# Cloning, Sequencing and Phylogenetic Analysis of β-DNA of *Cotton Leaf Curl virus*

B3O
3047

By

**Muhammad Irfan**

**Department of Microbiology**
**Faculty of Biological Sciences**
**Quaid-I-Azam University**
**Islamabad**
**2011**

# Cloning, Sequencing and Phylogenetic Analysis of β-DNA of *Cotton Leaf Curl virus*

A thesis submitted in partial fulfillment of the requirements for the Degree of

**Master of Philosophy**

**in**

**Microbiology**

**By**

**Muhammad Irfan**

**Department of Microbiology**
**Faculty of Biological Sciences**
**Quaid-I-Azam University**
**Islamabad**
**2011**

# Cloning, Sequencing and Phylogenetic Analysis of β-DNA of *Cotton Leaf Curl virus*

A thesis submitted in partial fulfillment of the requirements for the Degree of

**Master of Philosophy**

**in**

**Microbiology**

**By**

**Muhammad Irfan**

**Department of Microbiology**
**Faculty of Biological Sciences**
**Quaid-I-Azam University**
**Islamabad**
**2011**

# DEDICATION

*Dedicated to*

*My Loving*

*Parents and Family*

*Muhammad Irfan*

# Declaration

The material and information contained in this thesis is my original work. I have not previously presented any part of this work elsewhere for any other degree.

*Muhammad Irfan*

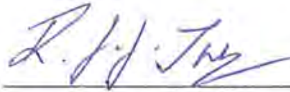
# Certificate

This thesis submitted by ***Muhammad Irfan*** is accepted in its present form by the Department of Microbiology, Quaid-I-Azam University, Islamabad, Pakistan; as satisfying the thesis requirements for the degree of Master of Philosophy in Microbiology.

**Supervisor:**

(Dr. Fariha Hasan)

**External Examiner:**



(Dr. Rani Faryal)

**Chairperson:**

(Dr. Fariha Hasan)

**Dated:** 25 Aug 2011

## List of Contents


| Sr. # | Title | Page # |
|---|---|---|
| 1. | List of Tables | i |
| 2. | List of Figures | ii |
| 3. | List of Acronyms | iii |
| 4. | Acknowledgements | v |
| 5. | Abstract | vii |
| 6. | Introduction | 1 |
| 7. | Aims and Objectives | 05 |
| 8. | Review of Literature | 06 |
| 9. | Material and Methods | 25 |
| 10. | Results | 30 |
| 11. | Discussion | 52 |
| 12. | Conclusions | 57 |
| 13. | References | 58 |
| 14. | Appendix | 68 |

# List of Tables


| Sr # | Titles | Page. No. |
|---|---|---|
| 1 | Genome, host range and vector for the genera of family *Geminiviradae* | 13 |

# List of Figures


| Sr # | Titles | Page # |
|---|---|---|
| 1 | Cotton growing area of Pakistan | 03 |
| 2 | Cryo electron microscopic image of *Maize streak virus* produce by Zhang *et al.*, 2001 | 08 |
| 3 | DNA Replication methods in *Begomovirus* | 11 |
| 4 | Gel image of DNA from cotton leaves | 31 |
| 5 | PCR amplification of cotton genomic DNA with actin gene primers | 31 |
| 6 | Summary of optimization experiment | 33 |
| 7 | PCR Amplification of cotton DNA at 60 °c | 35 |
| 8 | Growth of E.coli DH5 α competent cell on XGAL/IPTG LB Ampicillin agar plates | 36 |
| 9 | Plasmid gel image | 38 |
| 10 | Digested plasmid and clone with M13 | 39 |
| 11 | Purified plasmid gel image | 41 |
| 12 | Phylogenetic tree of all CLCuD associated *Begomovirus* (Evolutionary relationships of 103 taxa) | 44-45 |
| 13 | Phylogenetic tree of all CLCuV (Evolutionary relationships of 103 taxa) | 47-48 |
| 14 | NJ tree of CLCuV-PK-Bahawalnagar 2011-1 and CLCuV-PK-Bahawalnagar 2011-2 with all the available Pakistani Begomoviruses isolates | 50 |

# List of Acronym/abbreviations

| | |
|---|---|
| μL | Micorlitre |
| AD | Anno Domini |
| asRNA | Anti-sense RNA |
| AMP | Ampicillin |
| BC | Before Christ |
| BND | Benzoylated naphthoylated DEAE |
| $CaCl_2$ | Calcium chloride |
| CLCuD | Cotton leaf curl disease |
| CP | Coat protein |
| CR | Common region |
| CTAB | Cetyl trimethyl ammonium bromide |
| DEAE | Diethylaminoethyl cellulose |
| DNA | Deoxyribonucleic acid |
| DNAi | DNA interference |
| dNTP | Deoxyribonucleotide triphosphate |
| dsDNA | Double-stranded DNA |
| dsRNA | Double-stranded RNA |
| EDTA | Ethylene diamine tetraacetic acid |
| ESBL | Extended Spectrum β Lactamases |
| hpRNA | Hairpin RNA |
| HR | Hypersensitive response |
| ICTV | International Committee on Taxonomy of Viruses |
| IPTG | Isopropyl-beta-D-1-thiogalactopyranoside |
| IR | Intergenic region |
| $K_2HPO_4$ | Dipotassium phosphate |
| KCL | Potassium chloride |
| kDA | Kilo Dalton |
| kV | Kilo Volt |
| LB | Lauria broth |
| LIR | Large intergenic region |
| MCS | Multiple cloning site |
| mg | Milligram |
| $MgSO_4$ | Magnesium sulphate |
| $MgSO_4.7H_2O$ | Magnesium sulphate heptahydrate |
| miRNA | Microrna |
| mM | Millimolar |
| MP | Movement protein |
| mRNAs | Messenger RNA |
| NaCl | Sodium chloride |
| $NaH_2PO_4$ | Sodium phosphate |
| NaOH | Sodium hydroxide |
| ng | Nanogram |
| NH4Cl | Ammonium chloride |
| NLS | Nuclear localization signals |
| NSP | Nuclear shuttle protein |
| nt | Nucleotide |
| NW | New World |
| OD | Optical density |

| | |
|---|---|
| ORF | Open reading frame |
| OW | Old World |
| PCNA | Proliferating cell nuclear antigen |
| PCR | Polymerase chain reaction |
| PDR | Pathogen derived resistance |
| pH | Paviour of hydrogen |
| pre-miRNA | Precursor mirna |
| PVP | Polyvinyl pyrrolidone |
| RCA | Rolling circle amplification |
| RCR | Rolling circle replication |
| RDR | Recombination-dependent replication |
| RdRP | RNA dependent RNA polymerase |
| REn | Replication enhancer protein |
| Rep | Replication associated protein |
| RISC | RNA-induced silencing complex |
| RNA | Ribonucleic acid |
| RNAi | RNA interference |
| rpm | Revolutions per minute |
| SBS | School of Biological Sciences |
| SCR | Satellite-conserved region |
| SDS | Sodium dodecyl sulphate |
| SIR | Small intergenic region |
| siRNA | Small interfering RNA |
| SSC | Standard sodium citrate |
| SsDNA | Single-stranded DNA |
| TAE | Tris-acetate EDTA |
| *Taq* | *Thermus aquaticus* |
| ta-siRNA | Trans-acting sirnas |
| TrAP | Transcriptional activator protein |
| T-Rep | Truncated Rep |
| UV | Ultra violet |
| VIGS | Virus induced gene silencing |
| X-Gal | 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside |
| YMD | Yellow mosaic disease |

# ACKNOWLEDGEMENTS

*ALL Praises to Almighty Allah, the omnipotent, The most compassionate and His Prophet Muhammad (P.B.U.H) The Most perfect among all human beings ever born on the surface of earth, Who is forever a source of guidance and knowledge for humanity as a whole.*

*I feel great pleasure in expressing my ineffable thanks to my encouraging, inspirational and cool minded supervisor Dr. fariha Hassan, Assistant Professor, Department of Microbiology for his sense of devotion, creativity, affectionate criticism and keen interest in my work;*

*I tender my thanks to Professor Dr. Safia Ahmad, Chairpersan, Department of Microbiology, Faculty of Biological Sciences, QAU for providing me research facilities.*

*Words are inadequate to express my thanks to my co supervisor Dr. Muhammad Zeeshan Hyder Assistant Professor whose scholastic criticism made it possible for me to complete this hard job quite smartly. I am indeed humbly grateful to him for his kind attitude, valuable suggestions, most cooperative affectionate behavior, impetuous guidance and moral help. It was because of his inspiring guidance and dynamic cooperation during entire study program that I could complete this manuscript.*

*I feel great pleasure in expressing sincerest thanks to my class fellows Muhammad rafiq, Zia Ullah Khan, Fida Hussain, Imran Khan, Abdur Rahman and Muhammad Irfan Alam.*

*I would like to express my deep sense of gratitude, appreciation and sincere thanks to all of my Friends Asad Zia, Ikram ullh, Sahib Zada Yousafzai, Akhter Nazman, Amanullah, Fakhrul Hassan Taqweem ul Haq, Siddiq Akbar, Mati Ullah Khatak, ,*

*I would like to express my appreciation to my ex-class fellows and friends Muhammad Ayub, Muhammad Ismail, Iqbal Husain, Musghar Suhail, Sajid Mateen, Sadia Babar and my friend Sahar Naz.*

*In this stage how can I forget my sisters and brothers without their prayer I could not do this tough work.*

*Words do not come out easy for me to mention the feelings of obligations towards my affectionate parents. My father Fazal Ali proved the ocean of Love, care for me in which I saturated myself. In life he is the man who told me what life is, how to pass it, how to go through the problems, etiquettes and the every aspect of my life is incomplete without him. I am most earnestly appreciative to My Mother for the strenuous efforts done by her in enabling me to join the higher ideals of life and also for her financial and moral support, patience and prayers they had made for my success. May ALLAH ALMIGHTY infuse me with the energy to fulfill their noble inspirations and expectations and further edify my competence.*

**Muhammad Irfan**

# *Abstract*

# ABSTRACT

Viruses are very successful intracellular parasite infecting almost every life on this planet. Plants are also the hosts for many viruses including them are circular single stranded DNA (cssDNA) viruses, which have been classified in two different families i.e. *Geminiviridae* and *Nanoviridae*, both infecting plants and have cssDNA multicomponent genome. *Geminiviridae* is a very diverse and include many viruses which infect many host plant of agronomical importance worldwide. Cotton leaf curl disease (CLCD) is one of the most important disease economically which has resulted in huge losses in many epidemics in cotton growing countries around the world. The causative agents of this disease is not yet conclusively determined but so far about nice distinct bipartite begomoviruses are found associated with this disease. The genome of these viruses are composed of DNA-A which contains genes for DNA replication, coat protein and regulation of gene expression and associated DNA-β which contains gens for virus movement and plays role in virus pathogenicity.

Cotton is an important cash crop of Pakistan which plays important role in its economy. Many epidemics of CLCD have been reported which have resulted in huge economical loses despite the introduction of resistant varieties. Much of the scientific attention remained on understanding of DNA-A which are associated with CLCD and DNA-β remained poorly understood despite its role in virus movement and pathogenicity. In current investigation a PCR assay was developed and optimized to amplify DNA- from infected cotton plants along with actin gene as positive control from nuclear genome of cotton plants. The primer pair CLCuV DNA-β successfully amplified this component from infected plants originated from Bahawalnagar; a cotton producing area in Southern Punjab Pakistan. These products from two isolates were cloned and sequenced and sequence data was used for phylogenetic analysis to understand the evolution of DNA-β in Pakistan.

The phylogenetic analysis of DNA-β sequenced in this study along with sequences of DNA-β publically available in GenBank till March 2011 using NJ method revealed two major groups where isolates from Indian subcontinent and African continent grouped together in Indo-African group while other isolates from Indian subcontinent and China where grouped in Indo-China group. These groups indicate two distinct

Cotton is one of the most important cash crop of Pakistan. Cotton is a shrubby plant belongs to family Malvaceae and genus Gossypium. It has unique economic value all around the world. Cotton plants are sensitive to many environmental hazards, pests and diseases. Virus diseases of cotton have historically been of only sporadic importance to global cotton production. Worldwide Pakistan has a key role in cotton industry on behalf of its production (4th number in world), exporting of cotton in raw form (3rd) also a huge quantity of cotton is consume in Pakistan and stay as largest 4th consumer in world as well. Strength of Pakistan cotton industry is the exporting of fine yarn which leads the world cotton yarn.

Worldwide cotton is produce in large amount other than all other nonfood crops. Cotton is a multipurpose crop which provides 5 principle product including lint, food, oil, seed hulls and linters. The chief use of cotton is in Textile industry. Lint is use to achieve a high quality fiber from the textile industries.

The history shows cotton cultivation and use from very past like in the time of Mohenjodaro civilization. The seeds of cotton were also discovered in Mehr Garh Baluchistan which were from about 5th millennium BC (Ahmad *et al*, 1999). Some other reports show that the cotton growth as domesticated crop started from Africa, Mexico and Peru. Now a day more than 3.4 billion hectares area is cultivated (occupy about 2.5 % of the cultivated area across the world) on cotton crop and more than 25 billion people are associated with cotton industry accounting for about 7% of the labor in the arising world. Worldwide China and USA is the leading cotton producer followed by India, Pakistan and Uzbekistan. The economy of most growing nations largely depends on cotton production and export.

Oil is also extract from the seeds of cotton by crushing which is used for cooking, the seed covering called "hulls" is used for animal feed. Beside these the lint around the cotton seeds is use in plastic and explosive products formation.

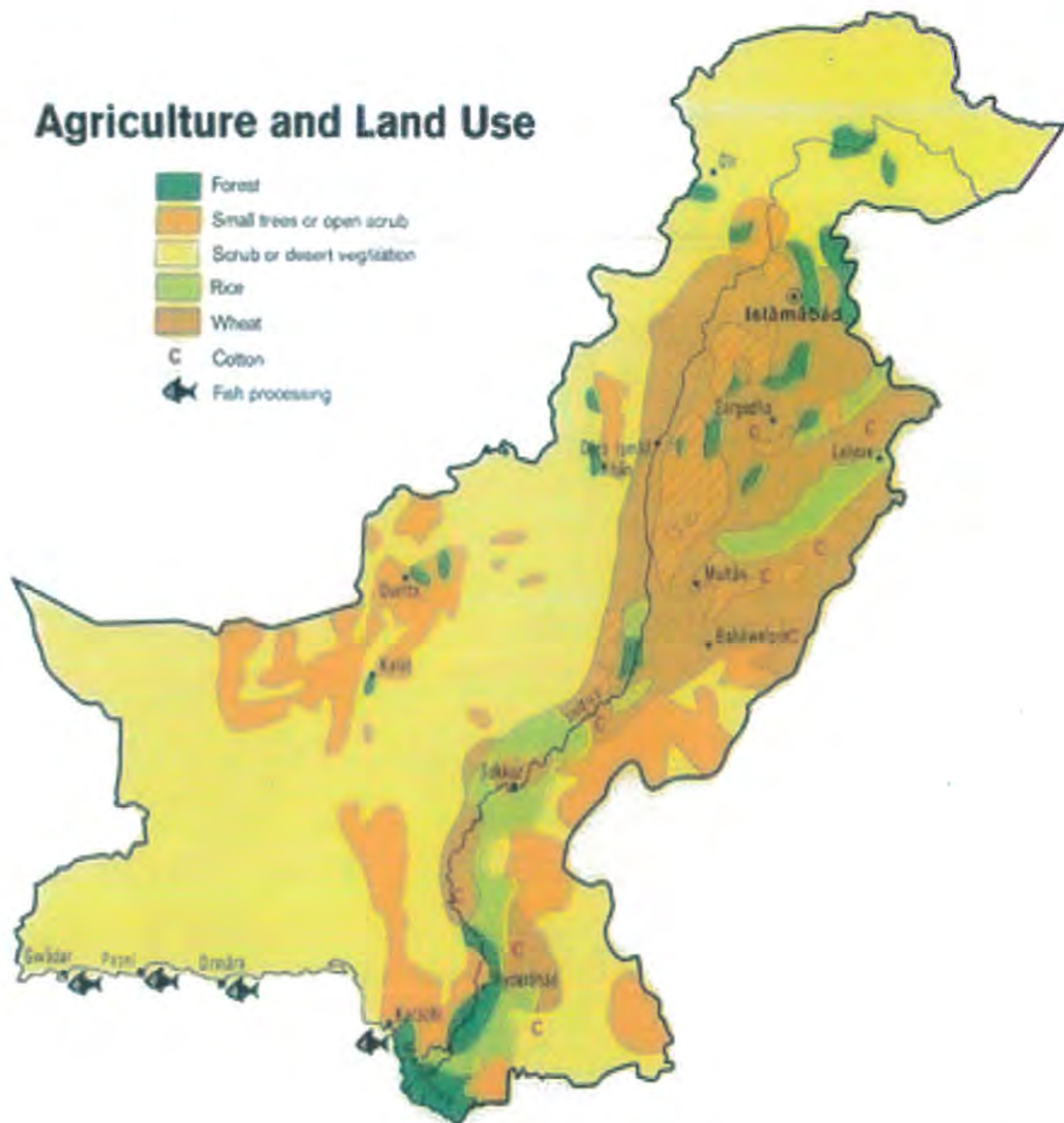



**Fig 1.1: Cotton growing areas of Pakistan** (Adopted from http://rehmananwar.blogspot.com/2010/03/agriculture-in-pakistan.html).

About 1.3 million farmers cultivate cotton on about 3 million acre area of Pakistan. So it provides opportunity to 75% farmers of Pakistan which cultivate about 15% of the cultivable area on cotton crop. Cotton's contribution in GDP is almost about 10% and contributes up to 57% of foreign exchange earning of Pakistan (Ahmad et al., 2010). About 30 – 40% is use domestically and the remaining is export in different form like

raw cotton, garments, cloth and yarn.

For several decades, cotton importance has increased in Pakistan economy .the most significant increase was that in early 1980, and 1990. The Punjab Province, with 2.5 million hectares under cotton, grows approximately three times the area grown in the Sindh to the south. The yield in the Punjab is typically twice that in the Sindh. However first time a considerable decrease in cotton production was observed due to the CLCuV in 1967 (Hussain and Ali, 1975). Cotton leaf curl virus is first time reported from Multan region in 1967 and consistently increase with time and the production is decrease with time to time. In 1993 a notable decrease in cotton production from 12.8 million bales in 1992 to 8 million bales (Aslam et al., 2000).

Geminiviridae genomes consist of single stranded DNA. It can infect dicotyledonous plants. Geminivirus have four genera based on genome organization, host range and insect vector. The four genera are *Mastrevirus, Begomovirus, Curtovirus* and *Topocuvirus*. Geminivirus is the second largest family among the plant pathogen viruses having 133 species. Begomovirus is the largest among other genera of the family geminiviruses comprising 117 species transmitted by whitefly *Bemisia tabaci* (Stanley *et al.*, 2005). Begomoviruses are transmitted by single species of whitefly typically the genome consists of two components DAN A and DNA B. The DNA A encodes all viral function required for viral replication, control of gene expression and insect transmission. DNA B code two gens necessary for viral movement within the plant.however a small number of monopartite begomoviruses have been identified which lack the B component. Satellite are viruses or nucleic acid that depend on a helper virus for their replication but do not have extensive nucleotide sequence similarity to their helper virus and are dispensable for its proliferation (Murant and Mayo,1982). DNA β are symptom modulating, single stranded DNA satellites associated with monopartite begomoviruses). DNA β is responsible for movement in plant, insect transmission. The function of DNA β in the development of infection remains unclear but responsible for movement of virus creating a cellular environment for viral replication and may overcoming host defense mechanism (Briddon *et al.*, 2003).

The DNA Beta codes for protein have role in the spread of infection. How the DNA beta is originated it is unclear till now (Briddon *et al*., 2010). The Beta component composed of 2 regions Virion-Sense Strand and Complementary Sense Strand and each encode a single protein. **BV1** is present on Virion Sense Strand also term as BR1 gene. BV1 gene carries the genetic information for the synthesis of Nuclear Shuttle Protein NSP. **BC1** also called as BL1 gene and encode for movement protein MPB respectively (Stanley et al., 2005; Seal et al., 2006).

The first satellite was found to be associated with *Tomato leaf curl virus* from Australia (Dry *et al.,* 1997). This 682 nt DNA depends on ToLCV for its replication and encapsidation but other geminiviruses can also support its replication. DNA β satellites are required for the induction of disease symptom in some host plants (Briddon *et al*., 2001).

The DNA β satellite are 1350 nt long, and it require a helper virus for their replication and encapsidation (Mansoor *et al*., 2003). DNA β satellite is discovered recently but it has existed for many centuries. The DNA β satellite associated with *Eupatorium yellow vein virus* (EupYVV) has associated with disease symptom described about 1250 year ago (Saundesrs *et al*., 2003).

## Aims and objective:

- Development and optimization of PCR for the amplification of CLCuV DNA-β.
- Sequencing and phylogenetic analysis of CLCuV DNA-β from Pakistan

Chapter 2

# Literature review

Viruses are controversial entities between living and nonliving organisms. These are obligate intracellular parasite that live and multiply in an appropriate host. The first plant virus was discovered by MartinusBeijerinck in 1898, the virus was coined "*Tobacco mosaic virus*". The most acceptable definition of viruses was done by Luria in 1978. According to Luria viruses are "entities whose genomes are elements of nucleic acid that replicate in living cells using cellular synthetic machinery and causing the synthesis of specialized elements that can transfer the viral genome to other cells" (Luria *et al.*, 1978). The virus may produce diseases in any living organism from unicellular like bacteria to multicellular organism having well sophisticated immune systems. Those viruses are of great interests which cause diseases in human and plants. Plant viruses are important because plants are essential for any form of life on this earth crust.

ICTV in its $8^{th}$ report approved 3 orders, 73 families, subfamilies 9, genera 287 and round about 1950 species of virus. (ICTV report 2005). Among these the plant pathogenic viruses have 20 families, 88 genera and about 750 species discovered up to now. The genome plant viruses composed of dsDNA, ssDNA, dsRNA and ssRNA. Most of these plant disease causing viruses are ssRNA.

## 1.1.Introduction to Family *Geminiviradae*:

The first statement about the *Geminiviradae* may be before 750AD, from current Japan area from the local poetry of the Man'yoshu (Japan) about damaging the *Eupatorium* plants which appears yellow in summer, this was linked with the yellow vein disease perennial plant (Inouye and Osaki, 1980). In 2003 it was determined that the causative agent that caused disease in *Eupatorium* was due to members of *Geminiviradae* or *Geminivirus* related beta component. The virus was named as *Eupatorium yellow vein virus (EpYVV)* (Saunders *et al.*, 2003). The Gemini viruses have worldwide importance because it can cause infection of many significant cash and food crops. *Geminiviruses* are tissue specific and mostly have access to phloem. This family causes many economically important viral disease in plants (Moffat, 1999), some important disease of *Geminiviruses* are *Maize streak virus* (Bosque-perez, 2000) and those that infect cassava (Legg and

Fauquet, 2004), tomato (Moriones and Navascastillo, 2000) and cotton (Briddon, 2003; Bridon and Markhan, 2000). This family is highly infectious and damages a wide range of crops. Due to this economical loss it got great attention worldwide therefore it has extensively studied to step down the rate of causing diseases and to improve the level of cash and food crop production (Buck and Coutts, 1983).

### 2.1.1 Morphology:

The electron microscopy of sap from the Maize streak and beet curly top diseases revealed a geminate or twinned quasi isometric particle (Bock *et al.*, 1974; Mumford, 1974). Due to this characteristic shape "GEMINI" the name of the virus family coined as *Geminiviradae* (Harrison *et al.*, 1977). This structural feature the viruses belonging to this family amalgamating the whole family members (Fig; 2.1). The size of *Geminivirus*virion is 22nm in diameter and 38nm long composed of two incomplete icosahedra (T=1) in which 110 coat protein (CP) are present in the form 22 pentameric capsomers (Zhang *et al.*, 2001).

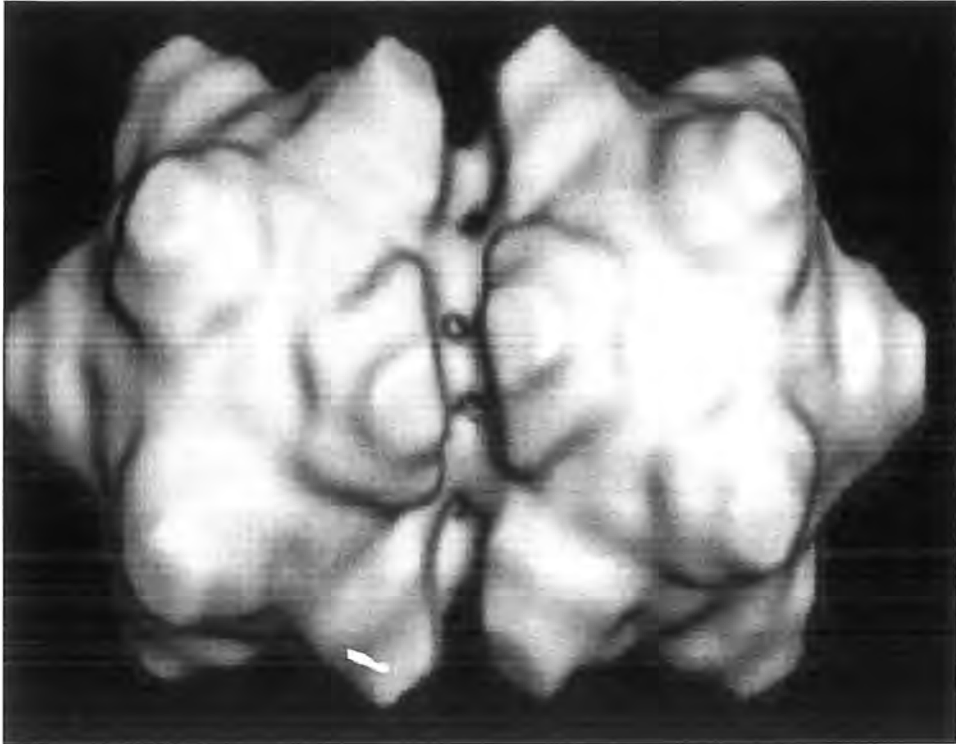


Figure 2.1 Cryo-electron microscopic image of *Maize streak virus* produce by (Zhang *et al.*, 2001.)

### 2.1.2 Nucleic Acid:

The genome of *Geminiviradae* viruses are composed of ssDNA circular molecule encapsidated in the geminate capsid. The geminate particle have a copy of Single Stranded DNA (ssDNA) having size of 2.5-3.0bp. Some specie have single piece of ssDNA called monopartite genome e.g. *Mastrevirus,* while some particle have an additional component of ssDNA called bipartite genome e.g. most of the *Begomovirus* have DNA-A and DNA-B (Hamilton *et al., 1983;* Hamilton *et al., 1984;* Lazarowitz and Lazdins, 1991).

### 2.1.3 Other Features:

The only reported structural protein associated with *Geminivirus*virion is CP. The size of CP is approximately 28-34 kDa. There is no any other protein reported till now with the virion. The presence of carbohydrate and lipid is also not reported (Fauquet *et al book*).

### 2.1.4 Genome Organization:

The genome of *Geminiviradae* genera is quite dissimilar to each other. Some genera contain a single component of ssDNA i.e. *Mastrevirus, Topocuvirus, Curtovirus* and few species of *Begomovirus.* While species exhibit the genome in two components of ssDNA i.e. DNA A and DNA B. The genome of *Mastrevirus* is about 2.6-2.8kb and encoded for four different proteins (Wright *et al* 1997). The CSS codes for two proteins Rep and Rep A, which play main role in the replication. The ORF Rep A and Rep B after splicing encoded Rep protein while the Rep A protein is encoded by ORF called Rep A. The VSS encodes for late infection proteins including Movement protein (MP) and CP. (Wright *et al.,* 1997)

### 2.1.5 Replication Cycle

*Geminiviruses* carried out their replication in the nucleus of host cell through dsDNA intermediate (Saunders *et al.,* 1991; Stenger *et al.,* 1991). The *Geminiviruses* follows the Rolling Circle Replication (RCR) and Recombination Dependent Replication (RDR) (Jeske *et al.,* 2001). In RCR mechanism first of all the ssDNA is converted into a dsDNA intermediate. All geminivirus encode a Rep protein which is essential for geminivirus DNA replication. This protein is unique for the initiation and termination of RCR activity when the viral DNA entered into the nucleus of the infected host cells (Laufs *et al.,* 1995a: Stenger *et al.,* 1991). The Rep protein is unique having no similarity with any known DNA polymerase (Laufs *et al.,* 1995c). There are significant 4 amino acid sequence motifs in different *Geminivirus* Rep protein. The function of motif 1 (FLTY) is yet to be clear while motif 11(HLH) is involved in metal ion concentration (Laufs *et* al., 1995c). Replication specificity of *Geminivirus* DNA is based on the Rep protein. RCR mechanism of replication was observed in both DNA A and DNA B. DNA B can't

replicate independently it needs the presence of DNA A for their replication while DNA A can replicate without DNA B (Yadava *et al.,* ).Rep Protein can start rolling circle replication by stand specific neck. RCR mostly completed by two steps firstly through secondary intermediate and by RCR Analogue process (Bisaro, 1996; Noris *et al.*, 1996; Saunders *et al.*, 1991; Stenger *et al.*, 1991; Timmermans *et al.*, 1994). Iteron found In Common Region (Hanley- Bowdoin *et al.,* 1999) is responsible for initiation of replication. Rep is covalently attached to 5'-terminus (Laufs *et al.,* 1995b). Rep and origin of replication (ori) of begomoviruses are usually not compatible. DNA polymerases synthesize the virion strand. It is cut and ligated to form a circular single stranded virion DNA. This circular ssDNA can either serve as template for replication or can be encapsulated into virion. The majority of begomoviruses are compatible with a recombination-dependent replication (RDR) mechanism (Jeske *et al.,* 2001). The RDR pathway is similar to the replication pathway of T4 bacteriophage, it is also terms as "join copy" pathway, "bubble migration synthesis" and "break induced replication" (Kreuzer, 2000: Mosig, 1998: Formosa and Alberts, 1986 and George and Kreuzer, 1996).

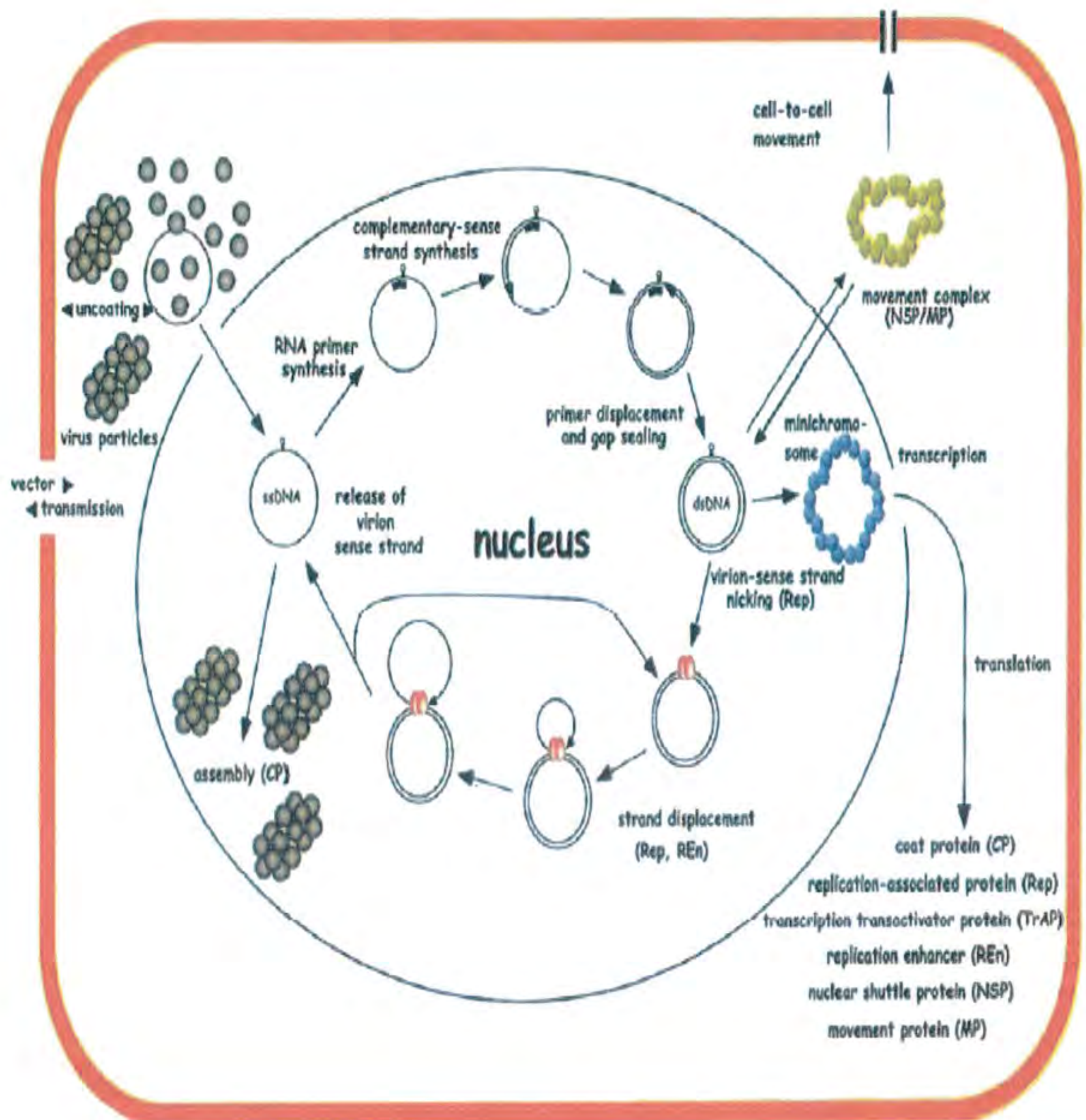



**Fig 2.2 DNA Replication methods in Begomoviruse**

Virus particle are introduced into the cell during whitefly feeding and the ssDNA is replicated in the nucleus.adsDNA intermediate is synthesized.ssDNA is synthesized by rolling circle mechanism (Stanley *et al.*,2005).

The RDR pathway is completed in three steps including

1. dsDNA was unwound to produce ssDNA which act as a 3' end.
2. Formation of "displacement loop" also called as D-loop or bubble loop in the DNA. The 3' end ssDNA act as primer for the replication process of DNA.
3. In this step the DNA heteroduplex extension is formed. Here the polymerase carry out the elongation of the DNA (Kreuzer, 2000: Mosig, 1998).

Different *Geminiviruses* were examined using benzoylatednaphthoylated DEAE (BND) cellulose chromatography in combination with improved high resolution two dimensional gel electrophoresis and was concluded that multitasking (utilization of both RCR and RDR mechanisms) in replication is diffused among *Geminiviruses* (Preiss and Jeske, 2003).

## 2.2 Classification of *Geminiviradae*:

The *Geminiviradae* is classified in to four different genera on the basis of their genome organization, biological properties, vector type by which it transfer and host range. The family *Geminiviradae* is further classified into four different genera on the basis of vector used for transmission of virus, biological properties, genome organization and host range. These four genera are *Begomovirus, Curtovirus, Mastrevirus* and *Topocuvirus* (Fontes *et al*., 1994; Rybicki *et al*., 2000; Fauquet, 2003; Stainly, 2005).

| Genera | Genome | Target plants | Vector |
|---|---|---|---|
| ***Begomovirus*** | Bipartite/Monopartite | Monocotyledonous | Whitefly (*Bemisiatabaci*) |
| ***Curtovirus*** | Monopartite | Dicotyledonous | Leafhopper |
| ***Mastrevirus*** | Monopartite | Monocotyledonous | Leafhopper |
| ***Topocuvirus*** | Monopartite | Dicotyledonous | Treehopper |

**Table; 2.1 Genome, host range and vector for the genera of family *Geminiviradae***

### 2.2.1 Genus *Mastrevirus*:

The main features of the genera *Mastrevirus* are the monopartite genome, the vector used for transmission is leafhopper and infect the monocotyledonous as well as Dicotyledonous plants (Boulton, 2002). The example of species infect the dicotyledonous plants are *Wheat dwarf virus* (WDV) and MSV, while the *Bean yellow dwarf virus* (BeYDV) and *Tobacco yellow dwarf virus* (TbYDV) are examples of dicotyledonous infecting species of the genera. The genome of *Mastrevirus* is composed of 2.6-2.8kbp which carried information for the coding of four proteins (Wright *et al* 1997). The ssDNA genome is divided into two regions that VSS and CSS each have having two ORFs. The CSS encodes for two proteins i.e. Rep and Rep A. the Rep A is encoded by Rep A ORF, while the Rep protein is the product of Rep A and Rep B ORFs, after splicing. The type species of *Mastrevirus* are *Wheat dwarf virus* (WDV) and *Tobacco yellow dwarf virus* (TbYDV).

### 2.2.2 Genus *Curtovirus*:

The *Curtovirus* exhibit the monopartite genome having tendency to infect the monocot plants. The vector used in transmission of *curtovirus* is leafhopper. The genome of *Curtovirus* is about 3.0kbp in ssDNA form (Hur *et al* 2007). There are seven genes reported in the genome of *Curtovirus* having an intergenic region (IR) which have the

origin for DNA replication (Baliji *et al.,* 2004; Briddon *et al.,* 1998; Klute*et al.,* 1996; Stanley *et al.* 1986a; Stenger, 1994a). *Spanish top curly virus and Beet top curly virus* are the type species of *Curtovirus* (Baliji *et al.,* 2004).

### 2.2.3 Genus *Topocuvirus*:

The genome of the genus *Topocuvirus*is monopartite in nature and it is the latest established genus of family *Geminiviradae.* The *Topocuvirus* can infect the dicotyledonous plants and the vector for transmission is treehopper *Micrutalismalleiffera.* The genome of TPCTV is consisting of two parts i.e. VSS and CSS. The VSS encodes for two proteins CP and V2 proteins, while the CSS carried out genetic information for four proteins i.e. Rep, C2, C3 and C4. The genome has an IR which has a hair pin like structure with Nona nucleotide TAATATTAC. Type specie of *topocuvirus* is *pepper yellow dwarf virus* (Lam *et al.,* 2009).

### 2.2.4 Genus *Begomovirus*:

The genus *Begomovirus* is the largest family causing diseases of important plants, cash crop and cereal/food crop. The name *Begomovirus* is derived from its type specie *Bean golden yellow mosaic virus* (BGYMV). The vector for the *Begomovirses* is a whitefly *Bemisiatabaci*Genn. The main reason of worldwide spread of *Begomoviruses* infections is the circulation of plant product worldwide and almost ubiquitous nature of their vectors (Rojas *et al.,* 2003; Seal *et al.,* 2005). The worldwide incidence of *Begomovirus* infection is also due the composite and transferable stretch of DNA (Sandrous *et al.,* 2001; Varma and Malathi, 2003; Bull *et al.,* 2004; Stanley *et al.,* 2004). The continuously addition of new species to the body of *Begomovirus* genus (Bull *et al.,* 2006) and high degree of genomic variability is due the high rate of mutation within the species of *Begomoviruses* in short time span (Gilbertson *et al.,* 1991; Stenger *et al.,* 1995; Ooi *et al.,* 1997; Sanz *et al.,* 1999; Duffy and Holmes, 2008). The *Begomovirus* is generally classified into two categories on the basis of phylogenetic characteristics and genomic arrangement. Those viruses which are found in the Eastern hemisphere like Europe, Africa, Asia and Australia region are called Old World (OW) Viruses. The viruses found in the western

hemisphere like in America are term as New World (NW) Viruses (Rybicki, 1994; Padidam *et al.*, 1999; Paximadous *et al.*, 1999). The main difference between OW and NW *Begomoviruses* are the presence of bipartite genome in the NW Virus while the OW Viruses have both bipartite and monopartite genome viruses. The genome of *Begomovirus* is composed of two components (bipartite) as well as of single component (monopartite). The transmission of *Begomovirus* species is facilitated by whitefly (*Bemisiatabaci*) and infects the dicotyledonous plants almost over the world (Pringle, 1999; Fauquet, 2008). The worldwide occurrences of the *Begomovirus* infections are due to the ubiquitous nature of the vector.

### 2.3. Genome Organization:

Bipartite species of family *Geminivirus*, genera *Begomovirus* have eight ORFs (Open Reading Frames) in their genome, while the monopartite species have six ORFs in genome. DNA A and DNA B have two parts Virion Sense (V) or Complementary sense (C) Strands. The VSS of DNA A encodes for four proteins and CSS encodes for two proteins. The DNA B codes for protein have role in the spread of infection. How the B DNA is originated it is unclear till now (Sanderfoot *et al.*, 1996). The B component also composed of 2 regions Virion-Sense Strand and Complementary Sense Strand and each encode a single protein. A sequence of about 200 nucleotides is shared between the both components. There is a common region "CR" between the two components of bipartite genomes i.e. CRA and CRB having the conserved sequence TAATATTAC located in either A or B component. The CR has more than 85% similarity (Bowdoin *et al.*, 1999; Astorga and Medrano, 2001). The type specie of *Begomovirus* is *Cotton leaf curl virus*. Most of the bagomoviradae specie is composed of bipartite (having 2 parts) both of DNA A and DNA B component, while some isolates consist of monopartite genome. Most of specie required 2$^{nd}$ component of DNA for the infection while some specie has the ability to infect the host with single part (Fauquet et al, 2003). There is an intergenic region between DNA-A and B term as Common Region (CR) which is highly conserved having about 200 nucleotide base (Lazarowitz, 1992). The origin of replication (ori) is bedded here in this GC rich region (Fontes *et al.*, 1994) having some constant repeats TAATATTA sequence which forms the hair pin like structure. Its main function is in

cleaving and joining of DNA molecule during replication process. The virus specific region required for replication is present at the upstream region of *ori* (hair pin like structure) which is particularly identify by the Rep and binds to it which leads to the replication process (Fontes *et al.*, 1994; Laufs *et al.*, 1995). Besides this the components may exchange some portion of genome by a process called pseudo-recombination or "Regulon Grafting". In regulon grafting the DNA A gives its part of CR region to DNA B. (Briddon *et al.*, 2010).

**DNA-A:**

DNA-A codes for protein required for capsid formation, replication, vector transmission, control of gene expression and overcoming host defense (Briddon *et al.*, 2010). It is further divided into 2 regions that are Virion Sense Strand exhibit 2 genes and Complementary Sense Strand has 4 genes. Six genes are present on DNA A. component which encodes different essential viral protein required for replication. The proteins encoded by the Virion Sense Strand are:

**AV1** gene is also known as AR 1 and the product of this gene is coat protein (CP). **AV2** is also named as AR2 and the product of this gene is movement protein (MP). *Note: The* $AV_2$ region is absent in the *begomovirus* of the new world (Ribicky, 1994; Stanley *et al.*, 2005). The Complimentary Sense Strand encodes 4 different genes these are: **AC1** also denoted by AL1 and coded for a protein that have a main role in the replication protein (Rep). **AC2** is known as AR2 and encodes for transcriptional activator protein (TrAP). **AC3** also called AL3 code for protein known as Replication enhancer protein (REn). This protein is involved in the assembly of viral particles in plants. **AC4** Also called AL4 and codes for AC4 protein.

**DNA β:**

The DNA Beta codes for protein have role in the spread of infection. How the DNA beta is originated it is unclear till now (Briddon *et al.*, 2010).The Beta component also composed of 2 regions Virion-Sense Strand and Complementary Sense Strand and each encode a single protein. **BV1** is present on Virion Sense Strand also term as BR1 gene.

BV1 gene carries the genetic information for the synthesis of Nuclear Shuttle Protein NSP. **BC1** also called as BL1 gene and encode for movement protein MPB respectively (Stanley *et al.*, 2005; Seal *et al.*, 2006).

## 2.4. Proteins encoded by *Begomoviruses*:

Genome of *Begomovirus* carry information for many proteins required for different function including, transmission, replication, intracellular and extracellular movement, assembly and encapsidation. A protein involves in a single function or it is responsible for multifunctional activity. Brief description is given about proteins and their respective function during life cycle of virus.

### 2.4.1 Coat Protein:

Coat protein (CP) is involved in a number of important functional activities including encapsidation, intracellular and extracellular movement and assembly of viral particles (ssDNA genome). The CP is also helpful in the insect transmission of viral particle. The name of CP protein is derived from its main function i.e. it surrounds ssDNA genome to maintain the integrity of the viral particle in geminate shape. In a geminate particle 110 subunits of CP are arranged as 22 pentamericcapsomers (Zhang *et al.*, 2001). CP protein is also involved in the movement of the viral particle within the cell and outside the cells to spread to spread the viral particle to other organs of the host (Boulton *et al.*, 1991). The CP is also responsible for the typical lesions produce by the species of *Geminivirus* due to the movement of virus during infection. It is also reported that the CP is involved in the nuclear shuttling to facilitate the movement of viral genetic material within the nucleus and nucleolus (Rojas *et al.*, 2001).

Guerra-peraza *et al.*, 2005 stated that the CPs were also involved in the transport of viral genome into different cells of host. They used the fluorescence protein tagging approach to prove the import of nuclear material of *Rice tungro bacilliform virus* (RTBV) and *Mungbean yellow mosaic virus* (MYMV). The result showed that the transport of CPs of both RTBV and MYMV were facilitated by a nuclear import factor protein called importin-α. They concluded that the internalization of the CP to the nucleus is dependent

by importin-α factor. The general symptoms due to the localization of CP are described by many scientists in different virus species like in TYLCV (Kunik *et al.*, 1998; Rojas *et al.*, 2001), MSV (Liu *et al.*, 1999) and ACMV (Unseld *et al.*, 2001). Any damage to CP of ToLCV can affect the spread of virus during infection (Rigden *et al.*, 1993). The CP is also involved in the systemic symptoms of the infection caused by the *Geminiviruses*. Any mutation deletion or insertion in the gene for CP caused infection without any symptoms (Boulton *et al.*, 1989). Another main function of the CP is its specificity for the vector. Bridon *et al* 1990 stated that the exchange of CP of the ACMV with the CP of BCTV switched the vector specificity from the whitefly to leafhopper.

### 2.4.2 AV2 or AR2 Protein (Pre-coat Protein):

It is also known as Pre Coat Protein. The gene AV2 or AR2 responsible for movement protein are present only in Old World *begomoviruses* viruses this gene is absent in the *Begomoviruses* viruses of New World (Ribicky, 1994; Stanley *et al.*, 2005). When the mutated *Tomato leaf curl New Delhi virus* lacking the gene responsible for AV2 protein was introduced into some host they didn't produce intense symptoms as well as low quantity of viral DNA. But when the virus was directly inoculated into protoplast the observed DNA concentration was equal to that of wild type DNA indicating that this is required for movement of virus within a host (Padidam *et al.*, 1996).

### 2.4.3 Replication Associated Protein (Rep)

The Replication Associated Protein (Rep) is code by the C1 gene also called AC1 or AL1. The product size is approximately 41kDa. The name "Rep" protein is due to its similarity with the rolling circle DNA replication initiator protein of some prokaryotes (Laufs *et al.*, 1995; Koonin and Ilyina, 1992). The main function of the Rep is the helping in the replication of virus as well as in self-suppression (Lazarowitz *et al.*, 1992; Fontes *et al.*, 1994; Eagle *et al.*, 1994).

Heyraud-Nitschke *et al.*, (1995) explained the mechanism of Rep protein action in a rolling circle replication (RCR) model. During their study they established that at repeated elements of stem loop structure the Rep binds and create a nick

(TAATATT↓AC) at the hairpin structure and initiate replication by binding to 5′ end with the help of tyrosine amino acid to the nicked DNA. The 3′-OH is used as a primer for the replication of virus DNA. The Rep is also responsible for helicase activity by forming large oligomeric complex (Choudhury *et al.*, 2006). The genome of genus *Mastrevirus* has genetic information for two Rep these Rep and RepA. The Rep is also responsible for the initiation of host replication machinery which probably the first step in viral replication inside the tissues and cells of host (Egelkrout *et al.*, 2002; Kong *et al.*, 2000). The multifunctional Rep has the ability to binds to retinoblastoma related proteins (pRBR) and altered cell cycle by binding to the transcription factor (Collin *et al.*, 1996). This binding suppress the pRBR and E2 promoter binding factor (E2F) which leads to the activation of transcription in host leaves. E2 binding to pRBR and repression of proliferating cell nuclear antigen (PCNA) a DNA polymerase processitivity factor, support the hypothesis through which the replication protein of *Geminivirus* alter the gene expression of host by pRBR/E2F pathway.

Zhang and dean (2001) stated that the E2F interact with PCNA and assembled pRBR which carry out remodeling activities i.e. histone deacetylases and some other enzymes, and generate a repressor complex. The Rep also binds to Replication factor C (RFC), a complex which helps in the transfer of PNCA to the replication site (Castillo *et al.*, 2003; Luque *et al.*, 2002). The Rep protein produces structural changes in host cells. The cells become enlarge three time than that of normal cell. The nucleus of Rep induced cells also became large and loosely appeared as compares to the cells where Rep is not induced.

### 2.4.4 Transcriptional Activator Protein (TrAP)

The TrAP is also called AC2 or AR2 protein. TrAP is a nuclear protein (Sanderfoot and Lazarowitz, 1995) and help in the expression of gene found on virion-sense strand of DNA-A (Sunter and Bisaro, 1992; Sunter *et al.*, 1994). In an experiment conducted by Noris *et al* (1996) the binding pattern of TrAP was observed in vitro. They find that the TrAP has the affinity to bind the ssDNA and dsDNA but it favors binding to ssDNA.

Another important function of TrAP is its role in the pathogenicity. As the plants have a sophisticated immune system and infectious agents are suppressed to cause infections unless and until these infectious agents have some elements which suppress, invade or inactivate the plant defense mechanism. The TrAP also has the ability to suppress some host defense factor and cause infection. An example of this suppression is the interaction of TrAP from TGMV and C2 from BCTV with the SNF1 which leads to the imuunocompromised condition and increase the susceptibility of host to infection. In such cases a decrease in latent period and low infectious dose was observed (Sunter *et al.*, 2001). Plants have another defense mechanism called RNA silencing used to inactivate the pathogenic elements. The TrAP and C2 protein were the first protein described which suppress the RNA silencing (Baulcombe, 2004; Ding *et al.*, 2004; Voinnet, 2005).

The suppression was done when the gene was expressed from both from the virus vector such as Potato virus X (PVX) and from plasmid internalized by means of agroinfiltration or bombardment (Trinks *et al.*, 2005; Vanitharani *et al.*, 2004; Voinnet *et al.*, 1999; Wang *et al.*, 2005). Researcher suggested that the silencing of gene by TrAP is by both transcription dependent and transcription independent mechanisms (Bisaro, 2006). In transcription dependent mechanism the host genes are activated which itself suppress the gene silencing mechanism (Trinks *et al.*, 2005). The second mechanism involved the inactivation of a cytoplasmic enzyme adenosine kinase (ADK) by interaction with TrAP results in a functionless enzyme (Wang *et al.*, 2005; 2003). The TrAP also involved in the suppression of plant immune system mechanism. The deficiency of ADK results in the impaired methylation process of target DNA. Methylation has been reported to reduce the replication of virus and also cut down the silencing of target gene (Buchmann *et al.*, 2009; Moffatt *et al.* 2002). Plants have protective mechanisms like hypersensitive (HR) and Program cell death (PCD). The TrAP involvement was reported in the counter activities against HR and PCD to protect the viral progeny inside the cell as well as spread from one tissue to other (Hussain *et al.*, 2007).

### 2.4.5 Replication enhancer protein (REn):

REn is a protein composed of 134 amino acids. The REn is small but have an extremely hydrophobic portion and remain constant among the genus *Begomovirus* as well in *Curtovirus* (Castillo *et al.*, 2003). REn helps in the accumulation of viral DNA in the host cells (Sunter *et al.*, 1990). The REn proteins are localized in the nuclei of host cells infected with virus. The REn and Rep are act synergistically and present at similar concentration (Nagar *et al.*, 1995). REn protein boosts the activity of Rep protein which leads to the enhanced viral DNA replication. REn has the ability of oligomerization and interaction with some host proteins mainly with proliferating cell nuclear Antigen PCNA and pRBR. Any mutation in REn gene causes decrease or completely diminished oligomerization and interaction activity with PCNA and pRBR. There are two parts hydrophobic and hydrophilic in the REn protein the hydrophobic part lies in the middle of protein and has a role in the self oligomerization, interaction with Rep and PCNA while the hydrophilic parts at C and N terminus of the protein involved in the interaction with pRBR.

### 2.4.6 AC4 Protein:

AC4 protein is coded by AL4 (AC4) gene. The silent feature of this protein is that it is present in only dicot infecting virus but the some species of *Mastrevirus* are exceptional case. The main function of this protein is development of infectious symptoms in plants (Stanley and Latham, 1992). They established stop codon for the gene AC4 which don't have any effects on other gene expression and protein. On introduction to *N. benthamiana,* the virus developed yellowing and downward curling of the leaves with stunt growth without leaves curling upward and swelling while the wild virus produce this typical swelling and upward curling symptoms. Different experiments showed that AC4 protein was responsible for the development of symptoms in the host (Rigden *et al.*, 1994; Jupin *et al.*, 1994). The AC4 protein accommodated in the infected cell periphery (Rojas *et al.*, 2001) and not required for spread of infection, DNA replication or viral assembly but it is involved in only symptoms development (Rigden *et al.*, 1994).

Vanitharani *et al.*, (2004) reported that AC4 protein of ACMV and *Sari Lankan cassava mosaic virus* (SLCMV) was involved in the suppression of PTGS.

### 2.4.7 Nuclear Shuttling Protein

The DNA beta of begomovirus composed of two regions Complementary sense strand and virion sense strand. The gene code for NSP is present at complementary sense strand. As the virus replicate inside the nucleus of host cell so the transformation of the virus to nucleus is needed. The NSP is a multifunctional protein carry out spread of infection inside cell and from one tissue to another, determination of host and transmission of virus from one host to another. The gene is responsible for the above mentioned functions but it mainly depends on the virus type and host to be infected. Mutation in in NSP gene affect the pathogenicity of virus but there is no net effect on the replication and encapsidation phenomenon (Brough *et al.*, 1988; Etessami *et al.*, 1988). The principle function of the NCP is the transportation of genomic materials (ssDNA) inside and outside the nucleus during late infection. Beside this the NSP also interact with ssDNA and transport the DNA into the cytoplasm of the infected cells here this bimolecular complex binds to the movement protein and form NSP-BC1-ssDNA complex and hence the DNA is transported to other healthy cells to produce infection (Gafni and Epel, 2002).

### 3.4.8 Movement Protein

MP is encoded by the BC1 gene responsible for cell to cell spread during infection. The MP involvement was reported in the plasmodesmatal movement of virus (Noueiry *et al.*, 1994; Sudarshana *et al.*, 1998). There are two hypotheses for the mechanism of action of MP.

***"Relay Race Model"***in this model the dsDNA present in nucleus is transported by NSP to cytoplasm where it binds to MP and dispatched for plasmodesmatal transportation.

***"Couple skating model"*** in this hypothesis it was stated that the ssDNA binds to NSP and the complex transported into the cytoplasm. In cytoplasm the MP attached and then moved the complex across the cells. Any mutation in NSP and MP blocked the

movement of virus from one cell to another cell (Sudarshana *et al.,* 1998). The MP protein is also involved in the increase of plasmodesmatal space to which the viral DNA is injected. Due to this increase in size and mediation by proteins the DNA transported into other cells (Noueiry *et al.,* 1994; Rojas *et al.,* 1998). Both the MP and plasmodesmatal are highly specific to the size of viral genome to be transported (Gilbertson *et al.,* 2003). MP was also reported in the symptoms development when express alone. The MP gene of SqLCuV was expressed SqLCV which give positive results.

## Cotton leaf Curl Disease:

Cotton leaf curl disease was first time observed in an African country Nigeria in 1912. Due to its mild effect, it did not get more attention. In 1924 the disease occurred once again with extended effects causing loss of production in the same country as well as in Sudan. The disease of similar signs and symptoms was also reported from Tanzania in 1926 (Kirkpatricka, 1931). Cotton leaf curl disease got great attention when the first epidemic of the disease occurred in Sudan in 1927-28 (Kirkpatricka, 1931). The disease was spread by means of a vector called white fly *Bemisiatabaci*Genn.It was hypothesized that the *Bemisiatabaci*was responsible for the transmission of "virus like particle" which might be the causative agent for cotton leaf curl disease. The electron microscopy revealed that a twinned particle in the infected tissue of cotton plants which was thought to be the culprit of the cotton leaf curl disease (Brown and Nelson, 1984). Later on it was discovered that the causative agent of the disease is a virus and named as "*Cotton leaf curlvirus*".

In Pakistan the *Cotton leaf curldisease* was reported first time in 1967 near Multan city in Punjab province but due to negligible harm it wasn't get much more importance (Hussain and Ali, 1975). In 1987-88 the cotton production was badly affected by *Cotton leaf curl virus* and epidemic outbreaks were observed in the followed years (Aslam *et al.*, 2000). In 1991-92 the cotton leaf curl disease became an epidemic in Pakistan and cause most prominent fall in cotton production per hectare (Briddon and Markham, 2000). The reports showed that about 7.7 million bales of cotton are lost since 1988 till 2002

(Akhtar *et al.*, 2005). The estimated loss due to CLCuD during 5 years from 1992 to 1997 was about 50 million US dollars (Briddon and Markham, 2000). The development of *Cotton leaf curl Virus* resistant cotton variety CIM-1100 in 1996 overcome the problem, but the reemerging in Burewala of *Cotton leaf curl virus* specie which breaks the existent resistant of cotton plants. The specie was reported from Burewala Pakistan so it was term as *BurewalaCotton leaf curl virus* ($CLCuV^{BUR}$). Due to breaking the resistant of cotton plant and highly variability it is considered the most dangerous virus for cotton production in Pakistan (Ahmad *et al.*, 2010; Mansoor *et al.*, 2003). The causative agent of cotton leaf curl virus is *Cotton leaf curl virus* belongs to the *Begomovirus* genus of family *Geminiviradae*. The members of family *Geminiviradae*are composed of paired icosahedral like particle also called geminate or twinned particle which encapsidated the genomic materials (Tahir *et al.*, 2011; Fonding *et al.*, 2000). Size of CLCuV is about 20nm in diameter 30nm long. Now a days it was determined that the CLCuD is not caused by only CLCuV but it is caused by a group of *Begomovirses* associated with CLCuD. This complex may cause the disease solely or together. This group of virus is called Cotton leaf curl disease associated *Begomovirus* complex. This complex is composed of about 10 distinct *Begomoviruses* and continuously emerging new virus due to mutation in different geographic regions. CLCuD associated *Begomovirus* complex is composed of

1- *Cotton leaf curl Multan virus*
2- *Cotton leaf curl Burewala virus*
3- *Cotton leaf curl Rajasthan virus*
4- *Cotton leaf curl Kokhran virus*
5- *Cotton leaf curl Alahabad virus*
6- *Cotton leaf curl Bangalore virus*
7- *Papaya leaf curl virus*
8- *Gossypiumpunctatum mild leaf curl virus*
9- *Tomato leaf curl Bangalore virus*
10- *Cotton leaf curl Gezira virus*

*Chapter 3*

# *Materials and Methods*

### 3.1 Sample Collection:

The samples of leaf of cotton plants showing clear symptoms of CLCDV were collected from different areas of lodhran and bahawalnagarPunjab Pakistan. Healthy symptomless cotton leaves were also taken from the same field as negative control. The leaves were kept on ice until reached in lab where they were stored at -20 °C until used.

### 3.2 DNA Extraction:

The DNA was extracted using the CTAB method (Khan *et al.*, 2004) with modifications as follows. Approximately 0.5 g of leaf tissue was pulverized using a pestle and mortar in 1.5 mL CTAB buffer (100 mMTris-HCl pH 8.0, 2% (w/v) CTAB, 20 mM EDTA, 1.4 M NaCl, and 1% (v/v) β-mercaptoethanol) and incubated at 65°C for 45 minutes. The supernatant was extracted with an equal volume of phenol: chloroform: isoamyl alcohol (24:24:1) and then by chloroform. DNA was precipitated by adding 0.1 volume of 3 M potassium acetate and an equal volume of isopropanol to the supernatant, followed by incubation for five minutes at room temperature. The resultant pellet was washed with 70% (v/v) ethanol and dissolved in 100 µL of Tris-EDTA (TE) buffer (10 mMTris-HCl, 1 mM EDTA, pH. 8.0) containing RNase (100 µg/mL) and stored at -20°C until used.

### 3.3 Primer Design:

Primers were designed on sequences obtained from GenBank. The conserve regions were identified by aligning the sequence by ClustalW (Thompson *et al.*, 1994) program implemented in Molecular Evolutionary Genetics Analysis Program (MEGA) version 4.0.1 (Tamura *et al.*, 2007) or version 3.0 (Kumar *et al.*, 2004).

### 3.4 PCR and Ligation:

A typical PCR reaction contained about 50 ng DNA template, *Taq* buffer (10 mMTris-HCl, pH 8.8, 50 mMKCl and 1% (v/v) Nonidet P40) 1.5 mM $MgCl_2$, 200 µM of each

dNTPs, 1.5 units *Taq* DNA Polymerase (recombinant) (Fermentas UAB Lithuania), and 25 pM of each primer. The PCR thermal profile for all primer sets were pre-PCR denaturation at 96°C for 3 minutes followed by 35 cycles of denaturing at 96°C for 20 seconds, annealing at 60°C for 20 seconds and extension at 72°C for 150 seconds, and a final extension at 72°C for 20 minutes. After PCR, the products were analyzed by electrophoresis using 1% (w/v) agrose gel in TAE buffer (40 mMTris-acetate, 1 mM EDTA pH 8.1) according to Sambrook and Russell (2001).

### 3.5 Ligation of PCR Products:

The PCR products were ligated directly into pTZ57R vector (InsTAclone™ PCR Product Cloning Kit, Fermentas UAB Lithuania) according to the manufacturer's instructions. A typical ligation mixture contained (40 mMtris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP, pH 7.8), 10% (w/v) PEG 4000, 2.5 U T4 DNA ligase, 0.15 µg pTZ57R/T vector and about 1 µg of PCR product. The ligation mixture was incubated at 4°C overnight and then used for electroporation into electrocompetent cells.

### 3.6 Competent Cell Preparation, Electroporation and Screening of Clones:

*Escherichia coli* DH5α (*supE*44, Δ*lac*U169 (Φ80*lac*ZΔM15), *hsd*R17, *recA*1, *endA*1, *gyrA*96, *thi*-1, *relA*1) cells were used for cloning purposes in this study. The cells were prepared according to Sambrook and Russell (2001). Briefly, a single colony of *E. coli* DH5α was cultured overnight in 3 ml Luria-Bertani (LB) medium (Tryptone 10g/L, Yeast extract 5g/L and NaCl 10g/L, pH 7.0) in a test tube at 37°C. The culture was shifted into 100 ml LB medium in a 250 ml flask at 37°C, until the O.D at 600 nm became 0.4-0.7. Then the cells were gently pelleted down by centrifugation at 1275xg for 15 minutes. The pellet was re-suspended in nanopure water based 15% (v/v) glycerol buffer. The cells were again pelleted down by centrifugation at 1275xg for 15 minutes. This washing step was performed three times and after the final washing step the cells were stored in 50 µL aliquots in eppendorf tubes frozen in liquid nitrogen and immediately placed in -80°C. All the handling and centrifugation steps were performed at 4°C.

The electrocompetent cells were mixed with ligation mixture and were transformed by giving the electrical pulse of ~10 milliseconds at medium voltage (2.5 kV) by using Cell-Porator *E. coli* Pulser, Life Technologies™, (GIBCO BRL USA) according to manufacturer's instructions. The cells were recovered from the electroporation cuvette by immediately adding 1 ml of liquid LB medium and mixing. The total mixture was transferred to a 1.5 ml eppendorf tube and incubated at 37°C for one hour on shaking incubator. After incubation 100 µl of transformed culture was spread on solid LB medium and selected using 50 µg/ml ampicillin, screened for functional β-glucoronidase expression, using 40 µL of 0.1M IPTG (isopropyl-β-D-thiogalactopyranoside) solution per plate with 40 µL of 20 mg/mL X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) solution per plate. The white colonies (indication of insertional inactivation of *lac*Z gene by cloned fragment) was picked up with sterile toothpicks and cultured in liquid LB medium containing 50 µg/mL ampicillin. The culture tubes were kept at 37°C in orbital incubated overnight with vigorous shaking for isolation of the plasmids.

**3.7 Plasmid Isolation:**

Plasmids from the selected clones were isolated using minipreparation protocol according to Sambrook and Russell (2001). In brief, from overnight cultures of a single *E. coli* colony in about 10 ml Lauria-Bertani (LB) broth containing 50 µg/mL ampicillin, about 3 mL culture was pelleted down by centrifugation at 3300xg for 3 minutes and the supernatant was removed. To resuspend the pellet, 100 µL of solution-I containing Tris, pH 7.4-7.6, 50 mM, EDTA 1 mM, RNase 100 mg/ml was added. Then 150 µL of solution-II containing 0.2N NaOH, 1% (w/v) SDS was added to and mixed well by inverting gently four or five times followed by addition of 200 µL of solution-III containing Potassium acetate 3 M, pH 4.8-5.0, and mixed comprehensively. The total mixture was centrifuged at 17,950xg for 5 minutes and the supernatant was taken in fresh eppendorf tube. DNA was precipitated by 1:1 volume of isopropanol. The mixture was centrifuged at 17,950xg for 5 minutes and pellet was washed with 70% (v/v) ethanol. Finally the plasmid DNA was dissolved in 50µL of sterile distilled water and stored at -

20°C. The plasmid concentration and size was estimated by agarose gel electrophoresis (Sambrook and Russell, 2001).

### 3.8 Confirmation of Cloning:

The clones were confirmed for the presence of cloned fragment by digestion with appropriate restriction endonucleases. A typical restriction reaction contained 10X respective restriction buffer diluted using nanopure water to 1X final concentration in reaction, 5.0 U of appropriate enzyme (Fermentas UAB Lithuania) and about 250 ng plasmid.

### 3.9 Sequencing of the Clones:

For sequencing, the plasmid DNA was purified by using Wizard® SV Gel and PCR Clean-Up System Kit (Promega Corp., Madison, WI, USA). Sequencing was commerically performed Macrogen Standard Custorm DNA Sequencing Services (Macrogen Inc. Seoul, Korea). For all sequencing reactions M13 F and M13 R and insert-specific primers were used to sequence both the strands.

## 3.10 Sequence Analysis:

### 3.10.1 Compilation of Sequencing Data:

The sequence data was compiled manually with the help of Blast 2 sequences (Tatusova and Madden, 1999; at http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi), for both the strand separately for each clone. Then the sequences from both the strands were complementary aligned to each other. In this way sequence from one clone was determined. A consensus sequence for each component was developed by aligning multiple sequence of every clone belonging to the same component of individual isolates by using ClustalW (Thompson et al., 1994). In this way all the sequence data was compiled.

### 3.10.2 Nucleotide and Amino Acid Identities and Homology Analysis:

The percentage identities were calculated for each pair of isolates and compared with each other using MetGat program (Campanella et al., 2003). Nucleotide homology search was conducted using the nucleotide blast program "Blastn" (Altschul, et al., 1997) available through World Wide Web at (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi).

### 3.10.3 Phylogenetic Analysis:

Full-length DNA sequences of the isolates from the family *Geminiviradae* were obtained from the nucleotide data base in the GenBank (Table 1) and used for different analyses. The sequences of genomic components were aligned in ClustalW Sequence Alignment program (Thompson *et al.*, 1994) using IUB matrix for DNA alignments in the Molecular Evolutionary Genetics Analysis Program (MEGA) version 4.0.1 (Tamura et al., 2007) or version 3.0 (Kumar et al., 2004). Neighbor-Joining (NJ) analysis was carried out using Maximum Composite-likelihood model with uniform rates among the sites, the 1000 bootstraps replicates were used to evaluate the significance of generated tree. Translated amino acid sequences of the corresponding components were obtained from the protein data base in the GenBank, aligned in ClustalW (Thompson *et al.*, 1994) using BLOSUM matrix for amino acid in the Molecular Evolutionary Genetics Analysis Program (MEGA) version 4.0.1 (Tamura *et al*,. 2007).

*Chapter 4*

# *Results*

### 4.1. Isolation of DNA:

The total DNA extracted from cotton leaves were analyzed by agarose gel electrophoresis and visualized by staining it with ethedium bromide (Figure 4.1). The analysis showed high molecular weight DNA indicating appropriate isolation of total genomic DNA.

### 4.2. PCR amplification of actin gene as positive control:

The extraction of DNA from the cotton leaves was further conformed for its used in PCR by amplifying actin gene from nuclear genome of cotton. The primers designed in coding region of actin gene were used in PCR to amplify DNA extracted from different infected and healthy cotton plants and the resultant bands were analyzed using agarose gel electrophoresis and visualized by staining with ethedium bromide (Figure 4.2). The PCR products gave an expacted band of about 700 bp from infected and healthy cotton plants. This provides confirmation that the DNA isolation procedure adopted is capable to isolating DNA from cotton plants and the resultant DNA has sufficient purity to be used as template in PCR analysis.

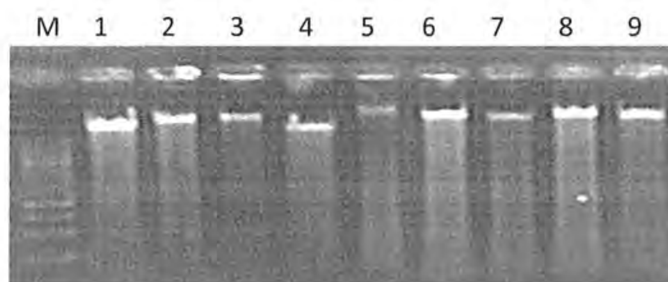



**Figure 4.1: Gel image of DNA from cotton leaves;**

Lane 1 – 9 is DNA extracted from different samples of cotton leaves collected from different areas. M is O'RangeRuler™ 500 bp DNA Ladder (Fermentas Lithuania UAB).

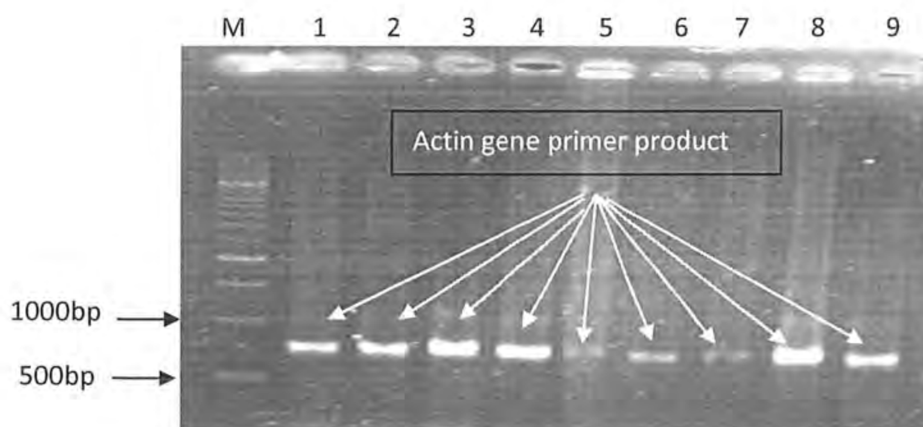



**Figure 4.2: PCR amplification of cotton genomic DNA with actin gene primers**

Lane 1 – 9 is cotton DNA amplified with actin gene primer. All samples gave positive result with actin gene primer.

**4.3. Optimization PCR amplification:**

The PCR assay was optimized to PCR amplify CLCuV DNA-β from infected cotton plants. The PCR assay was optimized for annealing temperature of primer at different temperatures.The annealing temperature has a great effect on optimization the amplification condition. The optimum temperature for amplification was obtained at $60^0$C (Fig 4.4) after the experimental trail at $52^0$C, $55^0$C, $58^0$C, $60^0$C and $62^0$C, The amplification at $60^0$C showed prominent and specific band while other temperature showed either faint bands or non-specific bands (Fig 4.3).

The primer optimization was carried at different annealing temperature and extension time to obtained the optimized condition for PCR amplification of targeted DNA at optimum level. At first the primers were used to amplified the targeted DNA at annealing temperature of $52^0$C and kept the extension time three minute. Amplification at $52^0$C gave specificity but the obtained bands were very faint which needs to be improved. Then annealing temperature of $55^0$C and extension time of three minutes was used to amplify the targeted CLCuV DNA but weak and multiple bands were obtained. For further optimization the annealing temperature was kept $58^0$C keeping the extension time same as previously. This gave an intense product along with some very weak nonspecific band. The specificity was improved to increase the annealing temperature to $60^0$C for same extension time. Very clear and intense band were obtained along with weak primer dimer. For more specification the annealing temperature was also kept at $62^0$C instead of better result the product was weak compare to the product of $60^0$C. These results showed that at low temperature the primers have low specificity and gave weak and multiple bands while as the temperature exceeded $60^0$C the specificity of primer increases but the bend strength decreases strongly (Fig 4.3).

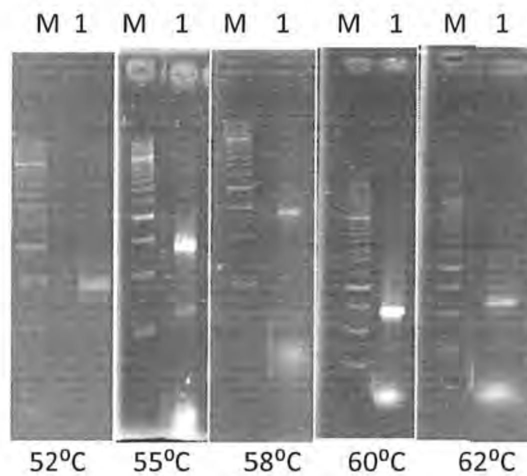



**Figure 4.3: Summery of all Optimization experiments;**

PCR optimization using same DNA sample, primer mix and extension temp i.e. 72°C, while changing the annealing temp for the primer. M is marker O'RangeRuler™ 500 bp DNA Ladder (Fermentas Lithuania UAB).

### 4.4. PCR amplification of DNA-β:

The isolated DNA confirmed with actin gene amplification was used as template to PCR amplifies DNA-β using Uni-CLCuVβ forward and reverse primer pair (appendix 4). The amplified products were separated on agarose gel electrophoresis (Figure 4.4). The PCR products showed an expected band of about 1.3 kb indicating the proper amplification of target DNA β from these infected samples.

## 4.5 Cloning:

After successful PCR amplification, the cloning of these products was performed. For the successful cloning of desire gene and DNA segment the followings steps were performed.

### 4.5.1 Refreshment of bacterial culture used for transformation:

The bacterial cultural strain *E. coli* DH5α were used for the transformation. The agar stock culture was refreshed on LB agar plates. The gram staining of the culture was done which revealed that the microbes were pink short rods and scattered in arrangement.

### 4.5.2 Culturing After Transformation:

The bacterial cells were made competent for transforming plasmid DNA using methods described in section (3.6). The ligated products were electroporated into competent bacterial cells and grown in LB agar plates containing ampicillin antibiotic for selection (Fig 4.5). The appeared colonies showed that the desired DNA transfer successfully to the competent cells because the transformed cells containing the vector which have the indicator gene coding Extended Spectrum β Lactamases (ESBL) which inactivate the ampicilin in the media and allow the growth of transformed cells selectively.

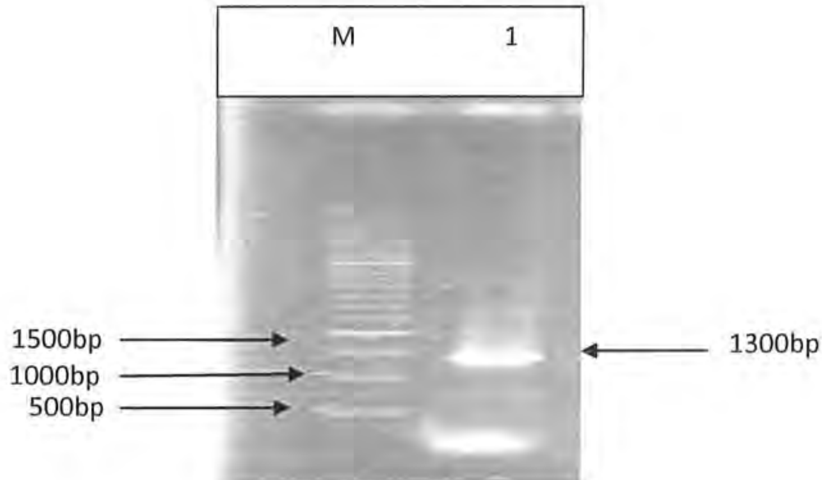



**Fig 4.4: PCR Amplification at 60°C**: Lane 1 is cotton DNA with CLCuV-β. Annealing temp was kept $60^0$C and extension was carried out at $72^0$C for 3minutes. : M is marker O'RangeRuler™ 500 bp DNA Ladder (Fermentas Lithuania UAB)

In this experiment, clear and prominent bands appeared as the annealing temperature was kept at 60°C. The extension temperature was kept the same i.e., 72°C for all the optimization experiments. Same primers and DNA samples were used, and at 60°C, they amplified well.

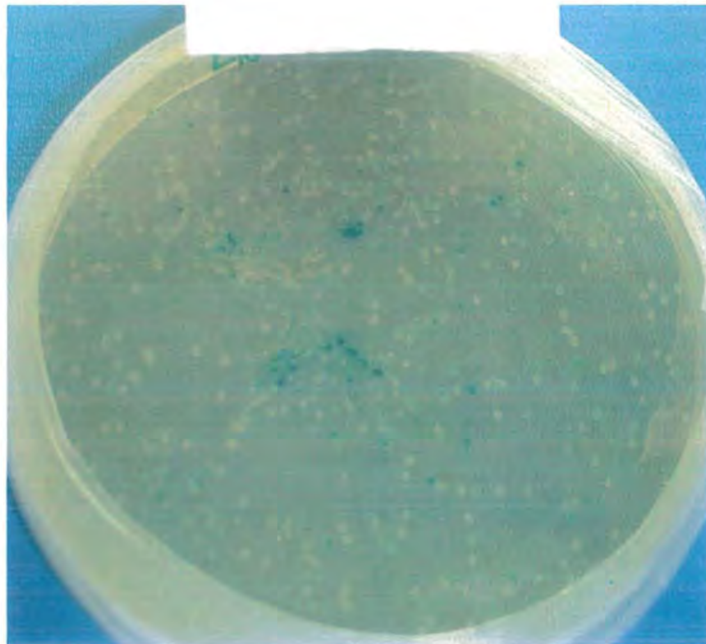


**Fig 4.5:** Growth of E.coli DH5 α competent cell on XGAL/IPTG LB ampicillin agar plates. Transformed cellof E coli appear in whitish colony while non transformed cell are in bluish colony

### 4.5.3 Plasmid isolation:

After transformation the plasmid was extracted to confirm the transformation of vector (vector + DNA of interest) to cells. The extracted plasmid's gel image shows the characteristic multiple bands. The multiple bands were due to different topology of supercoiled of plasmid DNA (Fig 4.6)

### 4.5.4 Plasmid Confirmation:

The extracted were confirmed by three different experiments which are described in following;

*4.5.4.1. Confirmation of Plasmid by restriction enzymes:*

The restriction enzymes were used which excised out the DNA fragment by cutting the vector at specific sites. This restriction cut results the formation of two bands one equal to the desire DNA bands size and the other is equal to the vector length. The restriction enzymes have their own specific sites lies near or upon the end of vector strand end. And the restriction cut release two sequences of different size. In this study the pTZ57R/T vector was used which is of about 2.8kb appeared on gel and the amplified product of 1.3 kb was observed (Fig 4.7).

*4.4.2 Confirmation by M13 primer amplification:*

The clones were also confirmed by a PCR amplification of extracted plasmid with M13 forward and reverse primer mix. The product size revealed that our DNA was successfully cloned into the transformed cells. The product size of M13 primer was almost equal to the size the desired DNA size (Fig 4.7).

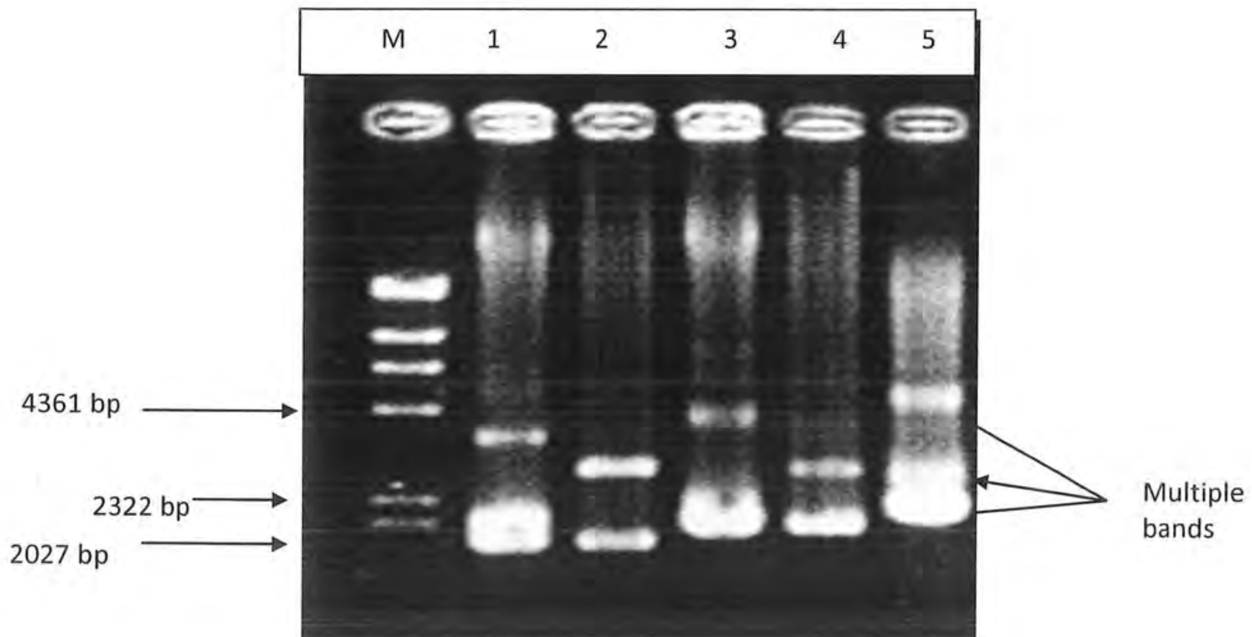



**Fig4.6:** plasmid gel image

Lane 1-3 is plasmid of CLCuV-Pk-Bahawalnagar-β-2011-1

Lane 4-5 is plasmid of CLCuV-Pk-Bahawalnagar-β-2011-2

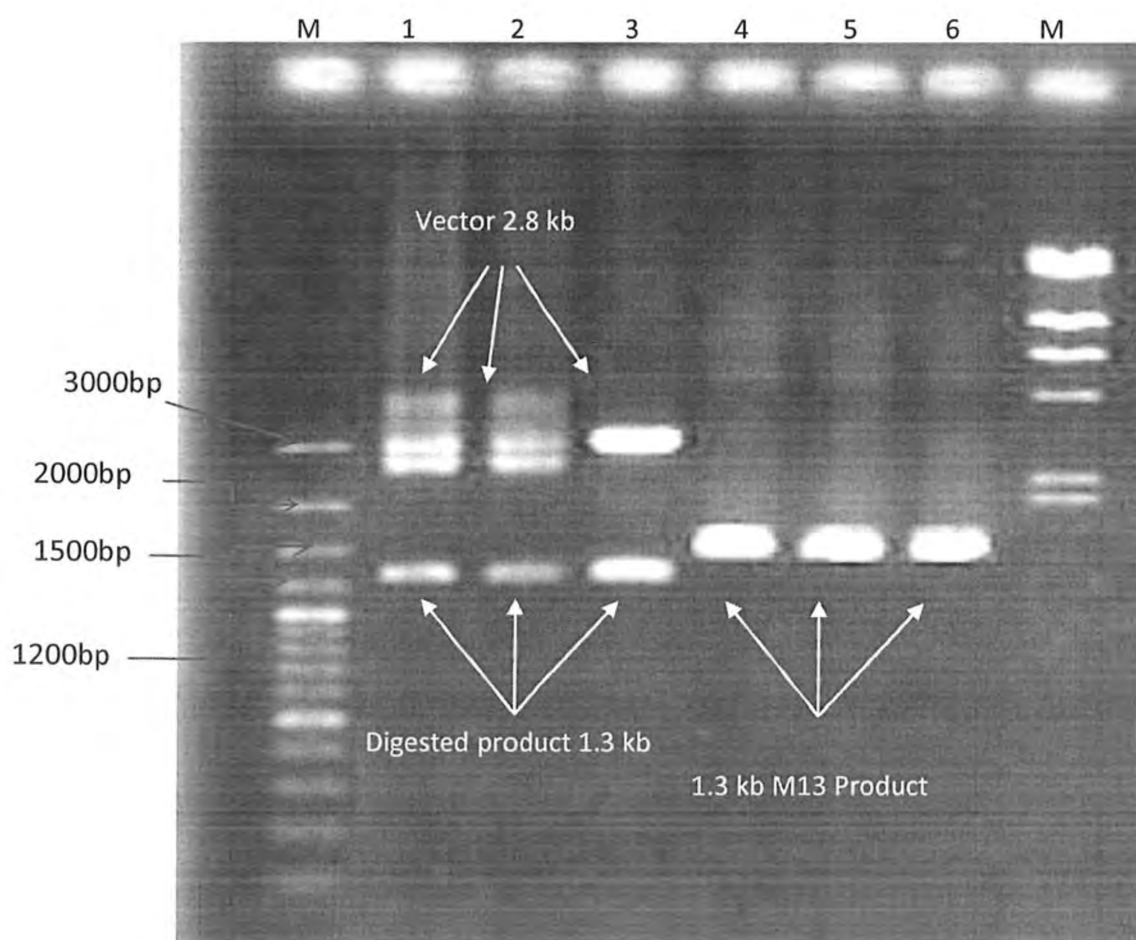



Fig 4.7: Digested plasmid and clone with M13

Lane 1-2 is digested plasmid of CLCuV-Pk-Bahawalnagar-β-2011-1

Lane 3 is digested plasmid of CLCuV-Pk-Bahawalnagar-β-2011-2

Lane 4-5 is CLCuV-Pk-Bahawalnagar-β-2011-1 with M13 primer

Lane 6 5 is CLCuV-Pk-Bahawalnagar-β-2011-2 with M13 primer

**4.5 Sequencing:**

For sequencing, the plasmid DNA was purified by using Wizard® SV Gel and PCR Clean-Up System Kit (Promega Corp., Madison, WI, USA). Sequencing was commerically performed Macrogen Standard Custorm DNA Sequencing Services (Macrogen Inc. Seoul, Korea). For all sequencing reactions M13 F and M13 R and insert-specific primers were used to sequence both the strands

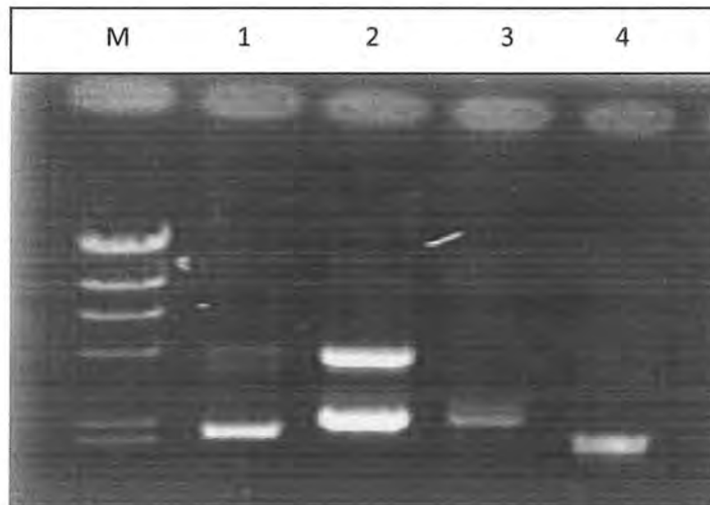



**Fig 4.8: purified plasmid**

M marker λ hind III

Lane 1-2 is purified plasmid of CLCuV-Pk-Bahawalnagar-β-2011-1 and Lane 3-4 is purified plasmid of CLCuV-Pk-Bahawalnagar-β-2011-2

### 4.6 Phylogenetic analysis with all CLCuD associated *Begomovirus*

Phylogenetic analyses of the isolated samples were done with the available sequences available in NCBI GenBank. The description of isolates, their abbreviated names and GenBanks accession number are presented in appendix 03.The sequences of isolates and sample sequenced in this study were aligned and analyzed by Neighbor Joining (NJ) method using maximum-like hood approach in MEGA 4.0 using homogenous evolutionary rated with uniform distribution and complete deletion. The resulted NJ tree (Figure 4.9) is based on all publically available CLCuD associated *Begomovirus* which included sequences of DNA-β isolated from cotton, okra, tomato crops from different geographic regions including China, Africa, Pakistan, India and Bangladesh. Most of the isolates of DNA-β were those which infected cotton crops while China was the country from where most of the sequences were reported.

The NJ tress resulted into five different clusters of isolates based on geographic regions where isolates showed close evolutionary relationships. On the basis of geography following clusters can be identified.

- Cluster I or Pakistani cluster I
- Cluster II or Indian cluster
- Cluster III or African cluster
- Cluster IV or Pakistani cluster II
- Cluster V or Chinese cluster

The Cluster I or Pakistani cluster I (with 95% bootstrap support) is consisting of CLCuV-Pk-Bahawalnagar isolates. The cluster also shows that this isolate is found in other regions of the world specifically in India. Some isolates grouped into "Cluster I" which was reported from different regions of India. The isolate CLCuV-Pk-Bahawalnagar may be transferred from Pakistan to India or vice versa through planting material or white fly vector. This cluster also have indian isolates which are CLCuV-I-(DQ364230), CLCuV-I-(AJ3160380), CLCuV-I-(AJ316037).The isolates sequenced in current study CLCuV-Pk-Bahawalnagar 2011-1 and CLCuV-Pk-Bahawalnagar 2011-2 are clustered into this

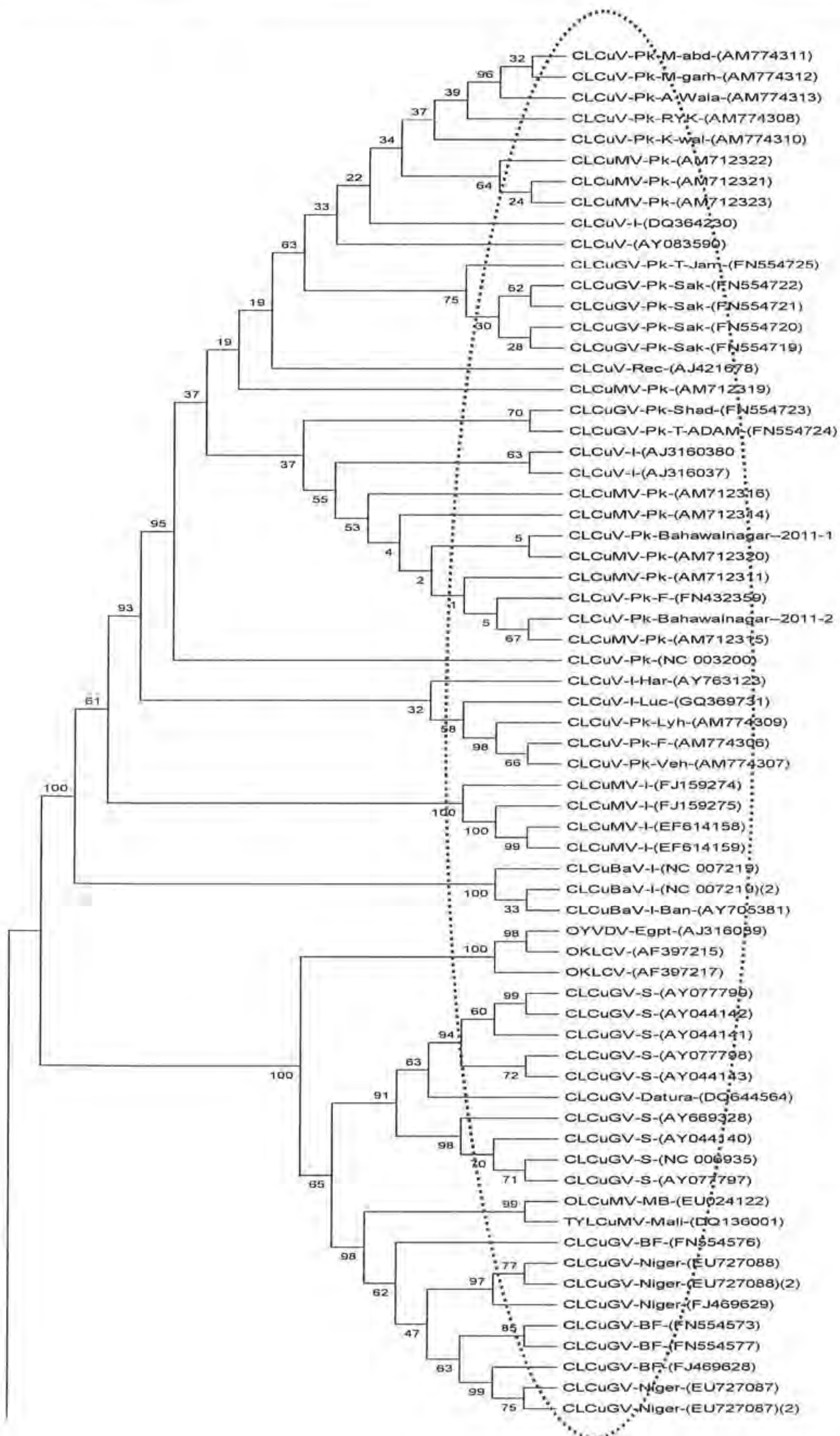
group. It means that these have close similarity with the member of this group. The clustering pattern showed that these isolate belong to CLCuMV. Geographically Bahawalnagar is present in cotton growing area and lies very close to Multan regions. These isolated species are much more close to the isolates from Multan and Faisalabad indicating a common ancestral relationship.the isolate in cluster I are from different region of Pakistan which are Faisalabad,Arifwala,Multan,Sangarh,and Rahim yar khan.

Cluster II or Indian cluster (with 100% bootstrap support) has isolates of India which are grouped in cluster II. The result shows that the similarity between the isolates of each country is much more. The isolates from India are grouped together in sub branches.the Indian cluster also have a Pakistani isolate from vehari and Faisalabad region.the indian isolate are from Haryana.Banglore and Lucknow.

Cluster III or African cluster (with 100% bootstrap support) is composed of mix isolates from different regions of Africa mostly from Sudan, Niger, Mali and Burkina Faso. This cluster has OKLCV,CLCuGV And TYLCuM. The result shows that the similarity between the isolates of each country is much more. The isolates from Sudan are grouped together in sub branches.

Cluster IVor Pakistani cluster II (with 100% bootstrap support) contained CLCuMV-PK isolates reported from Pakistan,TLCuV-B-G-(AJ542489) and CLCuV-I-LUC-(GU440531) are the two isolates which don't belong to Pakistan these belong to Bangladesh and India respectively. The virus in cluster IV can cause infection in cotton and tomato.

Cluster V or Chinese cluster (with 99% bootstrap support) contains isolate from china. The isolates from china are grouped together. It contain TYLCuChV isolate.




**Indo-Africa Cluster**

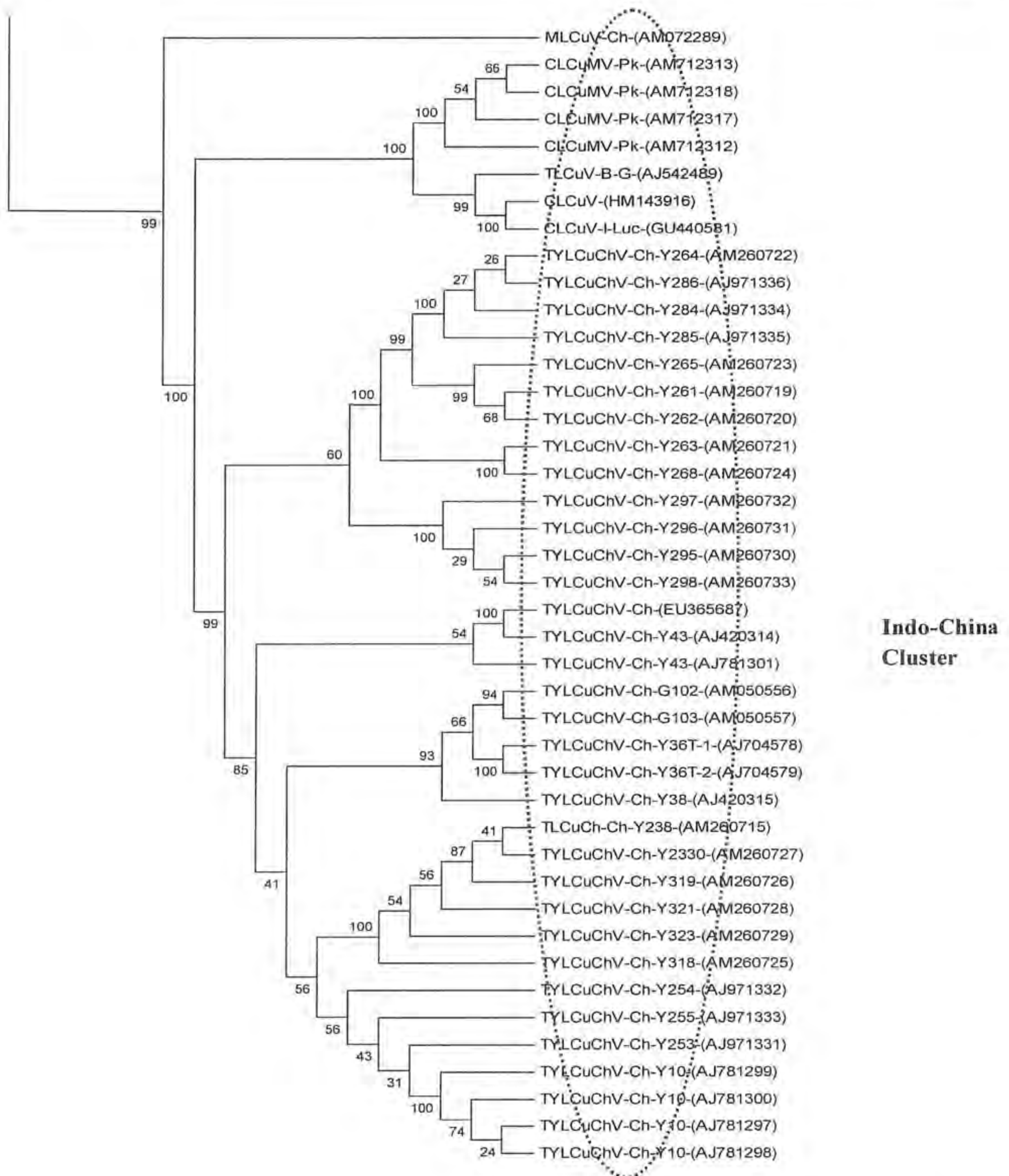
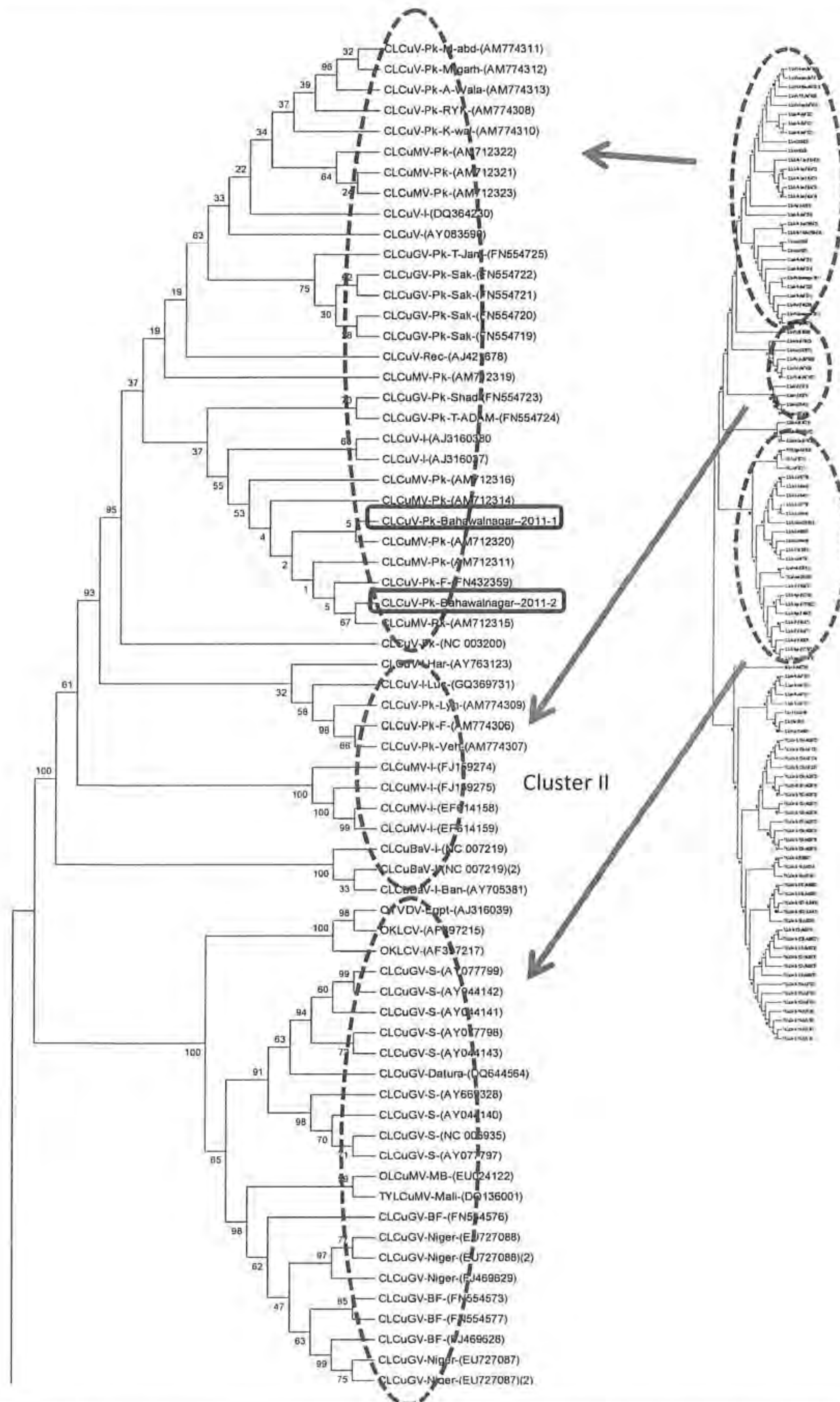



**Figure 4.9 NJ tree of CLCuV-PK-Bahawalnagar 2011-1 and CLCuV-PK-Bahawalnagar 2011-2** with all the available Begomoviruses isolates in NCBI genome data base.

Phylogenetic relationship of **CLCuV-PK-Bahawalnagar 2011-1 and CLCuV-PK-Bahawalnagar 2011-2** isolate with Begomovirus isolates clustered irrespective of the subgroups and isolate types bootstrap value are shown on the branches. Nomenclature of all the CLCuv isolates are given in Appendix 3



CLCuV-Pk-A-Wala-(AM774313) — excluded: the "W" is covered by the dashed circle
CLCuV-Pk-F-(FN432359)
CLCuV-Pk-Bahawalnagar--2011-1
CLCuV-Pk-Bahawalnagar--2011-2
CLCuV-Pk-Lyp-(AM774309)
Cluster II
CLCuBaV-I-(NC 007219)
CLCuGV-S-(AY077799)
CLCuGV-S-(AY044142)
CLCuGV-S-(AY044143)
CLCuGV-Datura-(DQ644564)
CLCuGV-S-(AY044140)
OLCuMV-MB-(EU024122)
TYLCuMV-Mali-(DQ136001)
CLCuGV-Niger-(EU727088)
CLCuGV-Niger-(EU727088)(2)
CLCuGV-BF-(FN554573)
CLCuGV-Niger-(EU727087)

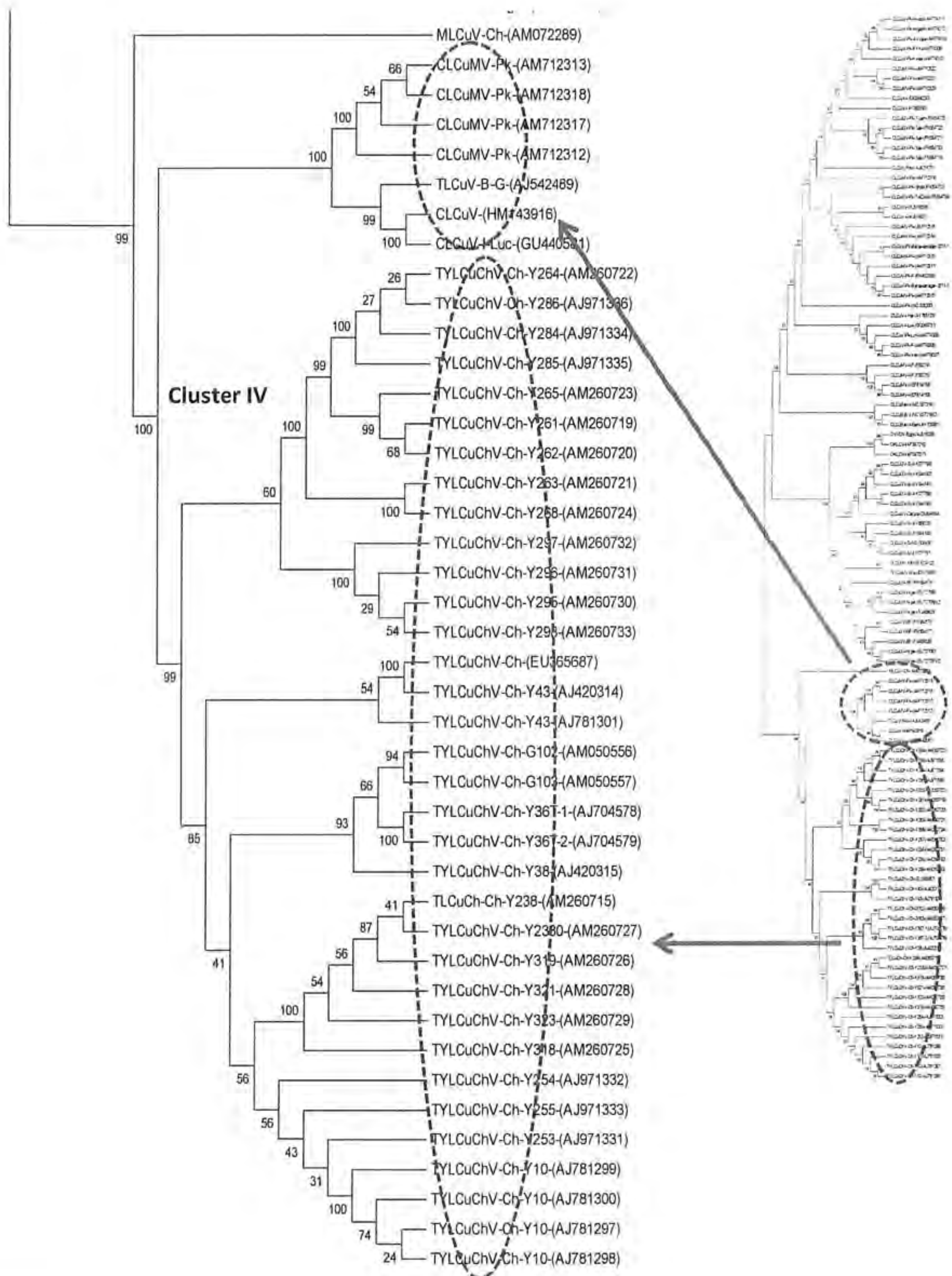



**Fig 4.10: Phylogenetic tree of all CLCuV (Evolutionary relationships of 103 taxa)**

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei **1987**1). The optimal tree with the sum of branch length = 1.95578885 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Tamura *et al.*, 2004). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004). and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The differences in the composition bias among sequences were considered in evolutionary comparisons. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analyses were conducted in MEGA4

In this figure the NJ tree was constructed using MEGA 4 of study isolates with the all other isolates. The similar genomic groups are cluster together irrespective of subgroups. The values of bootstraps are given at the braches. The sequences of all other isolates were downloaded from NCBI GenBank.

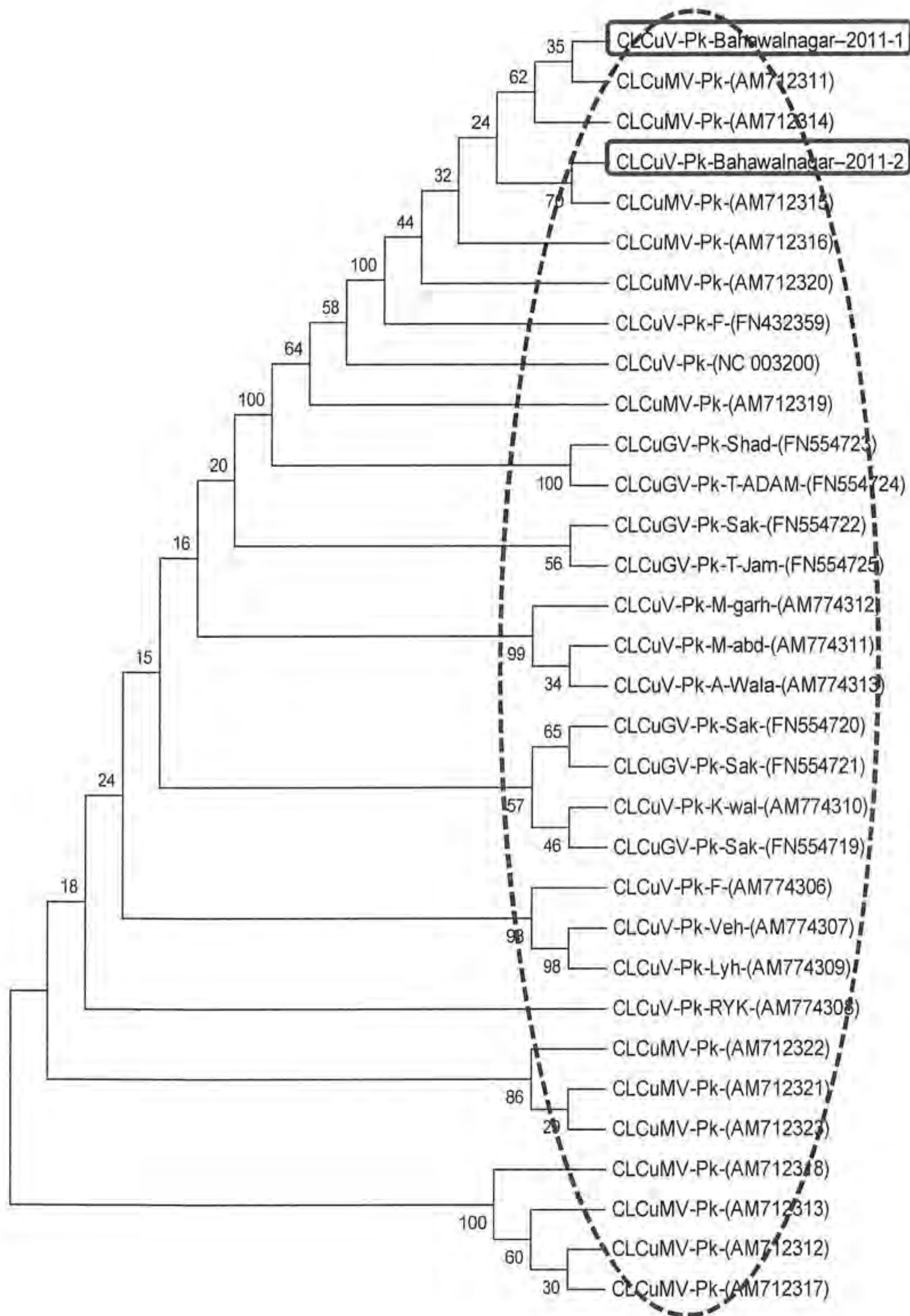



**Figure 4.11 NJ tree of CLCuV-PK-Bahawalnagar 2011-1 and CLCuV-PK-Bahawalnagar 2011-2** with all the available Pakistani Begomoviruses isolates in NCBI genome data base.

Phylogenetic relationship of **CLCuV-PK-Bahawalnagar 2011-1 and CLCuV-PK-Bahawalnagar 2011-2** isolate with Begomovirus isolatesare clustered irrespective of the subgroups and isolate types bootstrap value are shown on the branches. Nomenclature of all the CLCuv isolates are given in Appendix 3

*Chapter 5*

# *Discussion*

Cotton is the most important cash crop of the world. Especially it is of high value in the Subcontinent, more precisely in Pakistan as well as in India. Cotton has many uses according to it demand. The most value able use of cotton is in fabric industries. Millions of mills are working to convert the raw cotton fibers into highly economic valued garment stuffs. Beside this cotton seed, cotton lint, seed hulls, are also used for cooking oil extraction, animal feed. Worldwide cotton is produced in many countries. Initially the cotton was a wild plant which was domesticated as humans came to know the beneficial aspects of cotton plant and cotton became an important cash crop. As the growth and production of cotton increases some problems were also arose which negatively affect the cotton production. Due to continuous effort some were eradicated while some persistent constantly or reemerge after some time. One of the most crucial obstacles in the cotton production is the infectious agent virus. Viral infections adversely affect the quality and quantity of cotton. Besides the living and nonliving controversy of virus particle it is of great importance for human beings. Both the medical and agriculture side have deeply affected by this disputed thing called virus.

In agriculture sector the viral family *Geminiviridae* is the most important one. This family contains four genera in which *Begomovirus* is most significant. CLCuV belongs to this genus. Worldwide CLCuV can damage cotton quantity and quality as well. Due to CLCuV attack the cotton plant become stunt and leaf upward and downward, notation at leaves were also observed results a remarkable decrease in cotton production with low standard.

Kirkpatricka, (1931) first reported a mild CLCuD from Sudan in 1912, and after this mild incident the disease was reported in 1924 and 1926 in Niger and Tanzania with wider effect range. The disease was then spread to other countries. In Pakistan the CLCuD was detected in 1967 near Multan region (Hussain and Ali, 1975). CLCuD has a great threat for Pakistan cotton industry. Currently in Pakistan CLCuMV and CLCuBrV are most dangerous. The CLCuBrV is thought to be a recombinant isolate of CLCuMV and CLCuKoV. In 2001 the resistant cotton plants develops symptoms of CLCuD in

Burewala Punjab (Mansoor *et al.*, 2003). It is also hypothesized that a plant may contain more than one isolate to cause infection and hence the infectivity of viruses are boosted and therefore the emerging of new resistant strain are more due to the recombination process (Mansoor *et al.*, 2003).

The host range of begomovirus is quite broad as well and its vector white fly is known to transfer multiple begomovirus efficiently which makes the coinfection of multiple viruses in same host possible.ori of *Geminivirus* is located in LIR and is a site where a nick is produced to make an end for the start of rolling circle replication (koonin *et al.*,1992) and unit genome length are rejoined.the production of an end with highly conserved sequence makes homologous recombination more frequent.that is why the ori is well established recombination breakpoint in geminiviruses (lefeuvr *et al.*, 2009).

Satellite are viruses or nucleic acid that depend on a helper virus for their replication but do not have extensive nucleotide sequence similarity to their helper virus and are dispensable for its proliferation (Murant and mayo,1982).DNA β are symptom modulating, single stranded DNA satellites associated with monopartite begomoviruses).DNA β is responsible for movement in plant, insect transmission. The function of DNA β in the development of infection remains unclear but responsible for movement ,creating a cellular environment for viral replication and may overcoming host defense mechanism (Briddon *et al.*, 2003).

The current study was aimed to PCR amplify DNA-β from cotton growing areas in Pakistan and then sequence it so that evolutionary relationships of currently infecting viruses can be elucidated. To achieve this, we optimized DNA isolation procedure from cotton leaves. The primers were designed for actin gene from cotton genome as internal positive control as well as on conserved regions of CLCuV DNA-β. The PCR assay was also optimized for different annealing temperatures so that a single specific band suitable for cloning purposes can be obtained. The PCR assay showed best results when primer pair CLCuV DNA-β was used at 60°C (Figure 4.4).

These products from two isolates obtained from Bahawalnagar district were further ligated and cloned in vector for sequencing. The resulted sequence was analyzed to determine the phylogenetic relationships of these isolates with already available CLCuV DNA-β from other crops and countries around the world.

The study was carried out to find out the evolutionary study of β DNA of cotton leaf curl virus across the global level and to find out the depth knowledge about the status of β DNA of cotton leave curl virus associated begomovirus in Pakistan. The CLCuV sample was carried out and approximately 1300 base pair gene was amplified and sequenced. The amplified gene was aligned and it was highly similar with CLCuMV-PK-AM712320, CLCuMV-PK-AM712315.

The sequences of CLCuv β DNA downloaded from NCBI and accession numbers with regional detail given in Appendix 3. The phylogenetic analysis was carried out by Mega 4 shows in Fig 4.9.

The phylogenetic analysis using NJ tree (Figure 4.9) indicated that DNA-β isolated from different crops like cotton, okra and tomatoes splits into two major groups. One the bases of regions to which isolates from these groups belonged, these groups were named as Indo-Africa and Indo-China. The NJ tree also indicated five further subgroups or clusters, where isolates showed very strong clustering pattern based on geographic regions.

Cluster I mainly comprises of DNA-β isolates from Pakistan origin with few exception of DNA-β from Indian origin. The isolates in Cluster I are majorly isolated from cotton host. The DNA- in cluster I were associated with Cotton Leaf curl Multan virus.Cluster I mostly contain viruses from different cotton producing area of Pakistan.

Our finding lies in cluster I and is closely related with CLCuMV-PK-AM712320, CLCuMV-PK-AM712315 Pakistani type.CLCuBV isolated from bahawalnagar .but due to many reason it spread to other adjacent area very quickly.many factor facilitate the spread which include human transportation,cotton material,,increase cultivation and transportation of cotton germplasm for high productivity. Firstly the cotton crop was introduced in subcontinent in 1818.initially the subcontinent was free from CLCuD. But

when the cotton requirement is increase so the former adopted high productivity cotton from NW.the attack of CLCuV was reported (HafizurRahman *et al.*, 2001;Briddon and Markhan 2000).the result from phylogenetic analysis support the idea that virus easily attack where there host is abundant. NJ tree of all cotton leave curl virus was divided in to five clusters. Culster I (with 95% bootstrap support) consist of already reported cotton leaf curl virus isolates mostly from different cotton production region of Pakistan i.e Multan,Faisalabad.Rahimyarkhan,Tandoadam.our isolates lie in this cluster.

Here the result showed that CLCuMV (cluster 1V) are quite for away from CLCuV-Pk-Bahawalnagar-β-2011-1and CLCuV-Pk-Bahawalnagar-β-2011-2 (cluster 1) in the phylogenetic tree.

This study revealed that cluster I or Pakistani cluster I (with 95% bootstrap support) is consisting of CLCuV-Pk-Bahawalnagar isolates. The cluster also shows that this isolate is found in other regions of the world specifically in India. Some isolates grouped into "Cluster I" which was reported from different regions of India . The study isolates CLCuV-Pk-Bahawalnagar 2011-1 and CLCuV-Pk-Bahawalnagar 2011-2 are clustered into this group. It means that these have close similarity with the member of this group. The clustering pattern showed that these isolate belong to CLCuMV. Geographically Bahawalnagar is laid down in cotton producing region. The emerging of resistant specie of CLCuV-Pk-Bahawalnagar much more affects the cotton and spread in the area. These isolated species are much more close to the isolates from Multan and Faisalabad.so any one of them or recombinant of these both isolates may generate these new isolates. CLCuV-Pk-Bahawalnagar-β-2011-1is closely related to CLCuMV-PK(AM712320) and CLCuMV-PK (AM712314) and has a 98.4% similarity with CLCuMV-PK(AM712320) and has a 98.6%with CLCuMV-PK (AM712314).similarlyCLCuV-Pk-Bahawalnagar-β-2011-2 is closely related with CLCuV-PK-F(FN432359) and CLCuMV-PK(AM712315) and has a 98.3% similarity with CLCuV-PK-F(FN432359) and 98.4% similarity with CLCuMV-PK(AM712315).

Cluster II (with 100% bootstrap support) has isolates from indopak,mostly contain isolate from lucknow,Haryana from india and lyyah,Faisalabad and vehari from Pakistan.In

cluster II most of the isolates belonged to India with three from Pakistani origin. The isolates in Cluster II are majorly isolated from cotton host.

Culster III (Fig 4.8) (with 100% bootstrap) consist of cotton leave curl Geziravirus isolate mostly from the old world mostly from Africa like Kampala (Burkina Faso), Gezira (sudan), Sadore (niger) and Egypt.earlier CLCuVG isolated from Burkina Faso was clustered into same group (G1-G4) On the prescence and absence of recombination events (Tiendrebeogo *et al.*, 2010).these viruses have different host and can infect cotton, okra and tomato.

Culster IV (with 100% bootstrap) contain isolate from Multan (Pakistan).these isolate are slightly differ from cluster 1 pakistani isolate.this cluster have also Indian and Bangladesh isolate.the isolate of the cluster1V can also infect tomato beside cotton containing CLCuMV-PK,TLCuV-B-G isolate.

Culster V (with 99% bootstrap) contain isolate from China .the virus belonging from cluster V can cause infection in tomato.

Our analysis of *begomoviruse* showed that the virus cluster by geographical distribution. It is seen that the most of viruses which lie in same geographical area have same similar evolutionary rates. But there are some isolates which don't follow the same distribution. These abnormalities can be due to various factors.

Our analysis showed that the higher frequency in begomovirus is due to the large number of species in this genera.*begomoviruses* is a large group of viruses infecting host in broad spectrum.they infect many plant and vegetable crop.the host specificity in begomovirus is very broad and one virus can infect many host species from different families (sanz *et al.*, 2000).co infection is very common in begomoviruses which increase the possibility that different genes of co infecting viruses may recombine in a new way resulting in to the development of novel specie (Padidam *et al.*, 1999).

It was concluded that,

In this study we optimized DNA isolation procedure, designed primers for amplification of CLCuV DNA-β and cotton nuclear actin gene as internal positive control. We further cloned and sequenced DNA-β from two different isolates belonging to Bahawalnagar; a cotton producing district in Punjab, Pakistan. The phylogenetic analysis indicated two major groups of DNA-β worldwide. One containing isolates from Pakistan, India and Africa, while other majorly from China and few from Pakistan and Bangladesh. We further explored five distinct subgroups or clusters where isolates showed very strong relationships based on geographic regions. Pakistani isolates sequenced in this study belonged to major cluster Indo-Africa and within it to cluster I.The phylogenetic analysis revealed that DNA-β is evolving geographically where it could be associated with different begomoviruses, indicating mobile nature of this DNA which is consistent with the view of viral fluid genomes (Marinho *et al.*,2008) where viruses evolve around the core proteins, here DNA-A encoded proteins in case of DNA-β. Interestingly our analysis indicates the presence of two different clusters of DNA-β in Pakistan which necessitate further sequencing of more isolates from different cotton growing areas of Pakistan so that a comprehensive picture can be elucidated about the evolution of DNA-β in this region.

# *References*

# REFERENCES

**Ahmad, S., Noor-ul-islam., Mahmood, A., Ashraf, F., Hayat, H. and Hanif, M.** 2010. Screening of Cotton Germplasm against Cotton leaf curl virus, Pak. J. Bot; 42(5): 3327-3342.

**Akhtar, K. P., Azlam, M., Haq, M. A., Jamil, F. F., Khan, A. I. and Elahi, M. T.** 2005. Plant Pathology and Nematology Resistance to Cotton Leaf Curl Virus (CLCuV) in a Mutant Cotton Line. J. Cotton Sci; 9: 175-181.

**Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nuc. Acids. Res; 25(17): 3389-402.

**Ana, I. S., Aurora, F., Jesus, M., Gallego, J. Malpica, M., Fernando, G. A.** 1999. Genetic Variability of Natural Populations of Cotton Leaf Curl Geminivirus, a Single-Stranded DNA virus. J. Mol. Evol; 49: 672-681.

**Arguello-Astorga, G. R., Ruiz-Medrano, R.** 2001. An iteron-related domain is associated to motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. Arch Virol; 146: 1465-1485.

**Aslam, M., Jiang, C., Wright, R. and Paterson, A. H.** 2000. Identification of molecular markers linked to leaf curl virus disease resistance in cotton. J. Sci. I. R. Iran; 11(4):

**Balaji, V., Vanitharani, R., Karthikeyan, A. S., Anbalagan, S. and Veluthambi, K.** 2004. Infectivity analysis of two variable DNA B components of Mungbean yellow mosaic virus-Vigna in Vignamungoand Vignaradiata. J. Biosci; 29: 297-308.

**Baliji, S., Black, M. C., French, R., Stenger, D. C. and Sunter, G.** 2004. Spinach curly top virus: a newly described Curtovirus species from southwest Texas with incongruent gene phylogenies. Phytopathology; 94: 772-779.

**Baulcombe, D.** 2004. RNA silencing in Plants. Nature; 431: 356-363.

**Bosque-perez, N. A.** 2000. Eight decade of maize streak virus research. Virus Res; 71: 107-21.

**Boulton, M. I., King, D. I., Donson, J. and Davies, J. W.** 1991. Point substitutions in a promotor-like region and the V1 gene affect the host range and symptoms of maize streak virus. Virology; 183: 114-121.

**Briddon, R. W.** 2003. Cotton leaf curl disease, a multicomponent begomovirus complex. Mol. Plant. Pathol; 4: 427-434.

**Briddon, R. W., Bedford, I. D., Tsai, J. H., Markham, P. G.** 1996. Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. Virology; 219: 387-394.

**Briddon, R. W., Stenger, D. C., Bedford, I. D., Stanley, J., Izadpanah, K. and Markham, P. G.** 1998. Comparisons of a Beet curly top virus isolate originating from the Old World with those from the New World. Euro. J. Plant. Pathol; 104: 77-84.

**Brown, J. K.** 1994. The status of Bemisia tabaci (Genn.) as a pest and vector in world agroecosystems. FAO. Plant. Prot. Bull; 42: 3-32.

**Brown, J. K. and Nelson, M. R.** 1984. Geminate particles associated with cotton leaf crumple disease in Arizona. Phytopathology; 71: 699-703.

**Brown, J. K., Khan, J. A.** 2001. The molecular epidemiology of begomoviruses; Haworth Press, Inc., New York. J. Trends in Plant. Virology; 279-316.

**Buchmann, R. C., Asad, S., Wolf, J. N., Mohannath, G. and Bisaro, D. M.** 2009. Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. J. Virol; 83: 5005-5013.

**Castillo, A. G., Collinet, D., Deret, S., Kashoggi, A. and Bejarano, E. R.** 2003. Dual interaction of plant PCNA with geminivirus replication accessory protein (REn) and viral replication protein (Rep). Virology; 312: 381-394.

**Castillo, A. G., Collinet, D., Deret, S., Kashoggi, A. and Bejarano, E. R.** 2003. Dual interaction of plant PCNA with geminivirus replication accessory protein (REn) and viral replication protein (Rep). Virology; 312: 381-394.

**Choudhury, N. R., Malik, P. S., Singh, D. K., Islam, M. N., Kaliappan, K. and Mukherjee, S. K.** 2006. The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. Nuc. Acids. Res; 34: 6362-6377.

**Collin, S., Fernandezlobato, M., Gooding, P. S., Mullineaux, P. M. and Fenoll, C.** 1996. The two non-structural proteins from Wheat dwarf virus involved in viral gene expression and replication are retinoblastoma-binding proteins. Virology; 219: 324-329.

**Dickinson, V. J., Halder, J. and Woolston, C. J.** 1996. The product of Maize streak virus ORF V1 is associated with secondary plasmodesmata and is first detected with the onset of viral lesions. Virology; 220: 51-59.

**Ding, S. W., Li, H., Lu, R., Li, F. and Li, W. X.** 2004. RNA silencing: a conserved antiviral immunity of plants and animals. Vir. Res; 102: 109-115.

**Eagle, P. A., Orozco, B. M. and Hanley-Bowdoin, L.** 1994. A DNA sequence required for geminivirus replication also mediates transcriptional regulation. Plant Cell; 6: 1157-1170.

**Endrizzi, J. E., Turcotte, E. L. and Kohel, R. J.** 1985. Genetics, cytology and evolution of Gossypium. Adv. Genet: 23: 271-327.

**Fauquet, C. M.** 2002. The best sequence threshold criterion for geminivirus species demarcation. www.danforthcenter.org/iltab/geminiviridae/spdem.

**Fauquet, C. M., Sawyer, S., Idris, A. M. and Brown, J. K.** 2005. Sequence analysis and classification of apparent recombinant begomoviruses infecting tomato in the Nile and Mediterranean Basins. Phytopathology; 95: 549-555.

**Felsenstein, J.**1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution;39: 783-791.

**Fontes, E., Gladfelter, H. J., Schaffer, R. L., Petty, I. and Hanlly-Bowoin, L.** 1994. Geminiviradae replication origins have a molecular organization. Plant Cell; 6: 405-416.

**Gennadius, P.** 1889. Disease of the tobacco plantations in the Trikonia: The aleurodid of tobacco. Ellenike Georgia; 5: 1-3.

**Goodman, R. M.** 1981. Geminivirus in handbook of plant virus infection and Comparative Diagnosis. Elsevier; 879-910.

**Hamilton, W. D. O., Bisaro, D. M., Coutts, R. H. A. and Buck, K. W.** 1983. Demonstration of the bipartite nature of the genome of a single-stranded DNA plant virus by infection with the cloned DNA components. Nucl. Acids. Res; 11: 7387-7396.

**Hamilton, W. D. O., Stein, V. E., Coutts, R. H. A. and Buck, K. W.**1984.Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: Potential coding regions and regulatory sequences. EMBO. J; 3: 2197-2205.

**Hanley-Bowdoin, L., Settlage, S. B., Orozco, B. M., Nagar, S., Robertson, D.** 1999. Geminviruses: models for plant DNA replication, transcription, and cell cycle regulation. Crit. Rev. Plant. Sci; 18: 71-106.

**Howarth, A. J., Caton, J., Bossert, M. and Goodman, R. M.** 1985. Nucleotide sequence of bean golden mosaic virus and a model for gene regulation in Gemini virus. Proceeding of the National Academy of Science, U.S.A. 82: 3572-3576.

**http://ioupakistan.blogspot.com/2009/09/major-crops-of-pakistan.html**

**http://www.fibre2fashion.com/industry-article/15/1495/popup/popup1.png** cotton annual production and comparasion with other countries. Cited on 27.07.2011.

http://rehmananwar.blogspot.com/2010/03/agriculture-in-pakistan.html.

**Hur, J., Kenneth, J. B., Lee, S. and Keith, R. D.** 2007. Transcriptional activator elements for Curtovirus C1 expression reside in the 3′ coding region of ORF C1. Molecules and Cells; 23: 80-87.

**Hur, J., Kenneth, J. B., Lee, S. and Keith, R. D.** 2007. Transcriptional activator elements for Curtovirus C1 expression reside in the 3′ coding region of ORF C1. Molecules and Cells; 23: 80-87.

**Hussain, M., Mansoor, S., Iram, S., Zafar, Y. and Briddon, R. W.** 2007. The hypersensitive response to Tomato leaf curl New Delhi virus nuclear shuttle protein is inhibited by transcriptional activator protein. Mol. Plant. Micro. Inter; 20: 1581-1588.

**Hussain, T. and Ali, M.** 1975. A review of cotton diseases of Pakistan. Pak. Cotton; 19: 71-86.

**Inouye, T. and Osaki, T.** 1980. The first record in the literature of the possible plant virus disease that appeared in Manyoshu, a Japanese classic anthology, as far back as the time of the 8th century. Annual report of Phtopathological Society of Japan; 46: 49-50.

**James, J., Campanella.,Adebanke, F., Olajide., Volker, M. and Jutta, L. M.** 2003. A Novel Auxin Conjugate Hydrolase from Wheat with Substrate Specificity for Longer Side-Chain Auxin Amide Conjugates. Plant Physiol; 135: 2230-2240.

**Kirkpatricka, T. W.** 1931. Further Studies on Leaf-curl of Cotton in the Sudan.

**Koonin, E. V. and Ilyina, T. V.** 1992. Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. J. Gen. Virol; 73: 2763-2766.

**Kunik, T., Palanichelvam, K., Czosnek, H., Citovsky, V. and Gafni, Y.** 1998. Nuclear import of the capsid protein of Tomato yellow leaf curl virus (TYLCV) in plant and insect cells. Plant J. 13: 393-399.

**Lam, N., Creamer, R. R., Belfon, R.** 2009. Characterization of a new curtovirus, pepper yellow dwarf virus, from chile pepper and distribution in weed hosts in New Mexico. Arch. Virol; 154: 429-436.

**Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F. and Gronenborn, B.** 1995. Geminiviradae replicaton: Genetic and Biochemical characterization of Rep protein function. Rev. Biochem; 77:765-773.

**Lazarowitz, S. G. and Lazdins, I. B.** 1991. Infectivity and complete nucleotide sequence of the cloned genomic components of a bipartite squash leaf curl geminivirus with a broad host range phenotype. Virology; 180: 58-69.

**Lazarowitz, S. G., Wu, L. C., Rogers, S. G. and Elmer, J. S.** 1992. Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. Plant Cell; 4: 799-809.

**Liu, H., Boulton, M. I., Thomas, C. L., Prior, D. A., Oparka, K. J. and Davies, J. W.** 1999. Maize streak virus coat protein is karyophyllic and facilitates nuclear transport of viral DNA. Mol. Plant. Micro. Inter; 12: 894-900.

**Luria, S. E., Darnell, J. E., Baltimore, D. and Campbell, A.** 1978. General Virology, 3rd (edn). Wiley.

**Mansoor, S., Amin, I., Iram, S., Hussain, M., Zafar, Y., Malik, K. A. and Briddon. R. W.** 2003. The breakdowns of resistance in cotton to cotton leaf curl disease in Pakistan. New Disease Reports; 7: 10-20.

**Moffatt, B. A., Stevens, Y. Y., Allen, M. S., Snider, J. D., Periera, L. A., Todorova, M. I., Summers, P. S., Weretilnyk, E. A., Martin-Caffrey, L. and Wagner, C.** 2002. Adenosine kinase deficiency is associated with developmental abnormalities and reduced transmethylation. Plant. Physio; 128: 812-821.

**Noueiry, A. O., Lucas, W. J. and Gilbertson, R. L.** 1994. Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. Cell; 76: 925-932.

**Onuki, M. and Hanada, K.** 2000. Genomic Structure of a geminivirus in the genus Begomovirus from yellow Vein-affected Eupatorium makinoi. J. Gen. Plant. Pathol. 66: 176-181.

**Padidam, M., Sawyer, S. and Fauquet, C. M.** 1999. Possible emergence of new Geminiviruses by frequent recombination. Virology; 265: 218-225.

**Padidam, M., Sawyer, S., Fauquet, C. M.** 1999. Possible emergence of new geminiviruses by frequent recombination. Virology; 265: 218-225

**Paximadis, M., Idris, A. M., Torres-Jerez, I., Villarreal, A., Rey, M. E. and Brown, J. K.** 1999. Characterization of tobacco geminiviruses in the Old and New World. Arch. Virol; 144: 703-717.

**Pretheep-Kumar, P., Mohan, S. and Ramaraju, K.** 2004. Protein-enriched pea flour extract protects stored milled rice against the rice weevil, Sitophilusoryzae. J. Insect. Sci; 4: 26, Available online: insectscience.org/4.26.

**Pringle, C. R.** 1999. Virus taxonomy: The universal system of virus taxonomy, updated to include the new proposals ratified by the International Committee on Taxonomy of Viruses. Arch. Virol; 144: 421-429.

**Rigden, J. E., Dry, I. B., Mullineaux, P. M. and Rezaian, M. A.** 1993. Mutagenesis of the virion-sense open reading frames of Tomato leaf curl virus. Virology; 193: 1001-1005.

**Rigden, J. E., Krake, L. R., Rezaian, M. A. and Dry, I. B.** 1994. ORF C4 of tomato leaf curl geminivirus is a determinant of symptom severity. Virology; 204: 847-850.

**Rob, W., Briddon., Basavaprabhu, L. P., Basavaraj., Bagewadi., Muhammad, S. N., Claude, M. F., Aslam, M., Jiang, C., Wright, M. and Paterson, A. H.** 2000. Identification of molecular markers linked to leaf curl virus disease resistance in cotton. J. Sci. I. R. Iran; 11(4): 234-242.

**Rojas, M. R., Jiang, H., Salati, R., Xoconostle-Cazares, B., Sudarshana, M. R., Lucas, W. J. and Gilbertson, R. L.** 2001. Functional analysis of proteins involved in movement of the monopartite begomovirus, Tomato yellow leaf curl virus. Virology; 291: 110-125.

**Rybicki, E. P.** 1994. A phylogenetic and evolutionary justification for 3 genera of geminiviridae. Arch. Virol; 139: 49-77.

**Rybicki, E. P., Briddon, R. W., Brown, J. E., Fauquet, C. M., Maxwell, D. P., Harrison, B. D., Markham, P. G., Stanley, J.** 2000. Geminiviridae. In: van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens

**Saitou, N. and Nei, M.**1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Bio. Evo;4: 406-425.

**Sambrook, J., Russell, D. W.** 2001. Molecular Cloning: A laboratory manual, 3rd ed, Cold Spring Harbor Laboratory Press, New York.

**Sanz, A. I., Fraile, A., Garcia-Arenal, F., Zhou, X., Robinson, D. J., Khalid, S., Butt, T. and Harrison, B. D.** 2000. Multiple infection, recombination and genome relationship among *begomovirus* isolates found in cotton and other plants in Pakistan. J. General. Virol; 81: 1839-1849.

**Seal, S. E., Vandenbosch, F. and Jeger, M. J.** 2006. Factors influencing begomovirus evolution and their increasing global significance: Implications for sustainable control. Critic. Rev. Plants. Sci; **25:** 23-46.

**Siobain, D. and Edward, C. H.** 2008. Phylogenetic Evidence for Rapid Rates of Molecular Evolution in the Single-Stranded DNA Begomovirus Tomato Yellow Leaf Curl Virus. J. Virol; 82(2): 957-965.

**Stanley, J., Bisaro, D. M., Briddon, R. W., Brown, J. K., Fauquet, C. M., Harrison, B. D., Rybicki, E. P., Strenger, D. C.** 2005. Geminiviradae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (edn), Virus Taxonomy. 8$^{th}$ Report of the ICTV, London, UK. Elsevier; 301-326.

**Stanley, J., Markham, P. G., Callis, R. J. and Pinner, M. S.** 1986. The nucleotide sequence of an infectious clone of the geminivirus Beet curly top virus. EMBO. J; 5: 1761-1767.

**Stenger, D. C.** 1994. Complete nucleotide sequence of the hypervirulent CFH strain of Beet curly top virus. Mol. Plant. Micro. Inter; 7: 154-157.

**Sudarshana, M. R., Wang, H. L., Lucas, W. J. and Gilbertson, R. L.** 1998. Dynamics of bean dwarf mosaic geminivirus cell-to-cell and long-distance movement in Phaseolus vulgaris revealed, using the green fluorescent protein. Mol. Plant. Micro. Inter; 11: 277-291.

**Sunter, G., Hartitz, M. D., Hormuzdi, S. G., Brough, C. L. and Bisaro, D. M.** 1990. Genetic analysis of Tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for efficient DNA replication. Virology; 179: 69-77.

**Sunter, G., Sunter, J. L. and. Bisaro, D. M.** 2001. Plants expressing Tomato golden mosaic virus AL2 or Beet curly top virus L2 transgenes show enhanced susceptibility to infection by DNA and RNA viruses. Virology; 285: 59-70.

**Tahir, M. N., Amin, I., Briddon, R. W., Mansoor, S.** 2011. The Merging of Two Dynasties identification of an African Cotton Leaf Curl Disease-Associated Begomovirus with Cotton in Pakistan. PLOS ONE 6(5): e20366. doi:10.1371/journal.pone.0020366.

**Tamura, K., Dudley, J., Nei, M. and Kumar, S.**2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Bio. Evo;24: 1596-1599.

**Tamura, K., Dudley, J., Nei, M. and Kumar, S.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Bio. Evo; 24: 1596-1599.

**Tamura, K., Nei, M. and Kumar, S.**2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA); 101: 11030-11035.

**Tatusova, T. A. and Madden, L. T.** 1999. Blast 2 Sequence, a new tool for comparing protein and nucleotide sequence. FEMS Microbiology letters; 247-450.

**Tatusova, T. A. and Madden, T. L.**1999. Blast 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS. Micro. Lett; 174: 247-250.

Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, New York, pp. 285-297.

**Thompson., Thompson, J. D., Higgins, D. G. and Gibson, T. J.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nuc. Ac. Res; 22: 4673-4680.

**Unseld, S., Hohnle, M., Ringel, M. and Frischmuth, T.** 2001. Subcellular targeting of the coat protein of African cassava mosaic geminivirus. Virology; 286: 373-383.

**Vanitharani, R., Chellappan, P., Pita, J. S. and Fauquet, C.** 2004. Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. J. Virol; 78: 9487-9498.

**Voinnet, O.** 2005. Induction and suppression of RNA silencing: insights from viral infections. Nat. Rev. Gen; 6: 206-220.

**Wang, H., Buckley, K. J., Yang, X., Buchmann, R. C. and Bisaro, D. M.** 2005. Adenosine kinase inhibition and suppression of RNA silencing by geminivirus AL2 and L2 proteins. J. Virol; 79: 7410-7418.

**Wang, H., Hao, L., Shung, C. Y., Sunter, G. and Bisaro, D. M.** 2003. Adenosine kinase is inactivated by geminivirus AL2 and L2 proteins. Plant Cell; 15: 3020-3032.

**Zhang, H. S. and Dean, D. C.** 2001. Rb-mediated chromatin structure regulation and transcriptional repression. Oncogene; 20: 3134-3138.

# *Appendix*

# APPENDIX

**Appendix 1:** Nucleotide sequence of study Isolate

>CLCuV-Pk-Bahawalnagar-β-2011-1

ACCGTGGGCGAGCGGTGTCCGATGGCTTCTTGGTGGGTCCCATTGCTGGTATTGACT
TGATTTGACTTATATTGGGCCATTTAATGGGTTGAAAATGTTTGGGCCTTTGCAAGA
GGGTCTTTTATGGATAGGCCTCTGTATATTGTTGGTCATATGTGTGTTTAATATGCA
TTGCTCGATTGTGTTTGGAATTTAAACGGTGTACTTTTTATTGTATACGTACGGTTC
GTTTACATCCATTCCCAATATCTCTGGGTTTTCAAGTACAAGTATATCAAGTCTGTG
AACTATATCTTCTATCTCGATCTCTTCTATCTTTGCCCCGTTGTATGCGAATAGGAA
ATTCGCTATGATGCTCCCTTCAAAGCCGTTGAAGTCGAATGGAACGTGTATGTCTTC
GTACGTGTACTGGACGATCCCTTCATACTTGATTAGCGATGGTGACTTTGTGGATAG
TATCCTCATGTGAATGAAGATCTTCATATTCTCCATGATGCGAACGTCGACTATGAA
CCTGACTCCCTCCTTGTTTGTTCCGCTCGTTGTCATTTCTGCTTATTTGATGGGAAT
GTAAAGTGTGCATGGAAATATACACTTAATGAACGACATATTTTGTGTAGTTGTGTG
GTTGTGAGTGATTTATTATTATGTGATTGTCCATTAAAGGGATAAAGTGATGATGGA
GACGTATTACACGTGTTGTCATGTTGGCTTGAAATCTTTATACATGGGTTTGTACCG
TATATACCTATATATACGGATAGAAAACGGATGAGAAAAGGAAAAACGGAACTGAAG
AGGAAAAACAAAGAAAAGAAACAAGGATATATTATTTATGAAAGAAATGGAGCGCA
GCGAATCGAAACAGGAAAACCCAAGGAAAGAGAAAAAATAAAAAGTAAAGAGAAAAA
AAATACAAATTCGAAAACGTCATCGTTTGAGAGTAATAAAAAAAAAGAAAAAACAAA
AACATATCCGAAAACGTTGTCGTTTGAAGGTGACTGTGTGGTTTTACCATTTACTGT
GTGGTAAATGGTAAGTGATTGAAGAATAGTTAAAAAATGGAGACACCGATAGGTAAA
TTGTACCCCAATATATTGGGGTACAATTGGGGACTCATAAATTGCCTTTCCTAAAAT
ACCCCCGCTTTTGTGTCTAAGAGGCGCGTCGGAGTGCGCCTATAAAGTTAACATTCT
CTCTCCTCTTTTGATCTCCAATACAATTTCCCGGTGATCGGAGTCGAATTTTCCGAC
ACGCGCGGCGGTGTGTACCCCTGGAGGGTAGGTACCACTACGCTACGCAGCAGCCT
TAGCTACGCCGGAGCTTAGCTCGCCCACGTTCTAATATT

## Appendix 2:Nucleotide sequence of study Isolate

> CLCuV-Pk-Bahawalnagar-β-2011-2

ACCGTGGGCGAGCGGTGTCCGATGGCTTCTTGGTGGGTCCCATTGCTGGTATTGACT
TGATTTGACTTATATTGGGCCAATTTAATGGGTTGAAAATGTTTGGGCCTTTGGAAG
AGGGTCTTTTATGGAATAGGGCTCTGTATATTGTTGGTGATATGTGTGTTATATATG
CATAGCTGGTAAGTGTTTGGAATATTAACGGTGAACTAATTATTGAATACGTACGGT
TCGTTTACATCCATTCCCAATATCTCTGGGTTTTCAAGTACAAGTATATCAAGTCTG
TGAACTATATCTTCTATCTCGATCTCTTCTATCTTTGCCCCGTTGTATGCGAATAGG
AAATTCGCTATGATGCTCCCTTCAAAGCCGTTGAAGTCGAATGGAACGTGTATGTCT
TCGTACGTGTACTGGACGATCCCTTCATACTTGATTAGCGATGGTGACTTTGTGGAT
AGTATCCTCATGTGAATGAAGATCTTCATATTCTCCATGATGCGAACGTCGACTATG
AACCTGACTCCCTCCTTGTTTGTTCCGCTCGTTGTCATTTCTGCTTATTTGATGGAA
TTGTTTAGTGTGCATGCATTTATAGACTTAATGAACAGTTATTAAGAGTAGTTTGGT
GGTTGTGAGTGACGTATTATTATGTGAAACGTCCAAAAAAGGGATAAAGTCATGATG
GAGACGTATTACACGTGTTGTCATGTTGGCTTGAAATCTTTATACATGGGTTTGTAC
CGTATATACCTATATATACGGATAGAAAACGGATGAGAAAAGGAAGAACGGAACTGA
AGAGGAAAAACAAAGAAAAGAAACAAGGATATATTATTTATGAAAGAAATGGGAGCG
CAGCGAATCGAAACAGGAAAACCCAAGGAAAGAGAAAAAATAAAAAGTAAAGAGAAA
AAAAATACAAATTCGAAAACGTCATCGTTTGAGAGTAATAAAAAAAAAGAAAAAACA
AAAACATATCCGAAAACGTTGTCGTTTGAAGGTGACTGTGTGGTTTTACCATTTACT
GTGTGGTAAATGGTAAGTGATTGAAGAATAGTTAAAAAATGGAGACACCGATAGGTA
AATTGTACCCCAATATATTGGGGTACAATTGGGGACTCATAAATTGCCTTTCCTAAA
ATACCCCCGCTTTTGTGTCTAAGAGGCGCGTCGGAGTGCGCCTATAAAGTTAACATT
CTCTCTCCTCTTTTGATCTCCAATACAATTTCCCGGTGATCGGAGTCGAATTTTCCG
ACACGCGCGGCGGTGTGTACCCCTGGAGGGTAGGTACCACTACGCTACGCAGCAGC
CTTAGCTACGCCGGAGCTTAGCTCGCCCACGTTCTAATATT

**Appendix 3: Origin and Accession number DNA β of CLCuD associated *Begomovirus* used in this study**:

| Geographical Distribution | | | Accesion Number | Host |
|---|---|---|---|---|
| Isolate | Region | Country | | |
| B.G | Gazipur | Bangladesh | AJ542489 | Tomato |
| C.Dali198 | | China | EU365687 | Artemisia carvifoliaBuch |
| C.Y238 | Yunnan | China | AM260715 | tobacco |
| C.Y261 | Yunnan | China | AM260719 | tobacco |
| C.Y262 | Yunnan | China | AM260720 | tobacco |
| C.Y263 | Yunnan | China | AM260721 | tobacco |
| C.Y264 | Yunnan | China | AM260722 | Tobacco |
| C.Y265 | Yunnan | China | AM260723 | tobacco |
| C.Y268 | Yunnan | China | AM260724 | tobacco |
| C.Y318 | Yunnan | China | AM260725 | tobacco |
| C.Y319 | Yunnan | China | AM260726 | tobacco |
| C.Y330 | Yunnan | China | AM260727 | tobacco |
| C.Y321 | Yunnan | China | AM260728 | tobacco |
| C.Y323 | Yunnan | China | AM260729 | Siegesbeckiaorientalis |
| C.Y295 | Yunnan | China | AM260730 | Tomato |
| C.Y296 | Yunnan | China | AM260731 | Tomato |
| C.Y297 | Yunnan | China | AM260732 | Tomato |
| C.Y298 | Yunnan | China | AM260733 | Tomato |
| C.G102 | Guangxi | China | AM050556 | Tobacco |
| C.G103 | Guangxi | China | AM050557 | tobacco |
| C.Y253 | Yunnan | China | AJ971331 | Siegesbeckiaorientalis |
| C.Y254 | Yunnan | China | AJ971332 | Siegesbeckiaorientalis |

| C.Y255 | Yunnan | China | AJ971333 | Siegesbeckiaorientalis |
|---|---|---|---|---|
| C.Y284 | Yunnan | China | AJ971334 | Siegesbeckiaorientalis |
| C.Y285 | Yunnan | China | AJ971335 | Siegesbeckiaorientalis |
| C.Y286 | Yunnan | China | AJ971336 | Siegesbeckiaorientalis |
| C.Y10 | Yunnan | China | AJ781297 | |
| C.Y10 | Yunnan | China | AJ781298 | |
| C.Y10 | Yunnan | China | AJ781299 | |
| C.Y10 | Yunnan | China | AJ781300 | |
| C.Y43 | Yunnan | China | AJ781301 | |
| C.G36T-1 | Guangxi | China | AJ704578 | Euphorbia pulcherrima |
| C.G36T-2 | Guangxi | China | AJ704579 | Nicotianatabacum |
| C.Y43 | Yunnan | China | AJ420314 | |
| C. Y38 | Yunnan | china | AJ420315 | |
| I.Har | Hisar, Haryana | India | DQ364230 | Gossypiumhirsutum |
| I.Ban | Banglore | india | NC_007219 | Gossypiumhirsutum |
| S.G | Gezira | sudan | NC_006935 | Cotton |
| P. | | Pakistan: | NC_003200 NC_013468 | Gossypiumhirsutum |
| P.M | Multan | Pakistan: | AM712311 | G. annumalum |
| P.M | Multan | Pakistan: | AM712312 | G. annumalum |
| P.M | Multan | Pakistan | AM712313 | G. annumalum |
| P.M | Multan | Pakistan | AM712314 | G. annumalum |
| P.M | Multan | Pakistan | AM712315 | G. annumalum |
| P.M | Multan | Pakistan | AM712316 | G. annumalum |
| P.M | Multan | Pakistan | AM712317 | Gossypiumsomalense |
| P.M | Multan | Pakistan | AM712318 | Gossypiumsomalense |
| P.M | Multan | Pakistan | AM712319 | G.letifolium |

| P.M | Multan | Pakistan | AM712320 | G.barbadense |
|---|---|---|---|---|
| P.M | Multan | Pakistan | AM712321 | G. barbadense |
| P.M | Multan | Pakistan | AM712322 | G. barbadense |
| P.M | Multan | India | AM712323 | G. barbadense |
| I.H | Hissar,Haryana | india | AY763123<br>AF534186 | |
| | | | AJ421678 | |
| | | India | AY083590 | |
| I.L | Lucknow (U.P.) | | HM143916 | Nicotianatabacum<br>(cv. White Burley) |
| I.L | Lucknow | india | GU440581 | cotton |
| BF.Tie.Okra2 | | Burkina Faso | FN554573 | Abelmoshusesculentus |
| BF.Kap.Okra3 | Kampala | Burkina Faso | FN554576 | Abelmoshusesculentus |
| BF.Kap.Okra5 | Kampala | Burkina Faso | FN554577 | Abelmoshusesculentus |
| N.S | Sadore(45km from Niamey) | Niger: | FJ469628 | Okra |
| N.S | Sadore(45km from Niamey | Niger: | FJ469629 | Okra |
| N.S | Sadore(45km from Niamey | Niger: | EU727087 | Okra |
| N.S | Sadore(45km from Niamey | Sudan | EU727088 | Okra |
| S. | | Mali | DQ644564 | Datura 1 |
| M.B | Bamako | Mali | EU024121 | Abelmoschusesculentus |
| M.B | Bamako | China: | EU024122 | Abelmoschusesculentus |
| C.G87 | Guangxi | Mali | AM072289 | Malvastrumcoromandelianum |
| M. | | Sudan | DQ136001 | |
| S.CLC46-C | | Egypt | AY669328 | cotton |
| S | | Sudan | AJ316039 | Hibiscus esculentis |

| S.G | Gezira | Sudan | AY077797 | cotton |
|---|---|---|---|---|
| S.CLC32-S | Gezira | Sudan | AY077798 | Sida |
| S.CLC43-O | Gezira | | AY077799 | okra |
| | | | AF397215 | |
| | | Sudan | AF397217 | |
| S.G | Gezira | Sudan | AY044140 | |
| S.G | Gezira | Sudan | AY044141 | okra |
| S.G | Gezira | Sudan | AY044142 | okra |
| S.G | Gezira | Pakistan | AY044143 | |
| P.S | Sakrand | Pakistan | FN554719 | |
| P.S | Sakrand | Pakistan | FN554720 | |
| P.S | Sakrand | Pakistan | FN554721 | |
| P.S | Sakrand | Pakistan | FN554722 | |
| P.SH | Shadadpur | Pakistan | FN554723 | |
| P.TA | Tandoadam | Pakistan | FN554724 | |
| P.TJ | Tandojam | Pakistan | FN554725 | |
| P.FA | Faisalabad,Samundri | Pakistan | FN432359 | Sonchusarvensis |
| P.FA | Faisalabad | Pakistan | AM774306 | |
| P.V | Vehari | Pakistan | AM774307 | |
| P.R | Rahim Yar Khan | Pakistan | AM774308 | |
| P.LY | Layyah" | Pakistan | AM774309 | |
| P.K | Khanewal | Pakistan | AM774310 | |
| P.MA | Muzaffarabad | Pakistan | AM774311 | |
| P.MG | Muzaffargarh | Pakistan | AM774312 | |
| P.A | Arifwala | India | AM774313 | |
| I.LU | Lucknow" | india | GQ369731 | Crotalaria juncea |
| I.R | East India, Raigunj | India | FJ159274 | Hibiscus cannabinus |

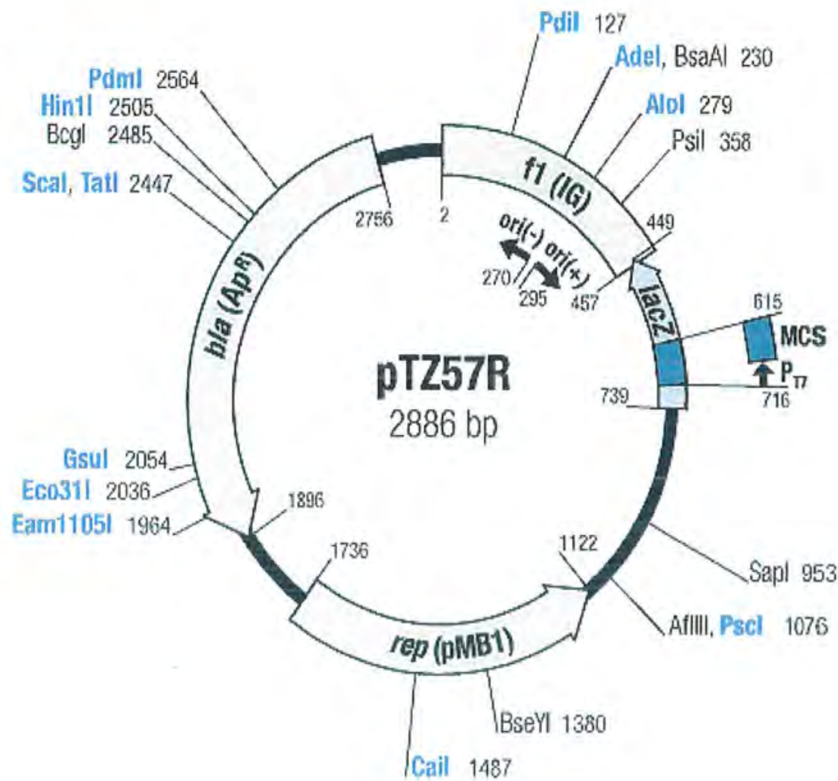
| I.B | Balurghat:East India | Niger: | FJ159275 | Hibiscus cannabinus" |
|---|---|---|---|---|
| N.S | Sadore | Niger: | EU727087 | okra |
| N.S | Sadore | | EU727088 | okra |
| I.B | East India, Bongaon | India | EF614158 | Hibiscus cannabinus |
| I.H | East India, Haringhata" | India | EF614159 | Hibiscus cannabinus |
| I. | | India | NC_007219 | |
| I.B | BANGLORE | India | AY705381 | Gossypiumbarbadense |
| I. | | India | AJ316037 | Gossypiumhirsutum" |
| I. | | | AJ3160380 | Gossypiumhirsutum |

## Appendix 4

Primer used for amplification of cotton leaf curl virus β DNA.

| | |
|---|---|
| UniCLCuV-β-forward primer | CTACGCCGGAGCTTAGCTCG |
| UniCLCuV-β-reverse primer | GCTAAGGCTGCTGCGTAGCG |

**Appendix 5: Map of pTZ57R and its multiple cloning sites**



Schematic representation of pTZ57R vector, showing the position of important genes and regulatory parts. Below is the sequence and restriction enzymes site which have their position in multiple cloning sites. Image is taken from Fermentas web site at http://www.fermentas.com/techinfo/nucleicacids/mapptz57r.htm