCTTGTCGGCCAATCATATGACGCGCTCA  
AAGCTTAAATAATTCTCCCGCCTATTATAAGTACTTCGTTGCTAAGTTTGC GTTTGAAAAATGTGGGATCCA  
CTGTTAAATGAGTTCCCCGACACCGTTCACGGTTTTAGGTGTATGTTAGCAGTTAAATATTTGCAGTTGGTA  
GAGAAACTTACTCTCCTGATACATTGGGTTACGATTTGATAAGGGATTTAATCCTGGTAATAAGGGCTAG  
GAATTATGTGCAAGCGACCAGCAGATATAATCATTCCACGCCCGCTTCGAAGGTACGCCGCCGTCTGAAC  
TTCGGCAGCCCATACGCGTGCCGTGCTGCTGCCCCATTGTCCGCGTCAACAAAGCAAAGCATGGGCGAA  
CAGGCCCATGAACAGAAAGCCAGGATGTACAGGATGTACAGAAGTCCAGATGTTCTAGAGGATGTGAA  
GGTCCATGTAAGGTTTCAGTCGTTGAGTCCAGACATGATATTCAGCATATAGGTAAAGTAATGTGTGTTAG  
TGATGTTACTCGTGGTACTGGGCTGACCCATAGAGTTGGTAAGAGATTTTGTGTTAAGTCTGTTTATGTGTT  
GGGTAAGATCTGGATGGATGAGAACATTAAGACGAAGAATCACACGAATAGTGTGATGTTTTCTTGTTA  
GAGATCGTAGACCTGTTGATAAACCTCAAGATTTTGGAGAGGTATTTAATATGTTTGATAATGAGCCCAGT  
ACGGCGACTGTGAAGAATGTTTCATCGTGATAGGTATCAAGTTCTGCGCAAATGGTATGCAACTGTCACCGG  
TGGACAATACGCTTCAAAGGAACAAGCTCTCGTGAAGAAATTTATTAGAGTTAATAATTATGTTGTGTATA  
ACCAGCAGGAAGCTGGCAAGTATGAGAATCATTCTGAGAATGCTTTAATGTTGTATATGGCGTGTACTCAC  
GCCTCTAACCCAGTGTATGCTACCTGAAAGATACGGATCTACTTCTATGATTCCGTGACAAATTAATAAAT  
ATTGAATTTTATTGAAGATGATTGGTCTACAAATACAACATGTTGTAATACATTCCATAATCCATGATCAA  
CTGCTCTAACTACATTATTAATACTGACAATTCCTAAGTTATTTAAATATTTAAGCACTTGAGTCCTAAAGA  
CCCTTAAGAAACGACCAGTCGGAGGCTGTGAGGTCATCCAGATTCGAAAAGCTATGAAACATTTGTGTATC  
CCCAACGCTTTTCTCAGGTTGTGATTGAACTGGATCCTGATCGTGAGTATATCCATATTCGTCGTGAATGGA  
CGGTTGACGTGGCTGATGATCTTGAATAGAGGGGATTTGGAACCTCCAGATATATGCGCCATTCCCTGC  
TTGAGCTGCAGTGATGGGTTCCCCTGTGCGTGAATCCATGGTTGTGGCAGTTGATTGACAGATACTAAGAA  
CACCTCATTCAAGATCTACTCTCCTCCTGTTGCGTCTTTCGCTTCCCTGTGCTGTACTTTGATTGGAA

GCTGAGTACAGTGGTCCTTCGAGAGTGATGAAGATTGCATTTTTTAAGGCCCAATTCTTTAATGCTGTGTTTT  
 TTTCCCTCGTTGAGGAATTCATTATAACTGCTGTTCCGACCAGGATTGCACAGGAAGATTGTCGGTATCCC  
 GCCTTTAATTTGAACTGGCTTTCCGTATTTTGTGTTGGACTGCCAGTCTCTTTGGGCCCCCATGAACTCTTT  
 AAAGTGTTTGAGGAAATGCGGGTCGACGTCATCAATGACGTTGTACCAGGCGTCGTTACTGTATACTTTGG  
 GACTCAGGTCCAGATGTCCGCACAAATAGTTATGTGGTCCCAATGATCTAGCCCACATCGTCTTCCCCGTT  
 CGACTATCTCCCTCAATTACTATACTCCGAGGTCTAAGGGGCCGCGCAGCGGCATCGACAACGTTATCGAT  
 GGCCAAACTTCAAGTCTTCTGGAACCTTGATCGAAAGAAGAAGGAGAAAAAGGAGAAAATAGGGAGC  
 CGGTGGCTCCTGAAAGATTCTATCTAGATTTGCATTTAAATTATGAAATTGTAGTACAAAATCTTTAGGAG  
 CTAGTTCCTTAATGACTCTAAGAGCCTTGACTTACTGCCTGCGTTAAGTGCTGCGGCGTAAGCGTCGTTGG  
 CTGTCTGCTGCTCCTCTTGTGATCTTGCATCTATCTGAAACTCTCCCCACTCGAGAGTGTCCCCGCTCTT  
 GTCGATGTAGGCCTTGACATCTGAGCTTGATTTAGCTCCCTGAATGTTCCGGATGGAAATGTGCTGACCTGG  
 TTGGGGATAcCAGGTGCAAGAATCTGTTATTCGTGCAGACGAATTTGCCCTCGAACTGGATGAGCACATGG  
 AGATGAGGGCTCCCATCTTCGTGTAACCTCTCTGCAGAGTTGATGTATTTTTTATTCGAGGGTGTGTTGATGG  
 CTTGAATTTGGGAAAGTGCTCCTCTTTAGTGAGTGAGCACTGTGGATAAGTGATGAAATAGTTTTTGGCTT  
 GTACTTTAAAACGTTTGGGGGGAGCCATTGACTTTGGTCAATTAGAGACAACCTGATGGGCTTTTACTCTG  
 GGAATTGGAGACTGGATAACAATTTATAGTGTCTCAAATGGCATATTCTGTAAATAACTAGAAGTTCGTTT  
 GAAATTCAAATTCACCTTTGGGGTCCAAAAGCGGCCATCCGTATAATATT

### **MV2B**

ACCGGATGGCCGCGCGATTTTTTTGTGGGCCCTACCATTAACTCTTGTCCGCCAATCATATGACGCGCTCA  
 AAGCTTAAATAATTCCTCCCGCTATTATAAGTACTTCGTTGCTAAGTTTGCCTTTGAAAAATGTGGGATCCA  
 CTGTTAAATGAGTTCCCCGACACCGTTTACGGTTTTAGGTGTATGTTAGCAGTTAAAATTTGCAGTTGGTA  
 GAGAAAACCTACTCTCCTGATACATTGGGTTACGATTTGATAAGGGATTTAATCCTGGTAATAAGGGCTAG  
 GAATTATGTGCAAGCGACCAGCAGATATAATCATTCCACGCCCGCTTCGAAGGTACGCCCGCTGACT  
 TCGACAGCCCATACGCGTCCCGTGTGCTGCCCCATTGTCCGCGTCACCAAAGCAAAGCATGGGGCGAA  
 CAGGCCCCAAGACAGAAAGCCCAGGATGTACAGGATGTACAGAAGTCCAGATGTTCTAGAGGATGTGAA  
 GGTCATGTAAGGTTTCAGTCTGTTGAGTCCAGACATGATATTCAGCATATAGGTAAGTAATGTGTGTTAG  
 TGATGTTACTCGTGGTACTGGGCTGACCCATAGAGTTGGTAAGAGATTTTGTGTTAAGTCTGTTTATGTGTT  
 GGGTAAGATCTGGATGGATGAGAACATTAAGACGAAGAATCACACGAATAGTGTGATGTTTTTCTTGTTA  
 GAGATCGTAGACCTGTTGATAAACCTCAAGATCTTGGAGAGGTATTTAATATGTTTGATAATGAGCCCAGT  
 ACGGCGACTGTGAAGAATGTTTCATCGTGATAGGTATCAAGTTCGCGCAAATGGTATGCAACTGTCACCCG  
 TGGACAATACGTTCAAAGGAACAAGCTCTCGTGAAGAAATTTATTAGAGTTAATAATTATGTTGTGTTATA  
 ACCAGCAGGAAGCTGGCAAGTATGAGAATCATTCTGAGAATGCTTTAATGTTGTATATGGCGTGTACTCAC  
 GCCTCTAACCCAGTGTATGCTACCTTGAAGATACGGATCTACTTCTATGATTCCGTGACAAATTAATAAAT  
 ATTGAATTTTATTGAAGATGATTGGTCTACAAATACAACATGTTGTAATACATTCCATAATACATGATCAA  
 CTGCTCTAACTACATTATTAATACTGACAATTCCTAAGTTATTTAAATATTTAAGCACTTGAGTCTAAAGA  
 CCCTTAAGAAACGACCAGTCGGAGGCTGTGAGGTCATCCAGATTCGAAAGCTATGAAACATTTGTGTATC  
 CCAACGCTTTCCTCAGGTTGTGATTGAACTGGATCCTGATCGTGAGTATATCCATATTCGTCTGTAATGGA  
 CGGTTGACGTGGCTGATGATCTTGAATAGAGGGGATTTGGAACCTCCAGATATATGCGCCATCCCTGC  
 TTGAGCTGCAGTGATGGGTTCCCCTGTGCGTGAATCCATGGTTGTGGCAGTTGATTGACAGATACTAAGAA  
 CACCCTCATTCAAGATCTACTCTCCTCCTGTTGCGTCTTTCGCTTCCCTGTGCTGTACTTTGATTGGAA  
 GCTGAGTACAGTGGTCCTTCGAGAGTGATGAAGATTGCATTTTTTAAGGCCAATTCTTTAATGCTGTGTTTT  
 TTTCCCTCGTTGAGGAATTCATTATAACTGCTGTTCCGACCAGGATTGCACAGGAAGATTGTCGGTATCCC  
 GCCTTTAATTTGAACTGGCTTTCCGTATTTTGTGTTGGACTGCCAGTCTCTTTGGGCCCCCATGAACTCTTTA  
 AAGTGTGTTGAGGAAATGCGGGTCGACGTCATCAATGACGTTGTACCAGGCGTCGTTACTGTATACTTTGGG  
 ACTCAGgTCCAGATGTCCGCACAAATAGTTATGTGGTCCCAATGATCTAGCCCACATCGTCTTCCCCGTTCC  
 ACTATCTCCCTCAATTACTATACTCCGAGGTCTAAGGGGCCGCGCAGCGGCATCGACAACGTTATCGATGG  
 CCCAAACTTCAAGTCTTCTGGAACCTTGATCGAAAGAAGAAGGAGAAAAAGGAGAAAATATAGGGAGCCG  
 GTGGCTCCTGAAAGACTTCTATCTAGATTTGCATTTAAATTTAAGAAATTGTAGTACAAAATCTTTAGGAGCT  
 AGTTCCTTAATGACTCTAAGAGCCTCTGACTTACTGCTGCGTTAAGTGCTGCGGCGTAAGCGCTGTTGGCT  
 GTCTGCTGCTCCTCTTGTGATCTTGCATCTATCTGAAACTCTCCCCACTCGAGAGTGTCCCCGCTCTTGT  
 CGATGTAGGCCTTGACATCTGAGCTTGATTTAGCTCCCTGAATGTTCCGGATGGAAATGTGCTGACCTGGTT  
 GGGGATACCAGTTCGAAGAATCTGTTATTCGTGCAGACGAATTTGCCCTCGAACTGGATGAGCACATGGA  
 GATGAGGGCTCCCATCTTCGTGTAACCTCTCTGCAGAGTTTGTGATGTATTTTTTATTCGAGGGTGTGTTGATGG  
 CTTGAATTTGGGAAAGTGCTCCTCTTTAGTGAGTGAGCACTGTGGATAAGTGATGAAATAGTTTTTGGCTT  
 GTACTTTAAAACGTTTGGGGGGAGCCATTGACTTTGGTCAATTAGAGACAACCTGATGGGCTTTTACTCTGG  
 GAATTGGAGACTGGATAACAATTTATAGTGTCTCAAATGGCATATTCTGTAAATAACTAGAAGTTCGTTT  
 AAATTCAAATTCACCTTTGGGGTCCAAAAGCGGCCATCCGTATAATATT

Table 1. Highest and lowest percentage nucleotide sequence identities for pairwise comparisons of the sequences of the clones MV2A and MV2B with selected sequences of other cotton leaf curl disease-associated begomoviruses.

	MV2A	MV2B	AF126406	AM421522	AM774295	HM461862	AM774296	AM774300	AM774301	AM774302	AM774303	AM774305	FN645929	FN645932
MV2A		99.2	53.9	98.6	98.4	98.7	98.8	98.8	98.9	98.8	98.2	98.7	98.8	98.9
MV2B	0.2		54.2	99.2	99.1	99.3	99.4	99.5	99.5	99.5	98.8	99.3	99.4	99.5
AF126406	46.4	46.6		54.1	54.5	54.1	54.3	54.4	54.5	54.4	54.2	54.5	54.4	54.3
AM421522	0.7	0.7	46.6		99.1	99.3	99.3	99.3	99.4	99.3	98.9	99.3	99.3	99.5
AM774295	0.9	0.9	46.1	0.9		99.2	99.2	99.2	99.3	99.2	99.1	99.2	99.3	99.4
HM461862	0.6	0.6	46.5	0.7	0.8		99.4	99.5	99.5	99.5	99.0	99.5	99.6	99.8
AM774296	0.5	0.5	46.4	0.7	0.8	0.6		99.7	99.8	99.7	99.0	99.5	99.5	99.6
AM774300	0.5	0.5	46.3	0.7	0.8	0.5	0.3		99.9	99.8	99.0	99.5	99.5	99.6
AM774301	0.4	0.4	46.2	0.6	0.7	0.5	0.2	0.1		99.9	99.1	99.6	99.6	99.7
AM774302	0.5	0.5	46.3	0.7	0.8	0.5	0.3	0.2	0.1		99.0	99.5	99.5	99.6
AM774303	0.9	0.9	46.5	0.9	0.7	0.8	0.8	0.8	0.7	0.8		99.0	99.0	99.2
AM774305	0.7	0.6	46.1	0.7	0.8	0.5	0.5	0.5	0.4	0.5	0.8		99.5	99.6
FN645929	0.6	0.5	46.1	0.7	0.7	0.4	0.5	0.5	0.4	0.4	0.8	0.5		99.7
FN645932	0.4	0.4	46.3	0.5	0.6	0.2	0.4	0.4	0.3	0.4	0.6	0.4	0.3	

Further analyses showed that the genes encoded by clones MV2A and MV2B had typical arrangement of begomoviruses. In clone MV2A nucleotide position of pre-coat protein (AV2) was 132-488, Coat protein (CP) 292-1062, Replication enhancer protein (REN) 1059-1463 and AC4 2244-2684. Rep was missing in this clone because of four mutations which were addition of T nucleotide at 1718 and 1785 and G nucleotide at 2592 and deletion of T nucleotide at 2466. In clone MV2B coat protein (CP) was fused with pre-coat protein (AV2) and they had nucleotide position of 132-1061. Nucleotide position of AC4 was 2241-2681, Replication enhancer protein (REN) 1058-1462 and Replication associated protein (Rep) 1504-2595.

## Discussion

Cotton leaf curl disease (CLCuD) is associated with cotton leaf curl begomoviruses and DNA betasatellites. The plants affected by CLCuD contain monopartite begomoviruses (Liu *et al.*, 1998a & Mansoor *et al.*, 1999). Seven species of begomoviruses have been reported, from which five were identified in Pakistan, one in India and one in Sudan (Amin *et al.*, 2006; Sharma & Rishi, 2007). Cotton leaf curl disease, which was formerly known as African leaf curl of cotton, was first recognized in Nigeria in native cotton species *Gossypium peruvianum* and *Gossypium vitifolia* (Farquharson, 1912). Jones and Mason (1926) investigated a second outbreak in Nigeria in 1924. The disease was soon after reported in Sudan, Tanzania (Golding, 1930; Kirkpatrick, 1931 & Prentice, 1972), Pakistan (Hussain & Ali, 1975) and India (Rishi & Chauhan, 1994). CLCuD was first reported in Pakistan (Multan) in 1967 and severe outbreaks happened only after 1988 when it was identified in 24 ha of cotton variety S-12 (*Gossypium hirsutum*). During the year 2001-2002, it was observed that a new variant of Multan cotton leaf curl virus (CLCuV) was appeared on all resistant cotton varieties at various locations in Vehari district especially in Burewala territory (Akhtar *et al.*, 2002 & Mansoor *et al.*, 2003a). Recent results also revealed that only a single begomovirus Cotton leaf curl Burewala virus (CLCuBuV) is prevalent in Pakistan, in contrast to the situation before its appearance (Amrao *et al.*, 2007). CLCuBuV is a recombinant that comprises of sequences derived from *cotton leaf curl Multan virus* and *cotton leaf curl Kohran virus*. In this study CLCuBuV was found in the sample tested for cotton leaf curl viruses which is affirmation of previous results that this virus is still present in the field. This sample was collected from Vehari which is very close to Burewala an area where CLCuBuV was first reported. CLCuBuV is also reported from other areas of Pakistan which are Muzaffarabad (Multan), Muzaffargarh, Lodhran, Layya, Behawalpur, Faisalabd, Arifwala and Khanewal and also from India from New Delhi and Punjab.

Clones MV2A and MV2B were amplified from same cotton sample and sequence analysis shows that clone MV2A lacks an intact Rep. Its presence in the sample indicates that it still replicates probably by hiring the Rep protein encoded by other homologous viruses and it seems to be common in geminiviruses as in a similar study on legume-infecting begomoviruses many isolates were found which were defective for one or two geminivirus-encoded proteins (Ilyas *et al.*, 2010). Genome of viruses mutate a lot

and in geminiviruses recombination of genome is very common and during all this process some molecules are produced which are highly virulent and this is the basis of evolution of viruses. CLCuBuV is an outcome of recombination and it is without an intact Transcriptional activation protein (*TrAP*) already (Amrao *et al.*, 2010). Probably this virus is under an immense pressure of selection which may be the reason for its high rate of mutation. However further study is required to confirm this hypothesis. Plants have different ways to restrict the virus from infection. One of the most considerable is the presence of so-called resistance (R) genes. Each R gene provide resistance to a peculiar virus by triggering localized areas of cell death around the infected cell, which could be seen by unaided eye as large spots. So infection is blocked from being spread by this way (Kumar *et al.*, 2000). RNA interference is also an effective defense mechanism in plants (Shors, 2008). When plants are infected, they usually produce natural disinfectants that kill viruses, such as salicylic acid, nitric oxide, and reactive oxygen molecules (Soosaar *et al.*, 2005). Plant defense against viral infection also constitute, the use of siRNA in response to dsRNA (Alberts *et al.*, 2002). Most of the plant viruses encode a protein to suppress this response (Ding & Voinnet, 2007). Plants may also reduce transport through plasmodesmata in response to injury (Oparka & Alison, 2001).

Study of viruses in the field is very important to develop resistance strategy against viruses. In order to develop resistance against Geminiviruses, different approaches have been made so far which are (1) pathogen derived resistance through the expression of viral protein such as replication associated protein and movement protein, (2) pathogen derived resistance without expression of viral protein through RNA interference and (3) resistance due to expression of the non pathogen derived antiviral agents such as virus induced cell death causing protein and DNA binding protein. Despite of all these efforts only moderate resistance has been developed which can be easily overcome by the high genetic recombination and mutation of Geminiviruses (Shepherd *et al.*, 2009).

This study has a number of pivotal benefits as far as the improvement of our particular crop is concerned. Most importantly the work may help in providing an overview of the field conditions if carried out on a broader scale. Besides, it gives us an idea as to which genotypes of a specific virus are prevalent in the field as well as the manner in which it affects the cotton crop. Moreover, the symptoms developed by the plant in case of infection can be recorded during the course of experiment. The study shows tremendous potential when it comes to presenting a complete picture of the field, once taken to a higher level. Eventually, based on our identification of the specific virus, effective strategies can be designed in order to eradicate it from the field.

## References

- Akhtar, K.P., M.A. Haq, M. Hussain and A.I. Khan. 2002. Whitefly transmitted geminiviruses and associated disorders in cotton: a review. *Pakistan Journal of Phytopathology*, 14: 140-150.
- Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter. 2002. Molecular Biology of the Cell. *Garland Science*, 7: 451-452.
- Amin, I., S. Mansoor, L. Amrao, M. Hussain, S. Irum, Y. Zafar, S.E. Bull and R.W. Briddon. 2006. Mobilisation into cotton and spread of a recombinant cotton leaf curl disease satellite. *Archives of Virology*, 151: 2055-2065.
- Amrao, L., I. Amin, M.S. Shahid, R.W. Briddon and S. Mansoor. 2010. Cotton leaf curl disease in resistant cotton is associated with a single begomovirus that lacks an intact transcriptional activator protein. *Virology*, 66: 333-355.
- Amrao, L., S. Mansoor, I. Amin, Y. Zafar and R.W. Briddon. 2007. Analysis of the components of the cotton leaf curl disease complex associated with resistance breaking. International Geminivirus Symposium and International ssDNA Comparative Virology Workshop, 20-26 May, Estalagem das Minas Gerais Ouro Preto, Brazil, 8.
- Arnim, A.V. and J. Stanley. 1992. Determinants of *tomato golden mosaic virus* symptom development located on DNA B. *Virology*, 186: 286-293.
- Bisaro, D.M. 1996. Geminivirus DNA Replication. *DNA Replication in Eukaryotic Cells*, 33: 833-854.
- Bisaro, D.M. 2006. Silencing suppression by geminivirus proteins. *Virology*, 344: 158-168.
- Brown, J.K. 2000. Molecular markers for the identification and global tracking of whitefly vector-begomo-virus complexes. *Virus Research*, 71: 233-260.
- Cahill, M., D. Johnson, K. Gorman and I. Denholm. 1994. Insecticide resistance in *Bemisia tabaci* from Pakistan. *Proceedings of Brighton Crop Protection Conference of Pest and Diseases*, 1: 431-460.
- Cohen, S.N., C.Y. Chang and L. Hsu. 1972. Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of *Escherichia coli* by R-Factor DNA (CaCl<sub>2</sub>/extrachromosomal DNA/plasmid). *Proceedings of National Academy of Sciences*, 69: 2110-2114.
- Ding, S.W. and O. Voinnet. 2007. Antiviral Immunity Directed by Small RNAs. *Cell*, 130: 413.
- Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus*, 12: 13-15.
- Farquharson, C.O. 1912. Report of Mycologist. *Annual Report*, 2: 196.
- Fauquet, C.M. and J. Stanley. 2003. Geminivirus classification and nomenclature: progress and problems. *Annals of Applied Biology*, 142: 165-189.
- Fauquet, C.M., R.W. Briddon, J.K. Brown, E. Moriones, J. Stanley, M. Zerbini and X. Zhou. 2008. Geminivirus strain demarcation and nomenclature. *Archives of Virology*, 153: 783-821.
- Golding, F.D. 1930. A vector of leaf curl of cotton in Southern Nigeria. *Empire Cotton Growing Review*, 7: 120-126.
- Gopal, P., P. Kumar, B. Sinilal, J. Jose, A.K. Yadunandan and R. Usha. 2007. Differential roles of C4 and βC1 in mediating suppression of post-transcriptional gene silencing: evidence for transactivation by the C2 of *Bhendi yellow vein mosaic virus*, a monopartite begomovirus. *Virus Research*, 123: 9-18.
- Gray and Banerjee. 1999. Mechanisms of arthropod transmission of plant and animal viruses. *Microbiol Molecular Biology Review*, 63: 128-148.
- Hameed, S., S. Khalid, E.U. Haq and A.A. Hashmi. 1994. Cotton leaf curl disease in Pakistan by a whitefly transmitted geminivirus. *Plant Diseases*, 78: 528.
- Harrison, B.D., Y.L. Liu, S. Khalid, S. Hameed, G.W. Otim-Nape and D.J. Robinson. 1997. Detection and relationships of *cotton leaf curl virus* and allied whitefly-transmitted geminiviruses occurring in Pakistan. *Annals of Applied Biology*, 130: 61-75.
- Hussain, T. and M. Ali. 1975. A review of cotton diseases of Pakistan. *The Pakistan Cottons*, 19: 71-86.
- Hussain, T. and T. Mahmood. 1988. Cotton leaf curl disease. *The Pakistan Cottons*, 32: 248.

- Ilyas, M., J. Qazi, S. Mansoor and R.W. Briddon. 2010. Genetic diversity and phylogeography of begomoviruses infecting legumes in Pakistan. *Phytopathology*, 91: 2091-2101.
- Jones, G.H. and T.G. Mason. 1926. Studies on two obscure diseases of cotton. *Annals of Botany*, 40: 759-772.
- Kirkpatrick, T.W. 1931. Further studies on leaf-curl of cotton in the Sudan. *Bulletin of Entomological Research*, 12: 323-363.
- Klinkenberg, F.A. and Stanley, J. (1990). Encapsidation and spread of *African cassava mosaic virus* DNA A in the absence of DNA B when agroinoculated to *Nicotiana benthamiana*. *Journal of General Virology*, 71: 1409-1412.
- Kumar, D.S.P., H.T. Wai and B.J. Baker. 2000. Structure—function analysis of the *tobacco mosaic virus* resistance gene. *Proceedings of The National Academy of Sciences*, 97: 4789-14794.
- Lazarowitz, S.G. 1992. Geminiviruses: genome structure and gene function. *Critical Review of Plant Sciences*, 11: 327-349.
- Liu, Y., D.J. Robinson and B.D. Harrison. 1998. Defective forms of *cotton leaf curl virus* DNA-A that have different combinations of sequence deletion, duplication, inversion and rearrangement. *Journal of General Virology*, 79: 1501-1508.
- Mansoor, S., I. Amin, S. Iram, M. Hussain, Y. Zafar, K.A. Malik and R.W. Briddon. 2003a. The breakdown of resistance in cotton to cotton leaf curl disease in Pakistan. *New Disease Report*, 7: 9-18.
- Mansoor, S., R.W. Briddon, Y. Zafar and J. Stanley. 2003b. Geminivirus disease complexes: an emerging threat. *Trends in Plant Science*, 8: 128-134.
- Mansoor, S., S.H. Khan and M. Saeed. 1997. Evidence for the association of a bipartite geminivirus with tomato leaf curl disease in Pakistan. *Plant Diseases*, 81: 958.
- Mansoor, S., S.H. Khan, A. Bashir, M. Saeed, Y. Zafar, K.A. Malik, R.W. Briddon, J. Stanley and P.G. Markham. 1999. Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. *Virology*, 259: 190-199.
- Moffat, A.S. 1999. Geminiviruses emerge as serious crop threat. *Science*, 286: 1835.
- Morales, F.J. and P.K. Anderson. 2001. The emergence and dissemination of whitefly transmitted geminiviruses in Latin America. *Archives of Virology*, 146: 415-441.
- Morra, M.R. and I.T.D. Petty. 2000. Tissue specificity of geminivirus infection is genetically determined. *Plant Cell*, 12: 2259-2270.
- Mubin, M., R.W. Briddon and S. Mansoor. 2009. Diverse and recombinant DNA betasatellites are associated with a begomovirus disease complex of *Digera arvensis*, a weed host. *Virus Research*, 142: 208-212.
- Nadeem, A. (1995). Molecular characterization of two cotton geminiviruses. *Virus Research*, 22: 9-18.
- Navot, N., E. Pichersky, M. Zeidan, D. Zamir and H. Czosnek. 1991. *Tomato yellow leaf curl virus*: a whitefly-transmitted geminivirus with a single genomic component. *Virology*, 185: 151-161.
- Noueiry, A.O., W.J. Lucas and R.L. Gilbertson. 1994. Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. *Cell*, 76: 925-932.
- Oparka, K.J. and G.R. Alison and Plasmodesmata. 2001. A Not So Open-and-Shut Case. *Plant Physiology*, 125: 123-126.
- Padidam, M., R.N. Beachy and C.M. Fauquet. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology*, 265: 218-225.
- Padidam, M., R.N. Beachy and C.M. Fauquet. 1996. The role of AV2 ("precoat") and coat protein in viral replication and movement in tomato leaf curl geminivirus. *Virology*, 224: 390-404.
- Polston, J.E., G.W. Simon, E. Hibert and M.T. Momol. 1999. Introduction of tomato yellow leaf curl virus in Florida and implications for the spread of this and other geminiviruses of tomato. *Plant Diseases*, 83: 984-988.
- Prentice, A.N. 1972. Cotton—With Special Reference to Africa. *Tropical Agricultural Series*, 5: 282.
- Rishi, N. and M.S. Chauhan. 1994. Appearance of leaf curl disease of cotton in northern India. *Journal of Cotton Research and Development*, 8: 179-180.
- Roberts, S. and J. Stanley. 1994. Lethal mutations within the conserved stem-loop of *African cassava mosaic virus* DNA are rapidly corrected by genomic recombination. *Journal of General Virology*, 75: 3203-3209.
- Rojas, M.R., H. Jiang, R. Salati, B.X. Cazares, M.R. Sudarshana, W.J. Lucas and R.L. Gilbertson. 2001. Functional analysis of proteins involved in movement of the monopartite begomovirus, *Tomato yellow leaf curl virus*. *Virology*, 291: 110-125.
- Rybicki, E.P., R.W. Briddon, J.E. Brown, C.M. Fauquet, D.P. Maxwell, B.D. Harrison, P.G. Markham, D.M. Bisaro, D. Robinson and J. Stanley. 2000. Geminiviridae. *Virus Taxonomy: Seventh Report on the International Committee on Taxonomy of Viruses*. Academic Press, San Diego, 285-297.
- Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2: Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Saunders, K. 2002. Characterization of *Sri Lankan cassava mosaic virus* and *Indian cassava mosaic virus*: evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology*, 293: 63-74.
- Saunders, K., I.D. Bedford and J. Stanley. 2001. Pathogenicity of a natural recombinant associated with ageratum yellow vein disease: implications for begomovirus evolution and disease aetiology. *Virology*, 282: 38-47.
- Saunders, K., R.W. Briddon and J. Stanley. 2008. Replication promiscuity of DNA- $\beta$  satellites associated with monopartite begomoviruses; deletion mutagenesis of the *Ageratum yellow vein virus* DNA- $\beta$  satellite localises sequences involved in replication. *Journal of General Virology*, 89: 3165-3172.
- Sharma, P. and N. Rishi. 2007. Cotton leaf curl disease, an emerging whitefly transmissible begomovirus complex. *Plant Viruses*, 1: 128-133.
- Shepherd, D.N., D.P. Martin and J.A. Thomson. 2009. Transgenic strategies for developing crops resistant to geminiviruses. *Virus Research*, 4: 22-45.
- Shors, T. 2008. Understanding Viruses. *Jones and Bartlett Publishers*, 223-299.
- Soosaar, J.L., T.M. Burch-Smith and S.P. Dinesh-Kumar. 2005. Mechanisms of plant resistance to viruses. *The National Review of Microbiology*, 3: 789-798.
- Stanley, J. 1983. Infectivity of the cloned geminivirus genome requires sequences from both DNAs. *Nature*, 305: 643-645.
- Stanley, J. and R. Townsend. 1985. Characterization of DNA forms associated with cassava latent virus infection. *Nucleic Acids Research*, 13: 2189-2206.
- Stanley, J., D.M. Bisaro, R.W. Briddon, J.K. Brown, C.M. Fauquet, B.D. Harrison, E.P. Rybicki and D.C. Stenger. 2005. Family Geminiviridae. In *Virus Taxonomy Classification and Nomenclature of Viruses Eighth Report of the International Committee on the Taxonomy of Viruses*. Elsevier Academic Press, 55: 301-326.
- Zhou, X., Y. Liu, D.J. Robinson and B.D. Harrison. 1998. Four DNA-A variants among Pakistani isolates of *cotton leaf curl virus* and their affinities to DNA-A of geminivirus isolates from okra. *Journal of General Virology*, 79: 915-923.