Chapter 1

1. Introduction

1.1. Family *Geminiviridae*

According to the 8th report of classification and nomenclature of viruses by the International Committee on Taxonomy of Viruses (ICTV), the plant viruses are categorized into 20 families, 88 genera and ~750 species (Stanley *et al*., 2005).

Geminiviruses, first described by Goodman in 1977, belong to the plant virus family *Geminiviridae*. They are obligate intracellular viruses, a feature characteristic of all viruses (Goodman, 1977). *Geminiviruses* have circular ssDNA genome encapsidated within a geminate particle.

Geminiviruses are accountable for drastic losses in agricultural yield as a consequence of causing diseases in a varied range of plants including wheat, maize, cassava, tomato and cotton (Hanley-Bowdoin *et al*., 1999; Morales and Anderson, 2001; Mansoor *et al*., 2003). The small genome size accounts for overlapping genes in *geminiviruses* to proficiently code the proteins required for DNA replication, regulation of gene expression, particle encapsidation and movement. The family *Geminiviridae* was established in 1995, previously categorized as a group named geminivirus (Murphy *et al.*, 1995).

1.1.1. Classification of Geminiviruses

The family *Geminiviridae* consists of seven genera i.e. *Begomovirus, Mastrevirus, Curtovirus, Topocuvirus, Turncurtovirus, Becurtovirus* and *Eragrovirus*. *Begomoviruses* are transmitted by single species of whitefly *(Bemisia tabaci*) and are the most extensively studied genera. *Mastreviruses* and *Curtoviruses* are transmitted by leafhopper while *Topocuviruses* are transmitted by treehopper. The classification is established upon

the genome arrangement, the host (monocot or dicot) and the taxonomy of the insect vector (Rybicki *et al.*, 2000; Fauquet and Stanley, 2005).

1.1.2. Symptoms

Geminiviruses are known to induce particularly variable symptoms, nonetheless, the symptoms can be classified into three categories: streak symptoms commonly for mastreviruses, golden mosaic symptoms for numerous begomoviruses, and leaf curl symptoms for curtoviruses, begomoviruses and topocoviruses and amalgamations of the last two types. Dwarfing and malformation of stem, leaves and flower are other common symptoms. While enations and vein thickening is rare, necrosis is exceptionally uncommon, yet these symptoms affect plants severely and tend to cause death (Brunt *et al*., 1996; Shepherd *et al*., 2011).

1.1.3. Evolution in Begomoviruses

Geminivirus diversification and speciation can be accredited to recombination. Recombination in begomoviruses is identified to lead to the evolution of more virulent viruses. Recombination is reported not only between different species but also across different genera of Geminiviruses. Recombination is a key factor in the evolution, diversification and speciation of emerging begomoviruses (Garcia-Arenal *et al.*, 2001).

Cotton leaf curl Burewala virus in Pakistan is a recombinant between Cotton leaf curl Multan virus and Cotton leaf curl Khokhran virus. It is a significant and self-explanatory example of evolution of resistance breaking strain of cotton leaf curl disease (CLCuD) associated begomovirus (Amrao *et al.,* 2010). The recombination reliant replication strategy has supported this mechanism (Preiss and Jeske, 2003) by common natural mixed infections (Torres-Pacheco *et al.,* 1996; Harrison *et al.,* 1997; Ribeiro *et al.,* 2003) and by the unique ability of plant viruses to co-infect the same cell (Morilla *et al.*, 2004).

This mechanism is suspected to contribute to the high number of virus species in the family Geminiviridae.

It is discovered that the disease complexes are spreading to new geographical regions and their host range is also diversifying. CLCuD was initially a limitation factor in cotton production in Punjab province of Pakistan but is now documented to cause yield losses in India and China. Detection of new diseases in this part of the world and the capability of the viruses to recombine genetic material declares begomoviruses a serious threat (Padidam *et al*., 1999; Saunders *et al.,* 2002).

1.2. *Begomoviruses*

The name *begomovirus* is derived from *Bean golden mosaic virus* (Van Regenmortel *et al*., 1997) which is now called *Bean golden yellow mosaic virus* (BGYMV). *Begomovirus* is the largest and economically most important genus of the family *Geminiviridae*. Virus species in this genus are transmitted by the whitefly *Bemisia tabaci* (Gennadius) in a persistent and circulative manner. The agricultural output of many tropical, sub-tropical and irrigated arid agricultural countries including Pakistan is suffering at the expanse of *Begomoviruses* (Brown *et al*., 1995). The most devastating diseases of dicotyledonous plants, including cassava mosaic, cotton and tomato leaf curl, and bean golden mosaic with losses estimated up to several billions of dollars per year are caused by begomoviruses (Harrison and Robinson, 1997; Varma and Malathi, 2003).

Phylogenetic studies have shown that *begomoviruses* can be classified into the Old World viruses (OW) (those originating from Europe, Africa, Asia and Australasia) and the New World (NW) viruses (those originating from the Americas) (Rybicki, 1994; Paximadis *et al.,* 1999; Padidam *et al.,* 1999). Most of the bipartite *begomoviruses* are documented from NW, whereas in the OW most of the viruses are monopartite. Moreover, OW *begomoviruses* comprise AV2 gene in the DNA-A component, absent from the genome

of NW *begomoviruses* (Rybicki, 1994; Stanley *et al.*, 2005). The coat protein of NW *begomoviruses* comprises of a PWRLMAGT motif (a conserved sequence of amino acids) at the N-terminal, while the OW *begomoviruses* lack this motif (Harrison *et al.*, 2002). Moreover, OW *begomovirus*es commonly contain three iterons; Rep binding sites (Arguello-Astorga *et al.*, 1994).

Begomoviruses induce diverse disease symptoms in host plants, including leaf curl, vein thickening, vein yellowing, mosaic and in certain infections, yellow blotch and stunting (Harrison, *et al*., 1997; Briddon and Markham, 2001).

1.2.1. Genome Organization of Begomoviruses

Begomoviruses have been categorized as monopartite and bipartite by determining the genome organization. The genome of monpartite *begomoviruses* have one genomic component (DNA-A) while the genome of bipartite *begomoviruses* consist of two components (DNA-A and DNA-B), covalently closed, circular, ssDNA molecules of about 2.5kb to 2.9kb (Lazarowitz, 1992). Alphasatellites and betasatellites accompany monopartite *begomoviruses* (Briddon *et al.*, 2004). While alpha satellites can autonomously replicate in their hosts, betasatellites depend on the helper virus for replication. Betasatellites are discovered to contribute to symptoms induction. Even though alphasatellites can replicate independently, they depend upon the helper *begomovirus* for transmission by the vector and systemic infection in host cells. There is no reported contribution of alphasatellites in infectivity and symptom induction in host plants (Mansoor *et al*., 2003a; Cui *et al*., 2004; Saunders *et al*., 2008).

SsDNA viruses, with two genomic components (Rybicki *et al*., 2000) and contain six functional ORFs (Figure 1.1) are the basic identifiers of bipartite *begomoviruses*. DNA-A includes genes involved in viral DNA replication and particle encapsidation while the genes responsible for viral movement are present on DNA-B (Lazarowitz, 1992). DNA-A as well as DNA-B genomic components include a common region of approximately 200 nucleotides within the intergenic region (Sunter and Bisaro, 1991; Lazarowitz, 1992). The common region encompasses a 30 nucleotide conserved region (stem loop), which serves as the origin of replication (Sunter and Bisaro, 1991). The two deviating promoters regulating the expression of viral genes are also present in the common region (Lazarowitz, 1992).

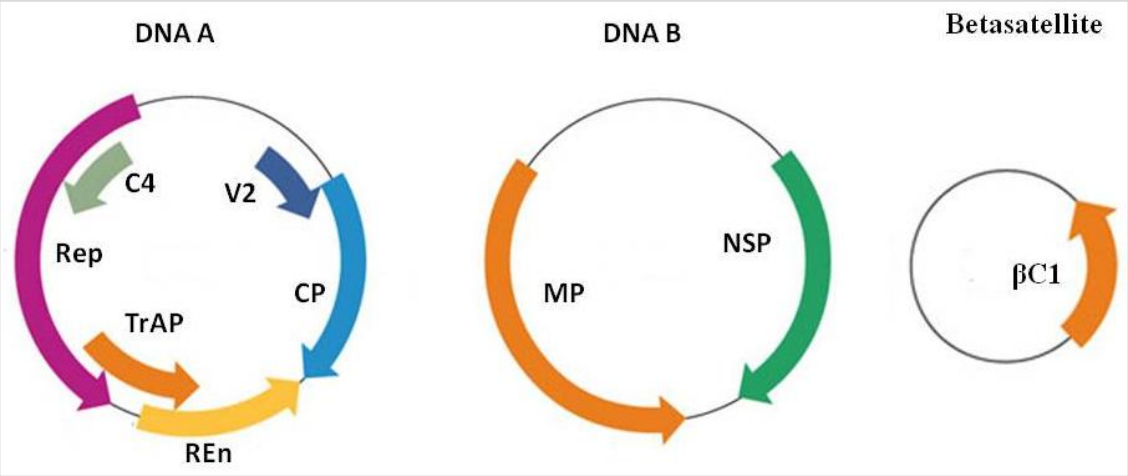

Figure 1.1: Genomic components DNA A, DNA B and betasatellite of begomoviruses along with the genes they encode. DNA A have 6 ORFs, DNA B have 2 genes whereas, betasateliite encodes only 1 gene βC1.

1.2.2. Proteins encoded by Begomoviruses

A single type of coat protein subunit, which has various functions including genome stability, infectivity, transmission by insect vectors and systemic spread (Lazarowitz, 1992), and ssDNA accumulation (Qin *et al*., 1998) is encoded by all begomoviruses.

There is irrefutable evidence that coat protein is a determining factor for the specificity of insect vector in all *begomoviruses*.

The pre-coat protein encoded by V2, restricted to Old World *begomoviruses*, was proved to be involved in the movement of bipartite *geminiviruses* by Padidam *et al.* (1996) employing mutation analysis.

The multifunctional Rep is accountable for extensive variety of functions throughout viral DNA replication. The functions include binding to iteron sequences found in the conserved intergenic region (Fontes et al., 1994), repression of its own promoter (Eagle et al., 1994) and DNA helicase activity (Pant et al., 2001). The plant protein retinoblastoma homologue, which controls the cell cycle and differentiation, was discovered to interact with Rep (Arguello-Astorga et al., 2004).

Transcriptional activator protein (TrAP) modulates the expression of the coat protein and the movement protein encoded by DNA-B component in bipartite begomoviruses (Sunter and Bisaro, 1991). TrAP is a nuclear protein (Sanderfoot and Lazarowitz, 1995) that transactivates virion-sense gene expression (Sunter et al., 1994). Moreover, TrAP is a pathogenicity factor that suppresses more than one host defense pathways (Sunter et al., 2001).

Replication enhancer protein (REn) modulates the viral DNA replication by activating CP gene (Azzam et al., 1994). With expression level equal to Rep, REn is found in nuclei of infected host cells (Nagar *et al.*, 1995), suggestive of its interaction with Rep in viral DNA replication (Table 1.1).

C4 protein is the symptom expression determining factor in monopartite *begomoviruses* (Rigden *et al*., 1994). C4 is located entirely within CP coding region but in an altered frame (Hanley-Bowdoin *et al.*, 1999). C4 proteins of bipartite begoviruses are highly variable and no function has been assigned to this gene product (Table 1.1).

The nuclear shuttle protein (NSP), one of the two proteins encoded by DNA-B (Figure 1.1), is restricted to the host cell nucleus and contributes in binding to ssDNA, subsequently transporting it to the cytoplasm (Pascal *et al.*, 1994). The movement protein (MP) accelerates the viral movement in the host by altering the structure of plasmodesmata (Pascal *et al.*, 1993). The role of movement protein in determination of symptom severity is well recognized (Duan *et al.*, 1997).

1.2.3. Satellite DNA

Alphasatellites and betasatellites are commonly found in most of the monopartite *begomoviruses*. The size of each satellite is ~1.4kb. Satellites are described as subviral agents composed of nucleic acid that are reliant on on co-infection with a helper virus for productive multiplication. Satellite nucleic acids have significantly distinctive nucleotide sequences from the genomes of their helper viruses (Briddon *et al.*, 2001; Briddon *et al.*, 2003).

1.2.4. Betasatellites

Betasatellites genome has a single ORF, encoding protein, βC1, which plays a significant and well established role in pathogenicity determination, accelerates helper DNA replication and suppresses plant defenses (Saunders et al., 2004; Cui et al., 2005). Betasatellites depend on the associated helper begomovirus for DNA replication. Both satellites have an A-rich region while betasatellites, in particular, also have a region of conserved sequence characterized as the satellite conserved region (SCR).

Table 1.1: Different functions are performed by proteins encoded by DNA A and betasatellite

Isolate 3AK	Gene	Protein	Function	Reference
DNA A	AV2	V2 (Pre-coat protein)	Pre-coat protein is involved in the movement of the virus	Padidam et al., 1996; Rojas et al., 2001
	AV1	V1/CP (Coat protein)	Stabilizes genome, cause infection, transmission and systemic spread by insect vectors is mediated via CP protein and ssDNA accumulation	Lazarowitz, 1992; Qin et al., 1998
	AC1/Rep	Rep (Replication-associated protein)	DNA replication, binding to iterons, interaction with plant proteins and control cell cycle	Elmer et al., 1988; Fontes et al., 1994; Ach et al., 1997; Castillo et al., 2003
	AC2/TrAP	TrAP (Transcriptional activator protein)	Viral transcription, suppression of host defense pathway, regulates expression of coat protein and movement protein and also acts as silencing suppressor	Sunter et al., 2001; Sunter and Bisaro, 1991; Van Wezel et al., 2003
	AC3/Ren	REn (Replication enhancer protein)	Controls viral DNA replication, interact with Rep during viral replication ,increases binding affinity of Rep and also interacts with plant proteins	Azzam et al., 1994; Nagar et al., 1995; Settlage et al., 2005
	AC4	C4	Symptom determination, suppression of gene silencing, movement and also acts as a virulence factor	Rigden et al., 1994; Vanitharani et al., 2005; Selth et al., 2004; Jupin et al., 1994; Krake et al., 1998
Betasatellite	βC1	βC1	Up regulates viral replication and suppress host defense, movement, host range determination, systemic spread and symptom induction	Cui et al., 2005; Saunders et al., 2004; Saeed et al., 2007; Saundres et al., 2000; Zhou et al., 2003

The βC1 protein is involved in determination of host range, movement and accumulation of satellite and helper virus (Saundres *et al.,* 2000; Zhou *et al.,* 2003). Betasatellites are capable of modulating the symptoms in hosts by intrusion of helper begomovirus replication (Saunders *et al.,* 2000; Briddon *et al.,* 2001).

1.2.5. Alphasatellites

Alphasatellites are molecules encoding their own Rep protein, thus, capable of self-replication (Figure 1.2). An A-rich region of 153-169 nucleotides, downstream of the *rep* gene, is common to all alphasatellites. Alphasatellites are believed to have evolved from nanoviruses. It is documented that alphasatellites can associate with monopartite begomoviruses and modify the symptoms and decrease DNA accumulation, however the mechanism is still unclear (Idris *et al*., 2011; Briddon *et al*., 2004).

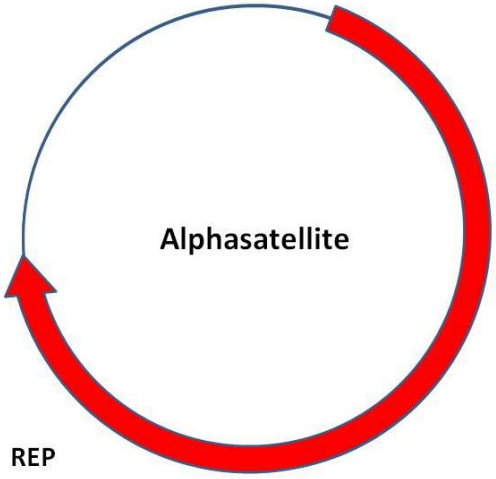

Figure 1.2: Alphasatellite comprises of single ORF encoding replication protein and are autonomously replicated.

1.3. Significance of Study

Hollyhock (*Alcea rosea*), family *Malvaceae*, is an ornamental plant widely grown in orchards, lawns and public gardens. The plant has been seen infected at a fairly high level i.e. 40-50% at different locations of Punjab province of Pakistan. Previously, Hollyhock has been found infected with a beta complex begomovirus from Egypt (Bigarre *et al*., 2001). Since hollyhock plant belongs to family *Malvaceae* same as of cotton, so there was a possibility that, it is harbouring Cotton leaf curl disease (CLCuD) associated

begomoviruses as an alternate host plant, which caused over US$5 billion losses to Pakistan economy during the mid to late 1990s (Briddon and Markham, 2000).

The present study is aimed at detection, identification and characterization of viruses from a begomovirus-like infected hollyhock plant sample collected from areas around Chakri (vein yellowing symptoms). The present study will enhance our knowledge about begomoviruses and help us to minimize the economic losses due to viral diseases by identifying alternate host plants of begomoviruses. Avoiding plantation of begomovirus host plants in and around crop cultivation areas can decrease the risks of begomovirus infection.

Literature Review

Begomoviruses cause serious disease of crop plants reducing both quality and quantity of plant product. They have been reported to infect vegetables, weed and ornamental plants. Plant viruses cause major losses to several agricultural and horticultural crops around the world, amounting to many million rupees. In Pakistan the major crops affected by Begomoviruses are Tomatoes, Beans, Cotton, Squash, Cassava and Chilli. Pakistan is an agricultural country and Begomoviruses are causing devastating yield losses in significant cash and food crops. Since virus is mutating/recombining and at a specific time there is a need of regular study for the identification of Begomoviral infection. Here previous work reported on begomoviruses is reviewed.

Shuja *et al.*, (2014), reported new species of CLCuBuV in cotton. They isolated and cloned both virus and its associated betasatellite from the sample using BurNF/BurNR and Beta01/Beta02 primers respectively. Analysis showed virus to be a recombinant containing sequences from two viruses, CLCuMuV and CLCuKoV. The virus was shown to lack an intact TrAP and it was encoding a 39 amino acid rather than 35 amino acid residual protein. Betasatellite was shown to be an isolate of CLCuMB but deeper analysis revealed it to encompass a sequence of tomato leaf curl virus which makes it an isolate of CLCuMBBur known to be involved in resistance breaking of CLCuD.

Marval *et al.*, (2014), provided an in-silico recombination analysis of Ageratum enation virus (AEV) isolated from an ornamental plant Merigold. Using Recombination Determination Program (RDP) AEV showed six recombination events based on the similarity of the sequence alignment proving AEV as a recombinant virus. First recombination was shown to have sequences from PedLCV and CrYVMV having starting and ending nucleotide coordinates at 2665-287. Second recombination event was

encounter at nucleotide positions 388 to 935 encompassing sequences from TLCuKV and CrYVMV. Third recombination was shown to occur at 938 to 1042 comprising sequences from AEV and TLCuRV. At 1042 to 1200 nucleotide position fourth recombination event was confirmed via RDP having sequences from AEV and TLCuRV.fifth recombination event was shown to begin at 1237 nucleotide position to 1366 nucleotide, major parent being TbCuSV and minor being CrYVMV. Final recombination was encountered at 2074 to 2210 nucleotide position having euphorbia leaf curl virus and bhendi yellow vein Bhubhaneswar virus.

Leke *et al.*, (2013), demonstrated recombinant virus infecting Okra plants in West Africa. The virus isolated from Okra plants was found to be a recombinant of cotton leaf curl Gezira virus (CLCuGeV) and okra yellow crinkle virus (OYCrV). The virus shared 87.8% nucleotide identity with CLCuGeV. Therefore, it can be regarded as a new species of begomoviruses. New strains of betasatellites and alpha satellites were also isolated from the infected plants. Betasatellite from CLCuGeV had 93.3% identity to CLCuGeV and was named as CLCuGeB whereas; alpha satellite shared 97.3% identiy and was named as CLCuGeA. Another alpha satellite had 95.2% okra leaf curl Burkina Faso alphasatellite so it was named as OLCuBFA. The results reflected diversity of begomoviruses and satellite complexes in infected okra plants in Africa.

Singh *et al.*, (2012), characterized two new monopartite begomoviruses infecting radish plants in India. First virus was Radish leaf curl Varanasi virus was a new recombinant species sharing 87.7% of identity with Tomato leaf curl Bangladesh virus.the other virus isolated from radish was an isolate of Croton yellow vein mosaic virus-India having 95.8% nucleotide identity. Furthermore, RDP analysis showed RaLCV is a recombinant having sequences from Euphorbia leaf curl virus and Papaya leaf curl virus. When these viruses were inoculated in plants they induced mild symptoms. However, when they were

coinfected with their betasatellites, resulted in symptom severity and increased in viral titres.

Ilyas *et al.*, (2012), from Pakistan first reported infection of begomoviruses in Catharanthus roseus. They isolated two viruses from the infected plant. Sequence analysis of both the viruses isolated proved them to be a recombinant of PedLCV and CrYVMV. First virus isolated KN4 was new specie having 86.5% identity to unpublished ChiLCIV and 84.4% identity to PaLCV. On the other side second virus KN6 was shown to have 99% identity with PaLCrV.

Fareed *et al.*, (2012), identified CLCuBuV for the first time from Ricinus communis. Virus and betasatellite was cloned using diagnostic primers of BurNF/BurNR and Beta01/Beta02. Virus showed 98.8% similarity to CLCuBuV lacking an intact TrAP gene which is typical of CLCuBuV. Betasatellite was found to have 96.3% identity to CLCuMB.

Anwar *et al.*, (2012), identified a new variant of ClYMV from infected croton plants which is grown as an evergreen shrub in gardens. Samples showing mild leaf curl and yellowing symptoms were collected from Lahore in 2012. Full length virus was amplified and sequenced showing 99.1% identity to ClYMV. Attempts to amplify DNA-B or β satellite component failed. It was a first report of ClYMV infecting croton and hence proving croton as a host plant of begomoviruses.

Kumari *et al.*, (2011), characterized a new species of tomato infecting begomovirus having 88.3% if identity with Tobacco curly shoot virus (TbCSV) and was thus named as Tomato leaf curl Ranchi virus (ToLCRnV). The betasatellite isolated showed 74.5% identity with Tomato leaf curl Bangladesh beta named as Tomato leaf curl Ranchi beta. Phylogenetic analysis showed ToLCRnV in close relationship with Tomato leaf curl Bangladesh virus, Tobacco curly shoot virus and Tomato leaf curl Gujarat virus. It was

also observed that β satellite was efficiently replicated by DNA A whereas, DNA B showed less accumulation. Leaf disk assay was also done which showed that DNA A can transreplicate homologous DNA B and β as compared to heterologous ones.

Pratap *et al.*, (2011), isolated a new variant of bipartite Tomato leaf curl New Delhi virus from eggplant in India. Virus was isolated and cloned. Further analysis of the sequences proved it to be a variant with virus having 97.6% nucleotide identity with Tomato leaf curl New Delhi virus [India:NewDelhi:Papaya:2005]. DNA B showed 94.1% identity with Tomato leaf curl New Delhi virus [India:Nagpur:Eggplant:2009]. Variants pathogenicity was confirmed through agroinfiltration showing yellow mosaic symptoms in eggplant and leaf curling symptoms in tomato plants.

Tahir *et al.*, (2010), identified new species of PepLCLV from capsicum spp. They reported it to be a recombinant of PaLCuV and ChLCV. Sequence analysis of the virus showed similarity of first ~2160 nucleotides to ChLCV having identity of 82.7% with these nucleotides and only 45.8% to the last 586 nucleotides. Whereas, PaLCuV showed 88.7% identity with last 586 nucleotides and only 82.2% identity with the first ~2160 nucleotides, proving it to be a recombinant. They also identified an asosociated betasatellite from PepLCLV. Being a monopartite virus it was a first report from Pakistan of an associated betasatellite from Capsicum spp.

Ohnesorge *et al.*, (2009), reported interaction of CP protein with small heat shock protein of transmission vector whitefly, Bemisia tabaci. CP was used to screen Bemisia tabaci cDNA library using the yeast two-hybrid system, in a search for interacting partners. A member of the small heat-shock protein family (termed BtHSP16) was identified and its interaction with the CP was verified by an in vitro pull-down assay.it was found that binding domain was located at the variable N-terminal part of the CP, while full-length BtHSP16 is required for the interaction.

Samad *et al.*, (2009), reported Tomato leaf curl Pakistan virus from Mentha samples. Total DNA was extracted from Mentha samples with and without symptoms from different locations. The presence of a begomovirus was detected by PCR using begomovirus coat protein gene-specific primers. A product 771 bp in size was amplified from samples with symptoms but not from symptomless plants. Amplicons were cloned into PCR-TOPO TA cloning vector and selected clones were sequenced in both orientations. Sequence analysis showed the highest levels of sequence identity (93%) with the begomovirus Tomato leaf curl Pakistan virus (isolate [-Pakistan:Rahim Yar Khan 1:2004]; DQ116884.

Fauquet *et al.*, (2008), proposed Geminivirus strain demarcation and nomenclature. Using the Clustal V algorithm (DNAStar MegAlign software), the distribution of pairwise sequence comparisons, for pairs of sequences below the species taxonomic level, identified two peaks: one at 85-94% nt identity that is proposed to correspond to "strain" comparisons and one at 92-100% identity that corresponds to "variant" comparisons. Guidelines for descriptors for each of these levels are proposed to standardize nomenclature under the species level.

Duffy *et al.*, (2007), estimated the evolution of single stranded Tomato yellow leaf curl virus by calculating mean genomic substitution rate. Sequences submitted in the GenBank during year 1988-2006 were analyzed for evolution studies. Mean genomic substitution rate was found to be 2.88 X 10-4 nucleotide substitution per site per year (subs/site/year). Calculating the substitution rate of CP gene, a conserved region, gave a result of 4.63 X 10-4 subs/site/year. Itergenic region which shows more frequent substitution gave a value of ~1.56 X 10-3 subs/site/year.

Kon *et al.*, (2007), reported a recombinant of Tomato leaf curl Java virus and Ageratum yellow vein virus-Java isolated from Ageratum conyzoides L. plants showing yellow vein

symptoms. DNA A of 2747 and β satellite of 1360 base pairs was amplified cloned and sequenced. Sequence analysis confirmed that virus is 91% similar to Tomato leaf curl Java virus. Whereas, ORFs C1 and C4 shared maximum identity of 91% and 95% with Ageratum yellow vein virus-Java respectively. Therefore, it was concluded that virus is a recombinant having sequence of Ageratum yellow vein virus- Java from 2389–2692. Infection with ToLCJV-Ageratum alone produced no symptoms but when coinfected with β satellite produced yellow vein symptoms. Nicotiana benthamiana plants were infected with the three viruses ToLCJV, AYVV-[Java] and ToLCJV-Ageratum showed similar symptoms which suggested exchange of C4 fragment which functions as pathogenicity determinant.

Briddon *et al.*, (2003), analyzed 26 different β molecules obtained from different geographical locations and revealed their architecture. They reported them as satellite molecules which function in symptom determination having an extremely conserved architecture comprising of single open reading frame of βC1 gene, adenine- rich region and satellite conserved region (SCR). Phylogenetic analysis of these betasatellites clustered them into two groups originating from Malvaceae and Solanaceae. It was concluded from these findings that β satellites coadapt on the basis of their host and geographical location.

Tahir *et al.*, (2006a), purified Bell pepper leaf curl virus, distinct monopartite begomovirus from infected bell pepper plants associated with leaf curl symptoms from Lahore. Total genomic DNA was isolated from symptomatic leaves and subjected to PCR amplification using a universal primer of coat protein gene, producing an expected size product of 750bp. The amplified product was cloned and sequenced. . The sequences analysis has the highest identity found from bell pepper to Tomato leaf curl New Delhi virus.

Tahir *et al.*, (2006b), screened a bipartite Begomovirus from infected leaf of Duranta repens exhibiting leaf curl symptoms. The whole genomic DNA was extracted from diseased leaf and PCR amplification was performed to confirm the presence of Begomovirus by using universal primer pair designed from conserved sequences of replication associated protein, producing an amplicon of approx. 1.5kb. This amplicon was cloned and sequenced. The DNA sequences revealed the maximum sequence identity tocroton yellow vein mosaic virus. DNA B produced approx. 2.8 kb fragment by PCR using specific primers pair. The sequence analysis of DNA B has the highest sequence identity with Tomato Leaf Curl New Dehli Virus.

Bull *et al.*, (2006), demonstrated the genetic diversity of begomovirus associated with Cassava mosaic disease (CMD) in major cassava growing region in Kenya. The genetic diversity among Begomovirus was identified by Restriction Fragment length Polymorphism. The full- length sequence reported the presence of East African cassava mosaic virus and East African cassava mosaic Zanzibar virus. These are novel species of begomovirus for which East African cassava mosaic Kenya virus is suggested .DNA-B was separated into eastern and western and eastern groups and much less diverse than DNA-A component. These studies suggested that diversification was encouraged by the movement of viruses across the adjacent countries.

Tahir *et al.*, (2005), reported for the first time isolation of tomato leaf curl New Delhi virus from bitter gourd samples in Pakistan. CP gene of the virus was amplified and cloned using universal primer for CP. Sequence analysis showed it to be 95% identical to tomato leaf curl New Delhi virus.

Haider *et al.*, (2005), detect begomoviruses from 5 different host plants, characterize them and built a relationship among them. The studies revealed that only SYLCV was mechanically transmissible to Nicotiana benthamiana while all viruses were be graft

transmissible. Serological studies showed their relevance to African cassava mosaic virus (ACMV). Indirect ELISA results clearly differentiated three viruses Ageratum yellow vein virus-Pakistan (AgYVV-PK), Solanum yellow leaf curl virus (SYLCV) and Eclipta prostrata yellow vein virus (EPYVV). The other two were confirmed as a strain. It was concluded that AgYVV and SYLCV were closely related to cotton leaf curl virus (CLCuV) than ACMV or EPYVV.

Tahir *et al.*, (2005), isolated a monopartite virus, Bell pepper leaf curl virus, from infected bell pepper plants showing leaf curl symptoms. The CP gene of virus of 750 bp was amplified cloned and sequenced. Later on sequence analysis showed its maximum identity to Tomato leaf curl New Delhi virus.

Cui *et al.*, (2005), demonstrated the binding activity and expression analysis of βC1 protein. Using electrophoretic mobility shift assays and UV- crosslinking experiments showed that βC1 is capable 0f binding both single stranded and double stranded DNA irrespective of any specificity of sequence or size. Nicotiana benthamiana plants were co-inoculated with GFP transgene alongwith TYLCCNV-Y10 and β. Patch agroinfiltration assay showed suppression of GFP transgene silencing as GFP mRNA accumulation was observed in leaf patches of N. benthamiana plants. Immunogold labeling provided evidence of nuclear localization of βC1 protein in infected N. benthamiana plants. Mutations in nuclear localization sequence of β, showed failure to induce symptoms, suppression of RNA silencing and nuclear accumulation. It was concluded that βC1 is required for symptom induction and RNA silencing suppression.

Gutierrez *et al.*, (2004), demonstrated geminivirus proteins interacting with host plant cell cycle proteins and DNA replication factors. To determine these interactions yeast two hybrid assays were utilized. The study showed interaction of wheat dwarf virus RepA protein interacting with host plant retinoblastoma related (RBR) protein. Moreover, it was

confirmed that viral Rep protein is not involved in RBR interaction, but deletion of C-terminal of the Rep protein imparts the same function as that of RepA protein suggesting interaction of this domain with RBR binding motif (LXCXE), which is also confirmed through secondary structure analysis.

Briddon *et al.*, (2004), observed the diversity of DNA-β in east and south East Asia. Cotton leaf curl virus and okra leaf curl virus from Pakistan and Ageratum yellow vein virus (AYVV) from Singapore were isolated. 17 β satellites were isolated, cloned and sequenced for diversity studies. Presence of DNA β was confirmed in almost all the virus complxes except for just two begomoviruses from Far East Asia. It was then concluded that β satellites are conserved in both their sequence and architecture.

Mansoor *et al.*, (2003), isolated small circular single stranded satellite DNA molecules from whitefly transmitted monopartite begomovirus infected plants. These satellites were named as β satellites necessary for virus accumulation and symptom expression whereas, itself it uses virus for its proliferation. Although these satellite molecules are diverse and have a vast host range but still have a much conserved architecture in terms of single ORF and satellite regulatory elements. Moreover, study highlights the importance of satellites as their host range is increasing and are distributed worldwide, they present a serious threat to agro-ecosystems worldwide.

Mansoor *et al.*, (2003), stated DNA B as an essential component of bipartite begomoviruses. Its replication depends on the Rep protein encoded by DNA A. the two DNA molecules interact by short motifs called iterons which are Rep binding sites. Further it was revealed that DNA β isolated from CLCuMV, a monopartite virus, lacks iterons but is trans-replicated by DNA A of the virus. Moreover, their interaction was confirmed by grafting a monopartite CLCuD sample in healthy plant, susceptible plant

and via whitefly in healthy plant. All three plants showed symptoms as that of bipartite viruses, confirming interaction of β molecules with viruses.

Idrees *et al.*, (2002), worked on samples of hollyhock and okra with leaf crumpling and curling symptoms respectively. Full length viruses with monopartite characteristics were cloned from both the samples named as HLCrV and OkLCV showing identity to old world begomoviruses such as CLCuV-SD having 84% and 95% identities. Individual alignments showed HLCrV to be 98% identical to AREV and OkLCV being 99% identical to OkEV. Alignments showed identity to viruses having enation symptoms and this can be explained by the amplification of satellite DNAs from both the viruses ranging from (741-1350 nucleotides).

Rojas *et al.*, (2001), analyzed proteins involved in the movement of TYLCV and suggested that CP and C4 are homologs of βC1. Monopartite begomoviruses have CP and C4 proteins which mediate movement of virus through mesophyll or epidermal plasmodesmata. Unlike bipartite begomoviruses movement protein βC1 which mediates cell to cell movement, CP and C4 have narrow capacity of doing that. They showed it via microinjection of Escherichia coli expressed proteins along with GFP fusion proteins expressed transiently. Expression of the proteins were examined at different growth stages of tomato plants resulting in the observation that only phloem cells of shoot apical, stem, leaves and floral tissue were infected.

Saunders *et al.*, (2001), demonstrated the importance of β satellite in inducing symptoms (a recombinant, having sequences from both DNA-A and β) by examining infection of AYVV in Ageratum conyzoides imparting yellow vein symptoms. They observed that DNA-A alone cannot induce symptoms and that satellite DNA is responsible for imparting symptoms during viral infection. For that they made a chimera similar to recDNA-Aβ17 and inoculate it into Nicotiana glutinosa. Inoculation resulted in mild

symptoms but similar to that seen in AYVV infection. On the basis of above mentioned observations they concluded that complex of recombinant DNA satellite molecules with viruses are responsible for symptom modulation in plants.

Briddon *et al.*, (2001), demonstrated that β satellite is necessary for inducing symptoms typical to that of begomoviruses. An infectious clone of CLCuV was developed and inoculated in cotton plants. There was an absence of typical vein swelling, vein darkening, leaf curling and enation symptoms from infected plants. They also described that complex of virus/ DNA-A/ β resulting in recombinant satellite having fragments from both DNA-A and β is totally a different molecule having the same function in symptom modulation.

Xie *et al.*, (2001), characterized a new species of begomoviruses, Tobacco curly shoot virus. Virus was isolated from tobacco plants showing curly shoot symptoms from Yunnan. DNA A of the virus was amplified cloned and sequenced. The sequencing showed virus to be of 2746 nucleotides having 6 ORFs. Similarity analysis showed 85% sequence identity with Tomato leaf curl virus from India. Coat protein showed 98% sequence identity with Cotton leaf curl virus from Pakistan.

Sanz *et al.*, (2000), suggested that multiple infections occur in cotton and other plants of malvaceous species along with recombination. It was identified that cotton plants having begomovirus infection in case of CLCuD encompass sequences from 2 or more begomoviruses. It was also observed that cotton plants contained sequences from non-malvaceous species and some non-malvaceous plant contained sequence of malvaceous species. Chimeric sequences were also seen, sequences having elements from malvaceous/non-malvaceous plants. A recombination site was also present at origin of replication of viruses. Hence it was concluded that recombination and multiple infection plays an important role in diversification of begomoviruses.

Polston *et al.*, (1999), studies tomato plants from Florida showing tomato yellow leaf curl virus symptoms. Three sets of primers were used to amplify full genome. 99% sequence identity was found with TYLCV-Is and was 98% identical to Israel isolate. They also found that virus spread within and outside the state was due to the transplants.

Materials and Methods

3.1. Sample Collection:

Althaea Rosea (Hollyhock) leaf samples showing typical begomovirus symptoms of vein yellowing were collected from Chakri, Punjab on 12 March 2014. Young leaves from symptomatic plants were collected and kept at -80^{o}C.

3.2. DNA Extraction:

For total DNA extraction protocol of Doyle & Doyle (1990), with some modifications was followed. Leaf sample of 1 g was crushed in mortar and pestle using liquid nitrogen until it turned into a fine powder. It was then transferred into falcon tubes containing pre heated 25ml CTAB isolation buffer solution (2% Cetyltrimethyl ammonium bromide, 1.4M NaCl, 20mM EDTA, 100mM Tris-HCl, and 0.2% mercaptoethanol) at 60^{o}C for 15 minutes. Falcon tube was placed in waterbath with constant shaking at 60^{o}C for 30 minutes.

After half an hour of heating and shaking, the mixture was cooled down at room temperature for 3 minutes and was transferred into 2 falcon tubes equally (12.5 ml each). Equal amount of Chloroform: Isoamyl Alcohol solution (24:1) was added in both the tubes. It was then mixed to form an emulsion. For separation of phases it was then centrifuged at 4000 rpm for 15 minutes at room temperature in Eppendorf centrifuge 5804R. Upper clear phase containing DNA from both the falcons was carefully transferred into a new falcon tube with pipette, leaving behind a few ml of clear phase to avoid contamination of DNA with proteins and RNA. Equal amount (almost 20 mls of chilled isopropanol) was added into the supernatant. Tube was mixed gently. White threads of DNA became visible. Precipitated DNA was centrifuged at 4000 rpm for 10 minutes. Supernatant was discarded and DNA was pellet down. DNA wash buffer (76% ethanol, 10mM ammonium acetate) was added to the pellet and pellet was transferred to a

new eppendorf tube alongwith the buffer. Eppendorf tube was centrifuged and supernatant was discarded. Pellet was incubated at 37^{o}C for 30-60 minutes and then dissolved into 1 ml of TE buffer (10mM Tris-HCl, 1mM EDTA) and stored at -20^{o}C.

3.3. DNA Analysis:

3.3.1. Agarose Gel Electrophoresis:

DNA concentration was determined by agarose gel electrophoresis. 1% agarose gel was made by adding 0.4 g of agarose in 40 ml of 1X TAE buffer (25 mM Tris, 5mM Glacial Acetic acid, 1mM EDTA, pH 8.0). Mixture was heated until it became clear and upon cooling Ethidium Bromide was added into it. Gel was poured into a gel caster having an appropriate comb and allowed to solidify. Solidified gel was placed into a tank containing 1X TAE buffer so that it is completely immersed into the buffer. Comb was removed carefully. 5µl of DNA mixed with 10X loading dye (0.4% bromophenol blue, 0.4% xylene cyanol FF and 25 % Ficol) was loaded into gel. 1 Kb DNA ladder (Fermentas) was also loaded into the well. Gel was run for 45 minutes at 80 volts and was then observed on Dolphin-Doc plus Image System.

3.4. PCR Amplification of DNA:

For all the 3 samples 50 µl reactions were made. The reactions consist of 10X Taq buffer [(NH4)2SO4, 25 mM MgCl2], 2 mM dNTP's, 100µM primers (SONAF 5′-GGGCCCCCATGAACTCTTTAAAGTG-3′, SONAR 5′-GGGCCCAAAGGGACTGG CAATC-3′) (Tahir et al., 2010), 1 µl *Taq* Polymerase and DNA depending upon its concentration

For betasatellite amplification, primers β01 and β02 were used (Briddon *et al.,* 2002). PCR reaction was carried out in Dual Block PCR. The conditions applied for amplification of required region are as follows (Figure 3.1 A and B):

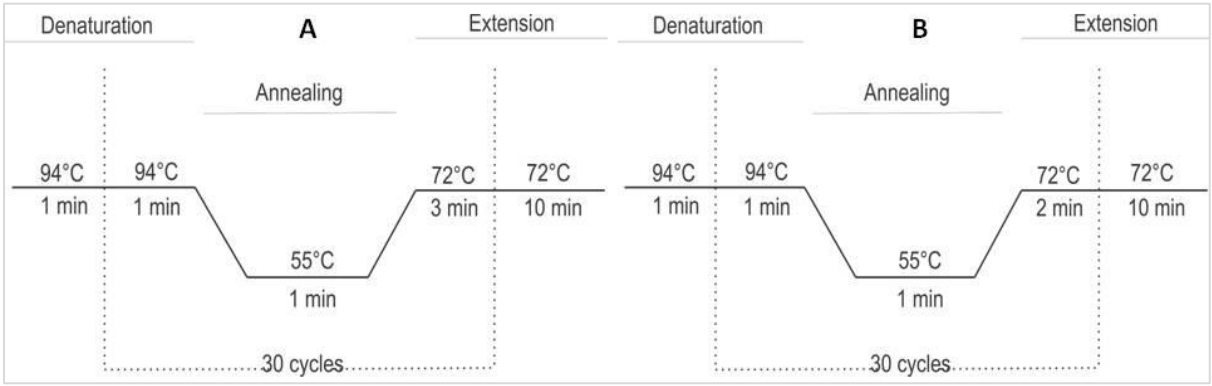

Figure 3.1: A, PCR conditions used for amplification of required fragment with SONAF/SONAR primers. **B,** PCR conditions used for amplification of betasatellite with β01/β02 primers. The difference is in primary extension time. For SONAF/R the time is 3 minutes while for beta01/02 extension its 2 minutes.

3.5. Gene Elution from PCR Product:

Required fragment was cut from the gel using razor blade and was kept in eppendorf tube. Silica Bead DNA Gel Extraction Kit (Fermentas) was used for gene clean according to manufacturer's instruction. Binding buffer in a 3:1 amount was added in eppendorf containing the excised band. It was then heated in a heat block for 15 minutes at 55^{o}C. The tubes were shaken after short intervals to mix the contents evenly. 7 µl of silica beads were added to the buffer solution and tubes were vortexed to distribute the beads evenly. Tubes were then heated again for 5 minutes at 55^{o}C. After heating they were centrifuged at 14800 rpm for 2 minutes to get a pellet. Supernatant was discarded and pellet was resuspended in 500 µl silica bead washing buffer provided with the kit. It was then again centrifuged at 14800 rpm for 1 minute. Supernatant was discarded and same washing step with silica bead washing buffer was repeated. Supernatant was again discarded and pellet was resuspended in 25 µl 1X TE buffer. Tubes were again heated at 55^{o}C for 5 minutes and were centrifuged for 3 minutes at 14800 rpm. Clear phase, the supernatant, was

carefully collected in a new labeled eppendorf. This gene clean product was used for ligation.

3.6. Cloning of PCR Product into pTZ57R/T Vector:

Gene clean product was ligated with pTZ57R/T vector using InsTAclone™ PCR Cloning Kit, Fermentas, according to manufacturer's instructions. Vector contains unique restriction sites (Figure 3.2). A total of 30 µl reaction was prepared in an eppendorf tube containing 1X ligation buffer, ~400 ng of purified DNA fragment, 1X PEG 4000 solution (5% w/v), 165ng restricted vector and 5 units of T4 DNA ligase. The ligation reaction was incubated overnight at 4oC in a water bath. Next day it was processed for transformation into DH5α competent cells.

3.7. Preparation of DH5α Cells:

Escherichia coli DH5α strain was used to prepare competent cells. A single colony from freshly grown plate was inoculated in 10 ml of Lauria Bertini Broth (LB Broth) (tryptone; 1% w/v, yeast extract 0.5% w/v and sodium chloride 0.5% w/v). Broth was autoclaved for 15 minutes at 121^{o}C before use. Inoculation was done in sterile conditions and was incubated overnight at 37^{o}C on continuous shaking. Next day culture was refreshed by transferring 2 ml of culture into 50 ml LB broth. It was again incubated at 37^{o}C for 2 hours in a shaking incubator. The culture was transferred into 50 ml falcon tube and was chilled on ice for 10 minutes. It was then centrifuged at 4000 rpm for 15 minutes at 4^{o}C to pellet down the cells. Supernatant was discarded and cells were resuspended in chilled 10 ml sterile 50 mM $CaCl_2$ solution. Cells were left on ice for 15 minutes and were again centrifuged as described previously. Supernatant was discarded and cells were resuspended in 2 ml of chilled 50 mM $CaCl_2$ solution. 50 µl of cells were aliquoted in eppendorf tubes and were stored at -80^{o}C for further use.

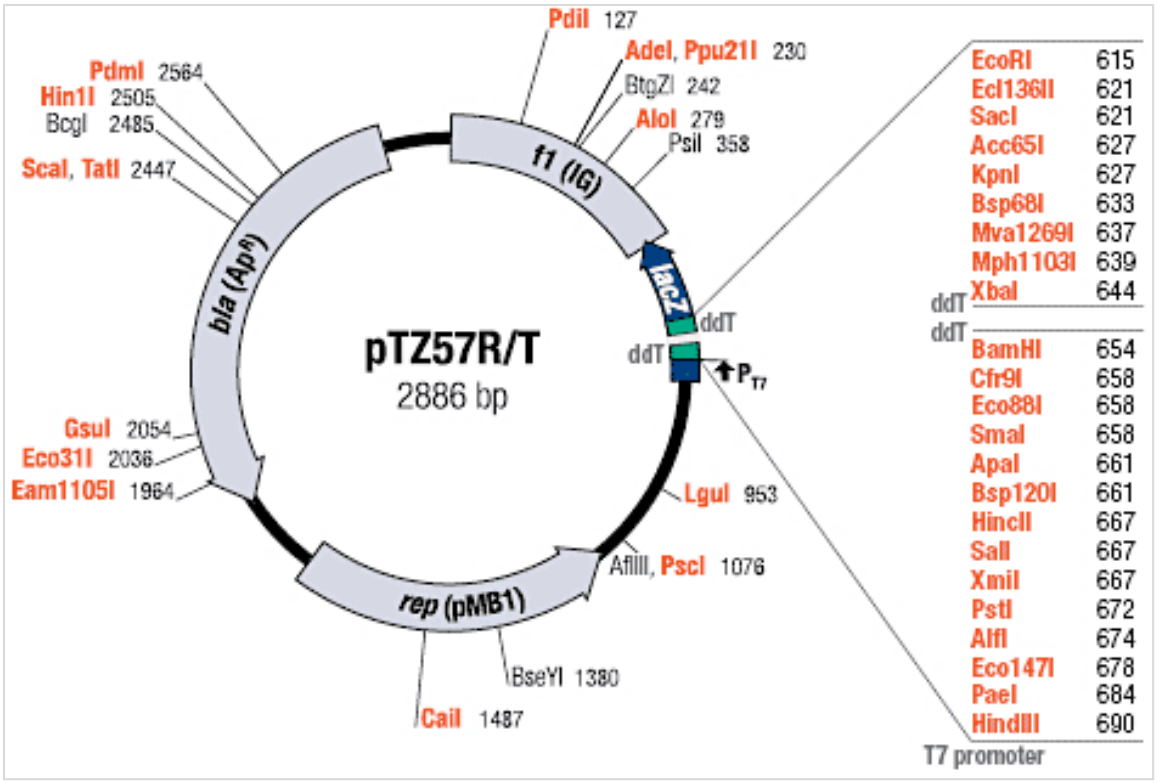

Figure 3.2: Map of the pTZ57R/T cloning vector. Unique restriction sites are indicated. (Image taken from http://www.fermentas.com/templates/files/tiny_mce/subsection_ images/k121_2. gif)

3.8. Transformation of Competent Cells:

Following the method described by Sambrook and Russell (2001) transformation of competent *E.coli* cells was carried out. Ligated product was transferred to eppendorf tube containing 50 µl of DH5α competent cells and was mixed gently. Cells containing ligated product was incubated on ice for 30 minutes. Heat shock was given to the cells for 2 minutes at 42^{o}C in a heat block. Cells were then placed on ice for 2 minutes. Adding 800 µl of LB medium to eppendorf tubes containing transformation mixture, cells were allowed to grow for 1-3 hours at 37^{o}C. Transformation mixture was spread on solid agar plates containing Ampicillin antibiotic and isopropyl-beta-D-thiogalactopyranoside (IPTG) and bromo-4-chloro-3-indolyl-β-galactopyranosidase (X-Gal) for blue/white screening of colonies. Plates were given overnight incubation at 37^{o}C.

3.9. Selection of Clones and Plasmid Isolation:

Next day white colonies were picked one at a time and were transferred to different falcon tubes containing 10 ml LB medium and 100 µg/ml Ampicillin. Falcon tubes were incubated in a water bath at 37^{o}C for 16 hours on continuous shaking. After incubation and growth 5 ml of culture was transferred to eppendorf tubes. Tubes were centrifuged at 14800 rpm for 1 minute. Supernatant was discarded and pellet was resuspended in 100 µl solution A (25 mM Tris-HCl, pH 8.0, 10 mM EDTA). Then 200 µl of solution B (1M NAOH, 10% SDS) was added and mixed by gently inverting the tubes. After that 150 µl soluction C (6 ml of 3 M potassium acetate, pH 4.8, 11.5 ml acetic acid, 28.5 ml H_2O) was added and vortexed light for even mixing. Tubes were then centrifuged at 14800 rpm for 10 minutes. Supernatant was transferred to a new eppendorf tube and 1 ml of chilled 100% ethanol was added to the supernatant. Tubes were incubated for 1 hour and again centrifuged for 10 minutes. Supernatant was again discarded and pellet was resuspended in 500 µl 70% ethanol. Following centrifugation for 5 minutes ethanol in supernatant was discarded and pellet was dried at 37^{o}C. After pellet had dried 40 µl of *RNase* A water (containing 20 µg/ml *RNase* A) was added to the pellet and pellet was resuspended. Extracted plasmid was confirmed by running 3 µl of plasmid DNA mixed with loading dye in 1% agarose gel.

3.10. Confirmation of Clones Via Restriction Enzyme Digestion:

Clones of bell pepper and hollyhock whose fragments were amplified with SONAF/SONAR primers were digested with *Apa1, EcoR1 and Hindi III* enzymes. While clones of hollyhock betasatellite amplified with Beta01/Beta02 primers were digested with *Kpn1, EcoR1* and *Hindi III.* Total of 30 µl reaction was prepared having 5 units of enzyme, 3 µl buffer and 5µl of plasmid DNA. Reaction tubes were incubated for 3 hours

at 37^{o}C. Digestion mixture was run in 1% agarose gel and confirmation of clones was done by looking at the size of fragments obtained.

3.11. Sequence and Phylogenetic Analysis:

For sequencing, plasmid isolation was done via kit (FavorPrep™ Nucleic Acid Extraction) according to manufacturer's instructions. For this purpose clone containing the desired transgene was streaked on agar plate. Again white colonies were picked and grown in LB medium before plasmid extraction. 25 µl of extracted plasmid was sent to Macrogen, Korea for sequencing.

After sequencing was done, resulting sequences were analyzed in BLASTN for local alignment based similarity search. On the basis of alignment complete sequences of viruses plus betasatellite were arranged using ClustalV, available in DNASTAR sequence analysis software package. Aligned sequences from BLAST were selected and retrieved to build a phylogenetic tree using CLUSTALX2. Trees were viewed and edit using tree view and Coral Draw software packages.

3.12. Recombination Analysis:

Recombination analysis of hollyhock virus was done using Recombination detection program 4 (RDP4) and Similarity plot (Simplot) software programs.

CHAPTER 4

4 RESULTS

4.1. Sample Collection:

Althaea Rosea (Hollyhock; isolate 3AK) leaves - symptomatic and symptomless were collected from areas around Chakri during 2014 and stored at -80^{o}C until processed for DNA isolation.

Figure 4.1: A, showing healthy symptomless plant **B,** shown is the infected hollyhock sample with vein yellowing symptoms.

4.2. DNA Isolation:

Total DNA was isolated from both symptomless and symptomatic samples and was stored at -20^{o}C until further processing. The good quality DNA was obtained while analyzed on to 1% agarose gel and compared with 1kb DNA ladder (Figure 4.2).

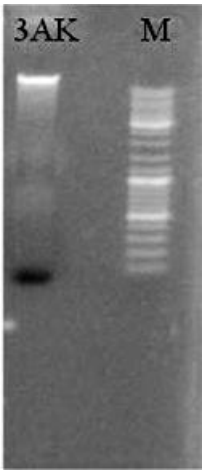

Figure 4.2: Agarose gel electrophoresis of total DNA of hollyhock. Lane 3AK showing the DNA extracted from symptomatic samples and lane M showing 1kb standard DNA marker. Total DNA band is above the ladder as its size is greater than the ladder.

4.3. PCR Amplification of DNA:

Abutting primers SonAF/SonAR (Tahir *et al.*, 2010) were used for the amplification of hollyhock DNA-A (section 3.4) and produced a band at approx. 2.8kb (Figure 4.3 B). Betasatellite was amplified using using abutting primer pair Beta01/Beta02 (Briddon *et al*, 2002; Section 3.4). An expected size band of 1350 base pairs in size was obtained using Beta01/Beta02 primers (Figure 4.3 C).

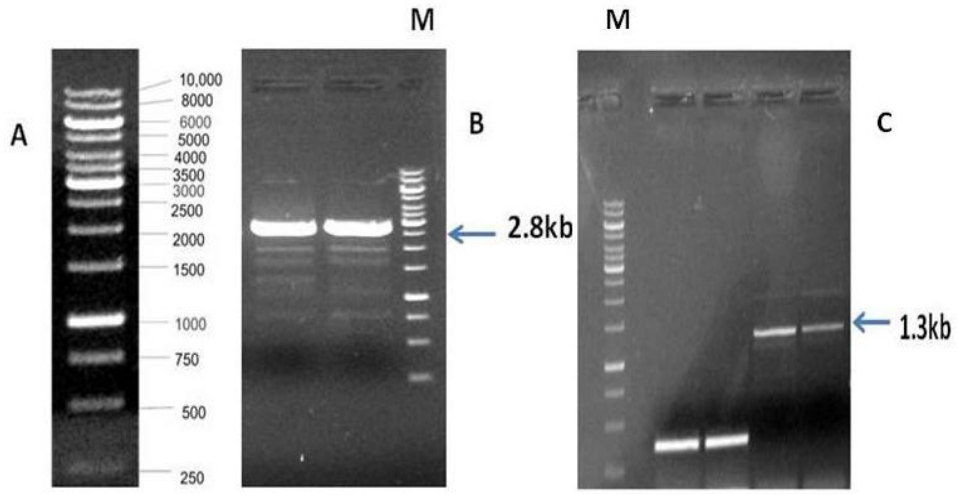

Figure 4.3: A, showing standard DNA marker of 1kb. **B**, Agarose gel electrophoresis showing full length amplification product of SonAF/SonAR of 2.8kb of isolate 3AK. **C,** Agarose gel electrophoresis showing amplified product at 1.3kb of beta01/beta02 primers.

4.4. T/A Cloning of begomovirus:

Required size PCR bands approx. 2.8kb (for DNA A) and 1.3kb (for betasatellite) were cut and were purified using Silica Bead DNA Gel Extraction Kit (see section 3.6) after running on 1% Agarose gel. The purified products were visualized under UV trans-illuminator. Thick bands of good quality for all the three samples were obtained. The resulted products were ligated in pTZ57R/T vector. *E.coli* DH5α competent cells were transformed with the ligated products. Transformed cells were spread on agar plates containing ampicillin, IPTG and X-gal.

4.5. Selection of Clones:

Many white and few blue colonies were appeared on the plates next day. White colonies were selected and cultured for their plasmids DNA isolation. Isolated plasmid DNA for both genomic component and satellite were subjected to restriction digestion analysis.

4.6. Confirmation of Clones through Restriction Digestion:

Clones of DNA A of isolate 3AK were digested with *Apa1* and *EcoR1* restriction enzymes. For genomic component, the bands obtained at 5.6kb with *EcoR1* restriction enzyme and with *Apa1* the bands obtained were at 2.8kb, 1.5kb and 1.3kb making total of 5.6kb (Figure 4.4 B). Positive clone of hollyhock DNA A was named as isolate 3AK. For betasatellite the bands obtained were at 4.1kb with *Eco*R1 and at 2.8kb and 1.3kb while digested with *Kpn*1 (Figure. 4.4 C). Betasatellite positive clone was named as isolate 4AK.

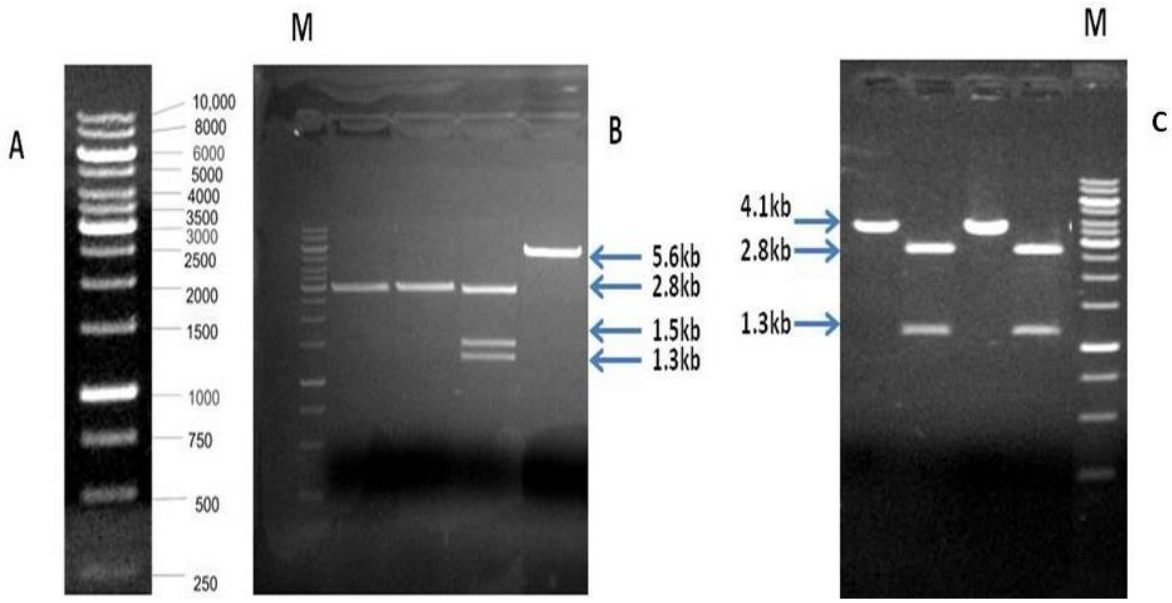

Figure 4.4: A, showing standard 1kb DNA marker. **B**, Agarose gel electrophoresis showing digestion of hollyhock DNA A with Apa-1 and EcoR 1 enzymes. With Apa1 bands obtained at 2.8kb, 1.5kb and 1.3kb and with EcoR1 single band at 5.6kb was obtained. **C**, Agarose gel electrophoresis showing digestion of hollyhock betasatellite with *Apa*-1 and *EcoR* 1 enzymes. Gel showing bands at 2.8kb, 1.3kb with *Kpn*1. *EcoR*1 digestion showing single band at 4.1kb.

4.7. Comparison of isolate 3AK with other Begomoviruses

The complete nucleotide sequence of isolate 3AK was determined to be 2741 nucleotides in length. Genomic component of isolate 3AK was compared to begomoviruses present in the databases. The analysis of isolate 3AK revealed its maximum nucleotide sequence identity of 86.6% with EuLCV (Table 4.1). Therefore, on the basis of species demarcation criteria (Fauquet *et al.*, 2003) isolate 3AK having less than 89% nucleotide seqeuence identity is a new species of begomoviruses for which named "*Hollyhock yellow vein mosaic Islamabad* virus" (HYVMIV) is proposed.

4.8. DNA Sequence Analysis

The isolate 3AK contains six (6) overlapping six ORFs with an arrangement typical of begomovirus (Table 4.2). Coat protein (CP) and precoat protein (V2) genes were present on virion sense strand while rest of the genes; Replication enhancer protein (REn) gene,

Transactivator protein (TrAP) gene, Replication associated protein (Rep) gene and C4gene were present on complementary sense strand Hairpin loop was present comprising the conserved nona-nucleotide sequence (TAATATTAC). TATATA box for the binding of Rep promoter was also present.

Betasatellite nucleotide sequence was determined to be 1352 nt and shared maximum identity (90%) with Kenaf leaf curl betasatellite (FN678779). As percentage identity was more than 78% therefore, it was characterized as an isolate of KLCuB. Betasatellite was comprising of single open reading frame, encoding *βC1* gene on complementary sense strand, showed 97% identity to Kenaf leaf curl βC1 gene (Table 4.3). Size of βC1 ORF was determined to be 356 nucleotides (coordinates 194 to 550) encoding 118 amino acid protein involved in symptom determination. Other important regions; A-rich region and nonanucleotide sequence TAATATTAC were also present.

Table 4.2: Residue position, length and protein size of different genes of virus and betasatellite

Isolate	Percentage Identity	Gene	Start Site	End Site	Total Residues	Protein Size	Sequence Identity (%)	Virus
3AK DNA A LK028571	86.6% EuLCV	AV2	140	505	365	121	89%	ToLCJV
		CP	300	1070	770	256	98%	RLCuV
		Rep	1519	2598	1079	359	94%	EuLCV
		TrAP	1212	1616	404	134	90%	CLCKV
		REn	1067	1471	404	134	93%	ChLCIV
		C4	2145	2447	302	100	89%	CLCuIV
Betasatellite LK028572	90% KLCuB	βC1	194	550	356	118	97%	KLCuB

Table 4.1: Sequence distance matrix showing percentage nucleotide sequence identities of isolate 3AK with the sequences of the most similar begomoviruses in the database

	HYVMV (2)	CLCuMV (2)	PedLCV (4)	ToLCKV (13)	ToLCKeV (3)	ToLCJoV (21)	ToLCuGNV (1)	ToLCuBDV (4)	TbCSV (16)	SHLDV (3)	RaLCuV (6)	PaLCuV (12)	EuLCuV (4)	CLCuKV (7)	AEV (28)	3AK (1)
3AK (1)	67.4-70.5	70.6-74.2	84.5-86.4	75.3-82.2	81.1-82.7	78-81.3	81.7	78.5-80.5	75.3-80.6	77.2-77.3	77.2-84.6	74.3-79.7	77.9-86.6	77.4-81.1	73.4-81.5	***
AEV (28)	65.2-72.3	68.4-73.4	74.5-86.5	71.4-84.2	73.5-82.4	72.3-81.9	77.4-84.9	72.9-85.5	74.1-87.4	72.2-82.1	70.9-87.3	66.1-84.9	73-81	70.7-81.9	80-100	
CLCuKV (7)	68.8-73.9	71.7-84.4	80.2-85.1	74-82.1	79.3-83	73.1-77	78.1-80.5	72.8-80.1	73.6-80.5	77.5-80.6	74.3-82	68.6-80.3	72.3-79.1	89.4-100		
EuLCuV (4)	63.7-71.3	68.6-74.1	77.6-84.3	73.2-80.4	76.1-80.6	76-80	77.1-81.7	76.2-79.7	76.2-79.7	76.1-80.2	71-83.9	72.3-79.4	89.7-100			
PaLCuV (12)	58.4-70.3	62.1-69.9	74-88.7	70.6-82.8	71.2-80.2	70.1-79.9	73-81.4	70.7-81.3	70.6-82.1	67.2-77.8	68.3-84.3	79.9-100				
RaLCuV (6)	66.5-70.4	71.4-75.6	80.6-88	70.7-81.9	75-83.2	72.3-81.9	79.6-85.5	72.8-84.4	62.5-86.9	67.4-87.6	78.5-100					
SHLDV (3)	64.5-73.2	66.3-74.6	83.4-83.8	73.4-78.7	81.4-82.4	74.3-77.3	78.9-79	74.9-81	78.1-81.5	100						
TbCSV (16)	63.2-66.6	67.6-72.8	77.9-84.2	76.5-84.9	77.4-80.4	77.8-87.5	77.7-81	77.1-89.5	84.2-100							
ToLCuBDV (4)	62.4-69.9	67.5-72.4	78.8-84.3	76.9-86.9	77.8-80.8	80-87.7	79.2-81.5	80.5-100								
ToLCuGNV (1)	63-69.7	67.5-72.2	83.9-87.1	80.1-87.5	85.8-87.8	78.9-81.3	100									
ToLCJoV (21)	68.5-73	70.6-72.8	77.8-83	77.2-83.4	76.9-79.3	84.2-100										
ToLCKeV (3)	63.2-69.4	67.2-71.8	82.5-87.1	80.1-85.7	95-100											
ToLCKV (13)	67-71.2	70.2-76.9	77.5-83.7	84.1-100												
PedLCV (4)	63.9-73.2	67.5-72.7	91-100													
CLCuMV (2)	74.4-77.4	88.1-100														
HYVMV (2)	89.2-100															

4.8.1. Multiple Sequence Alignment of C4 and βC1 Gene

Since both C4 and βC1 genes are involved in symptom determination therefore; nucleotide sequences of both the genes were aligned (Figure 4.8) to reveal any significant sequence similarity among both the sequences. Sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE), online multiple sequence alignment tool, available at EMBL-EBI website. The results showed 64.66% nucleotide sequence identity.

Figure 4.8: Multiple sequence alignment of C4 and βC1 gene showing 64.66% identity. Lane 1 represents the C4 and Lane 2 represents the βC1 gene sequence.

4.8.2. Analysis of Isolate 3AK Intergenic Region with Other Similar Begomovirus Intergenic Sequences

Sequence analysis of isolate 3AK intergenic region revealed maximum sequence identity of 97% with CLCMuV (JN558352). Sequences around TATA box almost last 143 nucleotides of intergenic region when aligned showed identity with CLCMuV. This suggests that it harbors iteronic sequences of CLCMuV. The length of intergenic region was 282 nucleotides. Two Rep binding sites (iterons) were also found to be present upstream of the TATA box. First unique iteron having sequence –TGGCAATCGGTG- occurred at 2601-2612 nucleotide position and second iteron occurred at 2628-2639 having the same nucleotide sequence (figure 4.6).

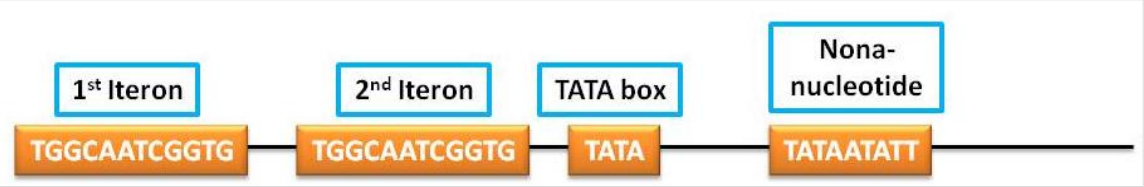

Figure 4.6: Position of iterons, TATA box and nona-nucleotide sequence in intergenic region

4.9. Phylogenetic Analysis of the Characterized Isolate 3AK

Phylogenetic tree was constructed based on the alignments of most similar begomoviruses sequences available in the database using neighbor joining (NJ) algorithm. The sequences considered here are shown in Figure 4.11. From the tree it is evident that isolate 3AK is closely related and cluster together with EuLCV. They have high sequence identity and a high bootstrap value. Isolate 3AK is highlighted and outgroup used in the tree, of distant begomovirus is *Cotton leaf crumple virus* (CLCrV), can be seen segregated from the other related sequences.

A phylogenetic tree was constructed to study relationship of betasatellite with other most similar betasatellite sequences available in the databases. Analysis showed that betasatellite is closely related to Eclipta yellow vein betasatellite (GQ478345), Kenaf leaf

curl betasatellite (KLCuB) clone HYBETA (FN678779) and KLCuB isolate Faisalabad (FR772083) and are present in the same cluster. Reference outgroup BhYVMA (FN658717) is segregated from other sequences (Figure. 4.13).

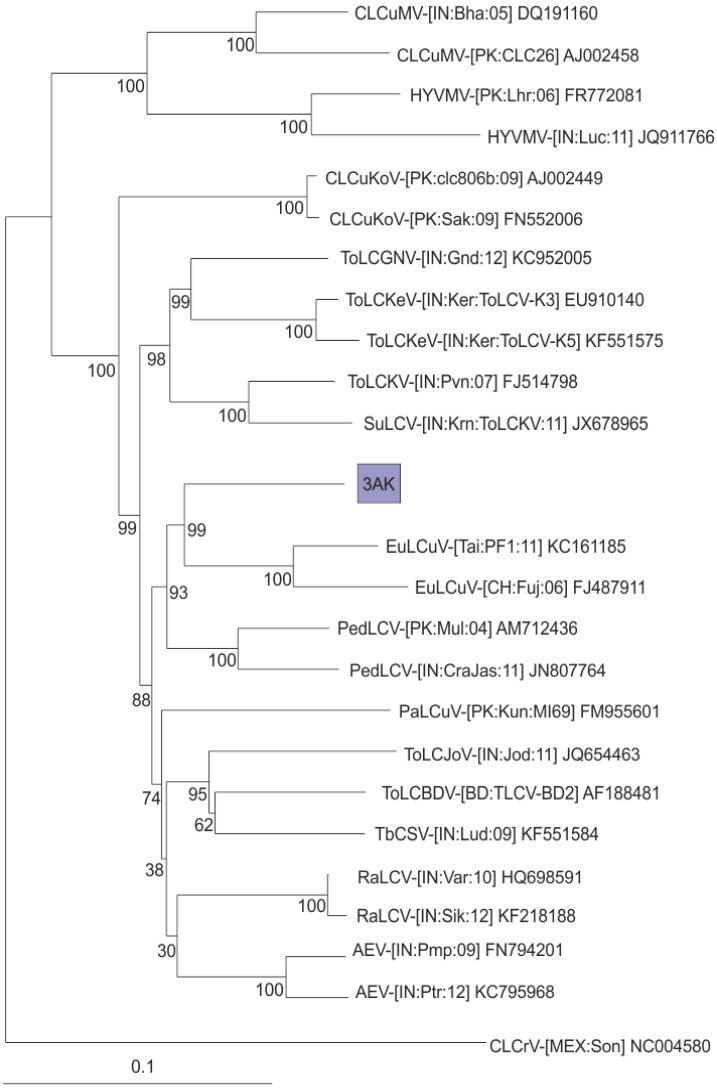

Figure 4.11: Phylogram built using Neighbour Joining algorithm showing isolate 3AK in cluster with closely related species. Euphorbia leaf curl virus and pedilanthus leaf curl virus are present in the same clade sharing common ancestors with isolate 3AK. Begomovirus isolate 3AK reported in this study is highlighted. The tree is rooted on *cotton leaf crumple* virus (CLCrV), a distantly related begomovirus. The numbers at the nodes represent bootstrap confidence values (1000 replicates). Viruses used are *Cotton leaf curl Multan virus* (CLCuMV), *Hollyhock yellow vein mosaic virus* (HYVMV), *Cotton leaf curl Kokran virus* (CLCuKV), *Euphorbia leaf curl virus* (EuLCuV), *Tomato leaf curl Pakistan virus* (ToLCuPKV), *Papaya laf curl virus* (PaLCuV), *Radish leaf curl virus* (RaLCuV), *Pedilanthus leaf curl virus* (PedLCV), *Ageratum enation virus* (AEV), *Tomato leaf curl Jodhepur virus* (ToLCJoV), *Tomato leaf curl Bangladesh virus* (ToLCBDV), *Tobacco curly shoot virus* (TbCSV), *Tomato leaf curl Karnataka virus* (ToLCKV), *Tomato leaf curl virus* (ToLCuV), *Sunflower leaf curl virus* (SuLCV),

Tomato leaf curl Gandhinagar virus (ToLCGNV) and *Tomato leaf curl Kerala virus* (ToLCKeV).

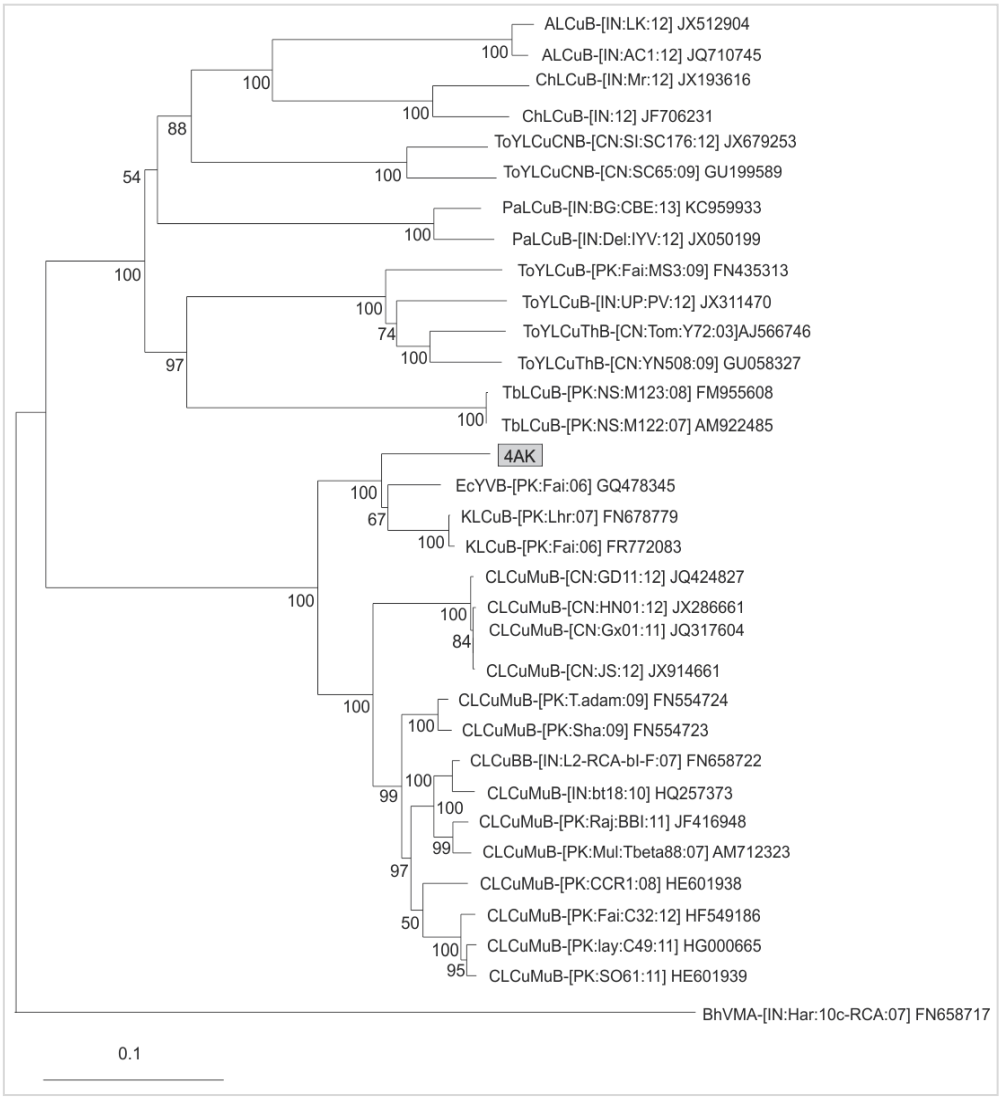

Figure 4.13: Phylogram showing betasatellite in cluster with closely related species. Betasatellite share same cluster with Eclipta yellow vein betasatellite and Kenaf leaf curl betasatellite. Betasatellite reported in this study is highlighted. Outgroup used is *Bhendi Yellow Vein Mosaic Alphasatellite* (BhYVMA) is segregated from other betasatellites. Values at the nodes represent the bootstrap values (1000 replicates). Viruses used are *Ageratum leaf curl beta* (ALCuB), *Chilli leaf curl beta* (ChLCuB), *Tomato yellow leaf curl China beta* (ToYLCuCNB), *Papaya leaf curl beta* (PaLCuB), *Tomato yellow leaf curl beta* (ToYLCuB), *Tomao yellow leaf curl Thailand beta* (ToYLCuThB), *Tobacco leaf curl beta* (TbLCuB), *Eclipta yellow vein betasatellite* (EcYVB), *Kenaf leaf curl betasatellite* (KLCuB), *Cotton leaf curl Multan betasatellite* (CLCuMB), *and Cotton leaf curl Burewala betasatellite* (CLCuBB).

4.1. Recombination Analysis via RDP4:

Recombination Determination Program 4 (RDP4) was used to confirm recombination events in isolate 3AK. Isolate 3AK sequence was found to harbor sequences at its start and end region from sequences from CLCuMV. Therefore, to confirm recombination, isolate 3AK sequence alongwith sequences of closely related begomoviruses PedLCV [DQ116884] and CLCMuV [AJ002459] were aligned using CLUSTALX. Alignment result was then used for RDP analysis.

RDP showed recombination at 1 point (Figure. 4.9). Recombination event was observed at the end of the sequence. Isolate 3AK having beginning breakpoint at 2386 nucleotide position and ending breakpoint position was at nucleotide 30. It was observed that PedLCV contributed as a major parent while fragment inserted at above mentioned position was of CLCMuV and hence contributed as a minor parent in recombination. Therefore, it was concluded that isolate 3AK is a recombinant, comprising sequences from two viruses (PedLCV and CLCMuV).

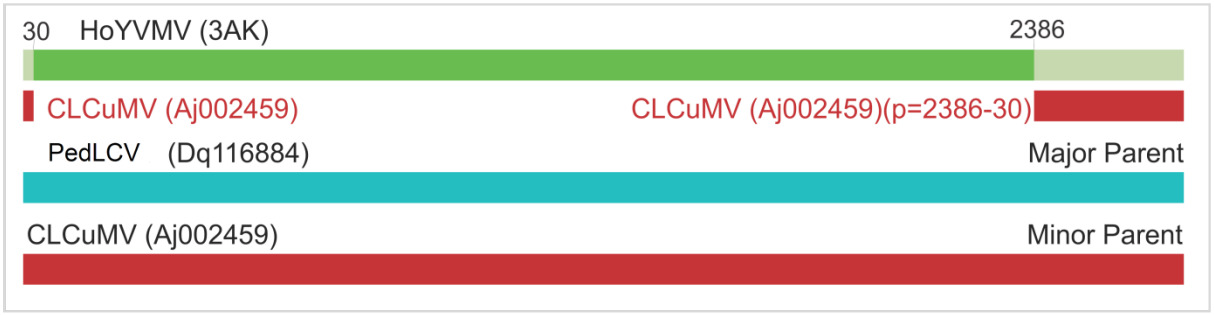

Figure 4.9: Recombination determination via RDP4. Analysis showed that isolate 3AK has single recombination event at the start and the end of the sequence harboring sequence from CLCuMV, contributing as a minor parent, at positions 2386-30 nucleotide position. Rest of the sequence is similar to that of major parent PedLCV.

4.2. Similarity Plot Analysis:

Similarity plot (Simplot) is a software tool for recombination analysis. Isolate 3AK was taken as a query and possible recombination was checked with closely related begomovirus sequences of PedLCV (DQ116884), Cotton leaf curl Multan virus-Okra (AJ002459) and Cotton leaf curl Multan virus-26 (AJ002458). Analysis showed that in the start there is a gap in the plot which means 100 nucleotides from the start of isolate 3AK are dissimilar and does not match with any of the three sequences. Straight blue line shows that PedLCV is 100% similar to our query sequence of isolate 3AK, while the two sequences from CLCuMV (overlapping green and red lines) are different from our query (Figure. 4.10). The 2 viruses PedLCV and CLCMV completely vary from each other.

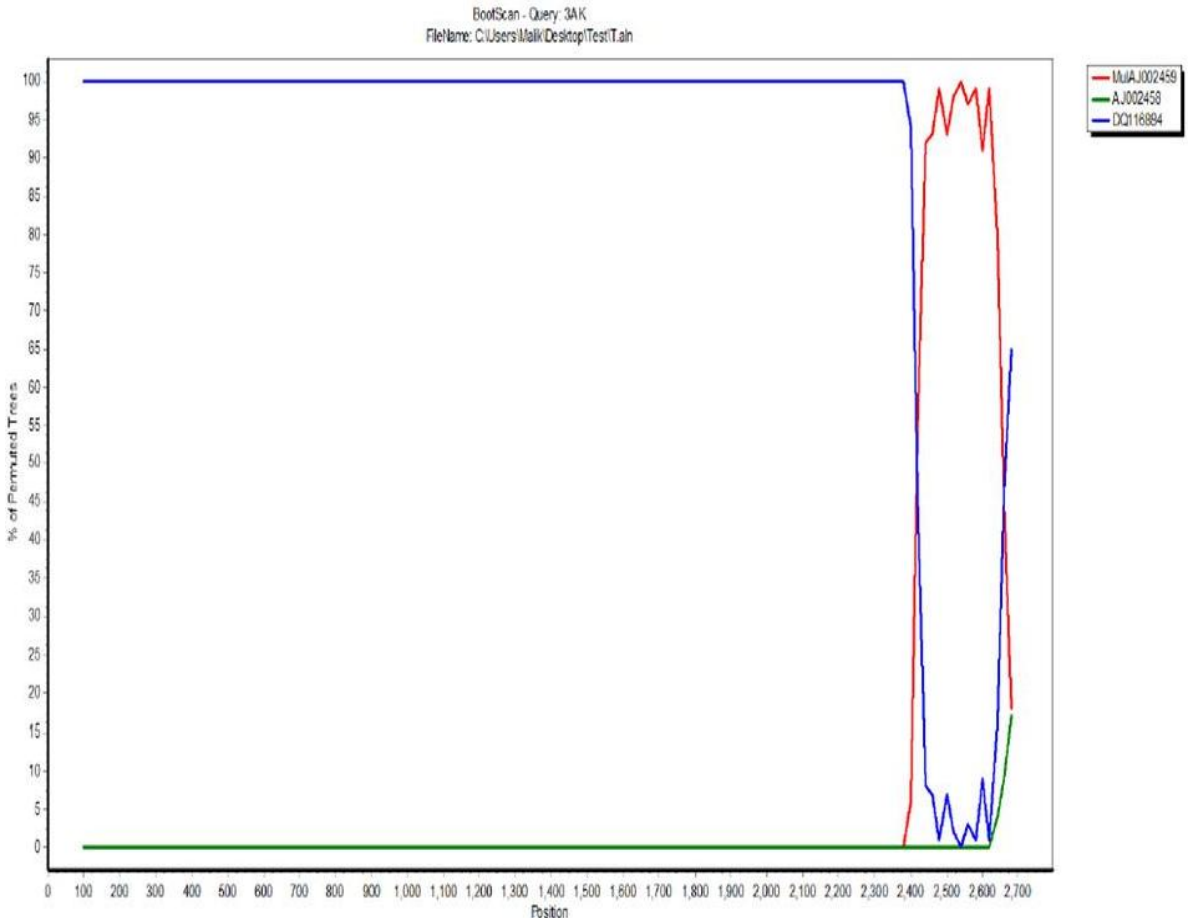

Figure 4.10: Simplot analysis showing exchange of fragments. Blue bar representing PedLCV is similar to the query sequence of isolate 3AK, is exchanging fragment at position 2360 with CLCuMV shown in red line, confirming recombination event in isolate 3AK

Chapter 5

Discussion

Begomoviruses have been infecting many crops, vegetables, ornamental plants and causing destruction at serious levels (Tahir *et al.,* 2010; Shahid *et al.,* 2007; Hussain *et al.,* 2011). In Pakistan begomoviruses are most destructive to cotton crop, as the losses estimated were upto 20% in 2010 due to cotton leaf curl disease (CLCuD). It has been observed that bipartite begomoviruses are also responsible for the emergence of new diseases in plant hosts. Satellite DNA molecules are responsible for varying symptoms in infected plants. On the other side there are increasing reports of emergence of new species of begomoviruses which suggest that this virus has a high mutation rate, infecting a wide range of host plants other than just cotton.

Infected Hollyhock samples were collected having vein yellowing and leaf crinkling symptoms. Hollyhock vein yellowing sample named as isolate (3AK) exhibited Hollyhock plant infected with a complex of begomovirus and a betasatellite. DNA sequencing showed virus to be of 2741 base pairs having maximum percentage identity with *Pedilanthus leaf curl virus* and also with *Euphorbia leaf curl virus* (89%) determined via BLAST analysis.

This is the first report of EuLCV infecting Hollyhock. Previously HYVMV, HLCrV, HLCuV, ZuYMV and MaVClV have been reported to infect Hollyhock. For this reason Hollyhock can be considered as an alternative host for CLCuV during off season when cotton is not cultivated (Singh and Misra, (1971); Salam *et al.,* (1998); Bigarre *et al.,* (2001); Choi *et al.,* (2003); Menzel *et al.,* (2010).

To confirm the virus as a new species based on the species demarcation criteria (Fauquet *et al,* 2003), sequence distance matrix was built using MegAlign which showed maximum of 86.6% identity with *Euphorbia leaf curl virus,* and 86.4% with PedLCV confirming it as a new species for which the name *Hollyhock yellow vein mosaic Islamabad virus* (HYVMIV) was proposed.

Phylogenetic analysis, using Neighbour Joining and Maximum Parsimony algorithms, of the virus also clustered HYVMIV with PedLCV and EuLCV having high bootstrap values.

Alignment analysis of the virus showed variation in sequence at the end and the start of the sequence (from last ~350 to first 50 nucleotides). These nucleotides vary from that of PedLCV. When these nucleotides were aligned they showed similarity with CLCuMV, suggesting recombination in virus. Recombination analysis via RDP4 and Simplot proved recombination event between the two viruses (PedLCV and CLCuMV). Begomovirus, HLCrV, infection in hollyhock was first reported by (Biggare *et al,* 2001) in Egypt. The two viruses (isolate 3AK and HLCrV) when aligned showed only 78% identity proving that the two viruses are distinct from each other. Not only in begomoviruses, recombinations have been reported across vast group of Geminiviruses. Frequent recombination events in begomoviruses accelerate the process of evolution and give rise to new species (Garcia-Arenal *et al*., 2001). These recombinations are indication of mix infections, two or more viruses infecting the same plant at the same time. Therefore, we can say that mix infection occurred either in Hollyhock or earlier in some other plant before the recombinant virus infected Hollyhock.

Recombination in begomoviruses has been reported previously as well. PepLCuLV is a recombinant of ChiLCuV and PaLCuV (Tahir *et al*., 2010). In 2001 resistant varieties of cotton were destroyed by cotton leaf curl Burewala virus (CLCuBV) which is a

recombinant and has found to infect a wide host range other than cotton (Mansoor *et al.,* 2003). Recombination in begomoviruses plays a role in giving rise to new much more vigorous species. Hollyhock in Pakistan has been found infected in various regions at high levels. Involvement of CLCuMV in recombination presents serious threats to cotton scientists. A new recombinant and resistant specie of the virus can nullify all the efforts made in controlling the disease.

Betasatellite molecule was also cloned from isolate 3AK which imparts typical begomovirus symptoms to the plant, i-e, of vein yellowing. Sequencing of betasatellite clone showed it to be of 1352 base pairs. Having 90% identity with *Kenaf leaf curl beta*, infecting *H. cannabinus* from India (Paul *et al*., 2008), it was proved to be an isolate of *Kenaf leaf curl beta*. Betasatellite molecule is contributing towards the vein yellowing symptoms of the sample plant typical of begomoviruses (Saunders *et al.,* 2001; Bull *et al.,* 2004).

In phylogenetic analysis betasatellite clustered with EcYVB and KLCuB having close relevance with CLCuMB and CLCuBB which suggests its close relationship with CLCuD causing begomoviruses.

Since both C4 and βC1 are responsible for symptom induction, their sequence analysis revealed 64.66% sequence identity which may contribute towards similar function. Further analysis is required to investigate the similar functions performed by both the genes present on different DNA molecules.

It was evident from the current results obtained that hollyhock is being infected by a recombinant, new species of begomovirus. In Pakistan a sharp analysis is necessary to keep a record of begomoviruses infecting cotton, their alternative hosts and the emergence of new species as a consequence of recombination in order to overcome the CLCuD.

These results also proved that hollyhock surely is an alternative host of CLCuV affecting cotton production in Pakistan. Hence, all these conditions should be kept in mind while devising an effective strategy against CLCuD.