Detection, Amplification and Sequence Analysis of Begomovirus(es) from different Plant Species

By

Salliha Khalid

Registration # 00000364461

Master of Science in Plant Biotechnology

Supervisor

Prof. Dr. Muhammad Tahir

Atta-Ur-Rahman School of Applied Biosciences

National University of Sciences & Technology

Islamabad, Pakistan

2023

Detection, Amplification and Sequence Analysis of Begomovirus(es) from different Plant Species

By:

Salliha Khalid

Registration # 00000364461

A thesis submitted in partial fulfilment of the requirement for the degree of Master of Science

in

Plant Biotechnology

Under the Supervision of:
Prof. Dr. Muhammad Tahir

Professor, Department of Plant Biotechnology

Atta-Ur-Rahman School of Applied Biosciences

National University of Sciences & Technology

Islamabad, Pakistan

2023

Thesis Acceptance Certificate

It is certified that final copy of MS Thesis written by Salliha Khalid (Registration no. 00000364461), of Atta-ur-Rahman School of Applied Biosciences (ASAB) has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations/MS policy, is free of plagiarism, errors and mistakes and is accepted as partial fulfilment for award of MS degree.

Supervisor: _____ 03 Aug

Prof. Dr. Muhammad Tahir
Dept of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Prof. Dr. Muhammad Tahir

Professor

Atta Ur Rahman School of Applied Biosciences

National University of Sciences & Technology

Head of Department: _____

Dr. Muhammad Faraz Bhatti
Head of Department (HoD)
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Prof. Dr. Muhammad Faraz Bhatti

Atta Ur Rahman School of Applied Biosciences

National University of Sciences & Technology

Principal: _____

Prof. Dr. Muhammad Asghar
Principal
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Dr. Muhammad Asghar

Atta Ur Rahman School of Applied Biosciences

National University of Sciences & Technology

TH-4 Form

National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation be prepared under our supervision by: Salliha Khalid, Reg No. 00000364461.

Titled: Detection, Amplification and Sequence Analysis of Begomovirus(es) from different Plant Species be accepted in partial fulfillment of the requirements for the award of **MS degree in Plant Biotechnology** with (A grade).

Examination Committee Members

1. Name: Dr. Muhammad Qasim Hayat.
   Department: Plant Biotechnology
   Signature: _____

2. Name: Dr. Najam – us – Sahar Sadaf Zaidi.
   Department: Industrial Biotechnology
   Signature: _____

Supervisor's Name: **Prof. Dr. Muhammad Tahir**

Prof. Dr. Muhammad Tahir
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Signature: _____

Date: 03 Aug 2023

Head of Department: **Dr. Muhammad Faraz Bhatti**

Dr. Muhammad Faraz Bhatti
Head of Department (HoD)
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Signature: _____

Date: 9-8-23

COUNTERSINGED

Prof. Dr. Muhammad Asghar
Principal
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Date: 9-08-23

Dean/Principal

Declaration

I certify that this research work titled "*Detection, Amplification and Sequence Analysis of Begomovirus(es) from different Plant Species*" is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources has been properly acknowledged / referred.

Salliha Khalid

Master of Science in Plant Biotechnology

Reg No. 00000364461

Copyright Statement

Plagiarism Certificate (Turnitin Report)

It is certified that MS Thesis titled "Detection, Amplification and Sequence Analysis of Begomoviruses from different Plant Species" by student Salliha Khalid, Registration Number 00000363051 has been examined by me. I undertake that:

a. This thesis has significant new work/knowledge as compared to already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph, or section has been copied verbatim from earlier work unless it is placed under quotation marks and duly referenced. As, English is not my first language, I have taken the help from AI in rephrasing the acknowledgement section.

b. The work presented is original and own work of the author (i.e., there is no plagiarism). No ideas, processed, results or words of others have been presented as Author's own work.

c. There is no fabrication of data or results which have been compiled/analyzed.

d. There is no falsification by manipulating research materials, equipment or processes or changing or omitting data or results such that research is not accurately represented in the research record.

e. The thesis has been checked using TURNITIN (Copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

Salliha Khalid

Date: 3-08-23

Supervisor: Prof. Dr. Muhammad Tahir

Signature:

Prof. Dr. Muhammad Tahir
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad
2023

My Dissertation is dedicated to the memory of my beloved Baba G (grandfather, late), who held a special place of being my first childhood friend.

Acknowledgement

First, I would like to extend my sincerest gratitude and acknowledgment to the Almighty Allah, the Lord of Lords, whose divine strength has empowered me to diligently complete my work within the given timeframe. May peace and blessings be upon the Prophet (□), whose profound teachings have perpetually served as a guiding light in my life.

I am deeply indebted to my esteemed Professor, **Dr. Muhammad Tahir**, for his selfless dedication and unwavering support. His expertise and wisdom have been instrumental in shaping my academic journey and enhancing my research skills. His selfless dedication and genuine interest in my progress have been a constant source of inspiration. **Throughout my master's journey, he has been my sole pillar of moral encouragement**. I am immensely honored to have had the privilege of being under his esteemed supervision, and I take great pride in calling him my **Ustaad** (Teacher).

Furthermore, I would like to acknowledge the profound influence of Dr. Allama Muhammad Iqbal's mystical poetry, which has consistently inspired me both in my academic endeavors and personal life. In addition to the invaluable contributions of distinguished researchers such as Dr. Stephen Winter, Dr. Judith Brown, and Dr. Rob Briddon in my field, Iqbal's poetry has been an additional wellspring of inspiration for me. I will surely share a few lines here.

زندگانی ہے صدف، قطرہ نیساں ہے خودی وہ
صدف کیا کہ جو قطرے کو گہر کر نہ سکے ہو اگر
خودنگر و خودگر و خودگیر خودی
یہ بھی ممکن ہے کہ تو موت سے بھی مر نہ سکے

(حیات ابدی، ضرب کلیم-025)

(Eternal Life, Zarb-e-Kaleem-025)

I am deeply grateful for the nurturing environment fostered by Atta Ur Rahman School of Applied Biosciences (ASAB), National University of Sciences & Technology (NUST), and the management of Amna Hostel, NUST. I would like to convey my heartfelt appreciation to all the instructors who have played a pivotal role in my journey, including Principal ASAB, **Dr. Hassnain Janjua**, GEC **Dr. Najam Us Sahar Sadaf Zaidi**, HoD Plant Biotechnology Department, **Dr. Muhammad Faraz Bhatti**, **Dr. Muhammad Qasim Hayat**, **Dr. Faiza Munir**, and **Dr. Alvina Gul**, for their unwavering dedication and support. I would also like to acknowledge the invaluable guidance and benevolent demeanor of **Dr. Hanna Rose**, a research scholar from Leibniz Universität Hannover.

Furthermore, I would like to extend my gratitude to the exceptional staff members, including **Sir Shahid** (Lab Technician), **Ms. Fozia Parveen** (Lab Technician), **Arshad Chacha** (Runner), and **Sajida Aunty** (Sweeper), for their amiable and responsible demeanor always.

I must also mention the name of my dear friend and roommate, **Ms. Moattar Fatma**, whose companionship proved to be a true blessing. Many thanks to my senior fellows, **Hafiz Muhammad Khalid**, for his support and encouragement, and **Saba Azeem**, for lending a compassionate ear to my concerns and aiding me in my work. Additionally, I would like to extend my gratitude to my colleagues, **Muhammad Hakeem**, for his valuable companionship during my lab work, and **Maira Shakeel**, for being such a delightful company.

I would like to acknowledge those individuals who, for their own reasons, chose not to assist me as their decision provided me with a valuable opportunity to learn independently.

Completing this degree would not have been possible without the unwavering financial and spiritual support from my beloved parents, **Mr. Khalid Hussain**, and **Mrs. Kubra Begum**. I hold them dear in my heart. I would also like to express my gratitude to my cousin brother, **Waleed Sabir**, for his consistent assistance in various tasks during my stay at the hostel. Thank you to my entire family for their untiring presence and support throughout my journey.

In conclusion, I would like to recognize and commend my own unwavering commitment and diligent work throughout this journey. I am proud of my ability to maintain focus and drive, even in the face of challenges and obstacles. This personal acknowledgement serves as a reminder of my own capabilities and reinforces the importance of self-belief and determination in attaining success.

Salliha Khalid

MS-PBT-10

Table of Contents

List of Abbreviations

A

Abutilon mosaic virus: AbMV, 33

B

Bean yellow mosaic virus: BYMV, 19

C

Cassava mosaic disease: CMD, 32
Cetyl Tri methyl Ammonium Bromide: CTAB, 39
Cherry tomato leaf curl virus: CTLCuV, 34
Chili leaf curl Pakistan virus: ChiLCPKV, 34
Chilli leaf curl India virus: ChiLCINV, 49
Citrus tristeza virus: CTV, 20
Codiaeum leaf curl betasatellite" (CoLCuB).: CoLCuB, 35
Cotton leaf curl Alabad virus: CLCuAlV, 63
Cotton leaf curl Bangalore virus: CLCuBaV, 63
Cotton leaf curl disease: CLCuD, 21
Cotton leaf curl Gezira virus: CLCuGeV, 35
Cotton leaf curl Kokhran virus: CLCuKoV, 31
Cotton leaf curl Multan betasatellite: CLCuMB, 34
Cotton leaf curl Multan virus: CLCuMuV, 31
Cowpea golden mosaic virus (CPGMV): CPGMV, 63
Cucurbit yellow stunting disorder virus (CYSDV): CYSDV, 36
Cucurbits chlorotic yellows virus (CCYV): CCYV, 36

H

Hollyhock vein yellowing virus" (HoVYV): HoVYV, 35

K

Kroton leaf curl betasatellite: KLCuB, 64

L

large intergenic: LIR, 22

M

Melochia mosaic virus (MelMV): MelMV, 36
Movement protein: MP, 22

N

new world: NW, 21
nuclear shuttle protein: NSP, 22

O

Okra yellow crinkle Cameroon alphasatellite: OYCrCMA, 35
old world: OW, 21
open reading frames: ORFs, 20

P

Papaya leaf crumple virus: PaLCrV, 49
Papaya leaf curl virus (PaLCuV): PaLCuV, 35
Pedilanthus leaf curl virus: PLCuV, 34
Plum pox virus: PPV, 20
Polymerase Chain reaction: PCR, 35

R

replication associated protein: Rep, 23
replication enhancer: REn, 22
Retinoblastoma related protein: RBR, 27
Rose leaf curl virus: RoLCuV, 49

S

Satellite conserved region: SCR, 23
Sida chlorotic leaf virus: SCLV, 33
small intergenic regions: SIR, 22
Squash leaf curl virus: SLCuV, 37
Sugarcane mosaic virus: SCMV, 19

List of Figures

List of Tables

Abstract

Begomovirus, the largest known genus of plant viruses, is a real challenge for plant virologists due to its great diversity, wider host range, and ability to undergo frequent recombination. Begomoviruses cause devastating yield losses in economically important crops. The present investigation aims to identify and amplify a *begomovirus* from commonly cultivated ornamental plants, specifically Duranta. repens, Hibiscus, and Jasmine. Leaf samples of one *Duranta repens,* showing leaf curl, two *Hibiscus rosa-sinensis,* exhibiting vein-thickening and a Jasmine, expressing leaf mottling and yellowing symptoms, indicative of begomoviruses and seemingly non-symptomatic samples, were collected in the vicinity of the National University of Sciences and Technology (NUST), Islamabad in 2022. Total nucleic acid was extracted via the CTAB method. Using diagnostic primer pair (CLCV1/CLCV2), expected size bands of approx. 1.1kb were obtained from symptomatic samples while non-symptomatic samples didn't yield any bands. An amplification product of approx. 2.8kb was obtained, using primer pair BurxF/BurxR, from Hibiscus (isolate HAHN), and was sequenced to its entirety. The total nucleotide sequence obtained had 2751 bases and shared maximum nucleotide identity with an isolate of *Cotton leaf curl Multan virus* (CLCuMuV), representing a variant of CLCuMuV. A sequence of 1892 bases was obtained from a full-length product, of Hibiscus (isolate HAH), which shared 97% sequence identity to CLCuMuV. Two full-length products (approx. 2.8kb each) were amplified from D. repens, using primers pairs BurxF/BurxR and SonF/SonR. The partial sequence, 1251 nucleotide bases, obtained from the product amplified with BurxF/BurxR primers shared 94% nt sequence identity to an isolate of *Papaya leaf crumple virus* (PLCV), which has not previously been reported from Pakistan. While partial sequence, 1608 bases, derived using SonF/SonR primers shared 95% nt sequence identity to an isolate of *Chili leaf curl India virus* (ChiLCINV). The partial sequence obtained from Jasmine, 1164 bases, amplified with primer pair SonF/SonR shared the highest nucleotide sequence identity of 98% to an isolate of ChiLCINV. Full-length betasatellite (approx. 1.3kb) were amplified from Hibiscus, and *D. repens* using primer pair Beta01/02 while no amplification was obtained from Jasmine. The complete sequence of betasatellite obtained from Hibiscus was determined to be 1346 nt and shared 96% nt sequence identity to Cotton leaf curl Multan betasatellite (CLCuMB) while the partial sequence, 358 bases, obtained from Duranta shared 91% nucleotide identity with CLCuMB. The study showed that Hibiscus, Duranta and Jasmine are reservoir host plants for economically important viruses, they may serve as sources of inoculum, and recombination sites

for the emergence of new begomoviruses. The study revealed the first report of CLCuMuV infecting *Hibiscus rosa-sinensis* in Pakistan and mixed infection of PLCrV and ChiLCINV in *Duranta repens*.

Keywords: *Cotton leaf curl Multan virus, Hibiscus rosa sinensis, Duranta repens, Chili leaf curl India virus*, Alternate host, Begomovirus diversity.

CHAPTER 1: INTRODUCTION

1.1. Plant Viruses: An Overview

Plant viruses are invisible infectious agents that cause serious turmoil inside a plant. The history of plant viruses' dates to the 19th century when even the word "virus" was unknown. A Russian researcher Iwanowski was inspecting the cause of a peculiar disease in a tobacco plant. During the filtration of plant extract, he noticed a very small agent that even managed to escape the pores of the filter membrane. Those agents were later specified as Tobacco *mosaic virus* (TMV) and it was the first virus to be labeled. Just like viruses of other kingdoms, plant viruses also rely on the host machinery for multiplication and gene expression. They are composed of a genetic material which could be either single or double-stranded DNA or RNA, encapsidated in a protein shell, known as coat protein (Gergerich & Dolja., 2011). Unlike human viruses, plant viruses do not have spikes or receptors that recognize the specific protein inside the host. Plant viruses are transmitted either via an insect vector, mechanical inoculation, or directly via the seed to the next generation. Therefore, the plant viruses directly enter the plant cell cytosol and take control of it. Plant viruses are also host-specific (Figure 1.1).

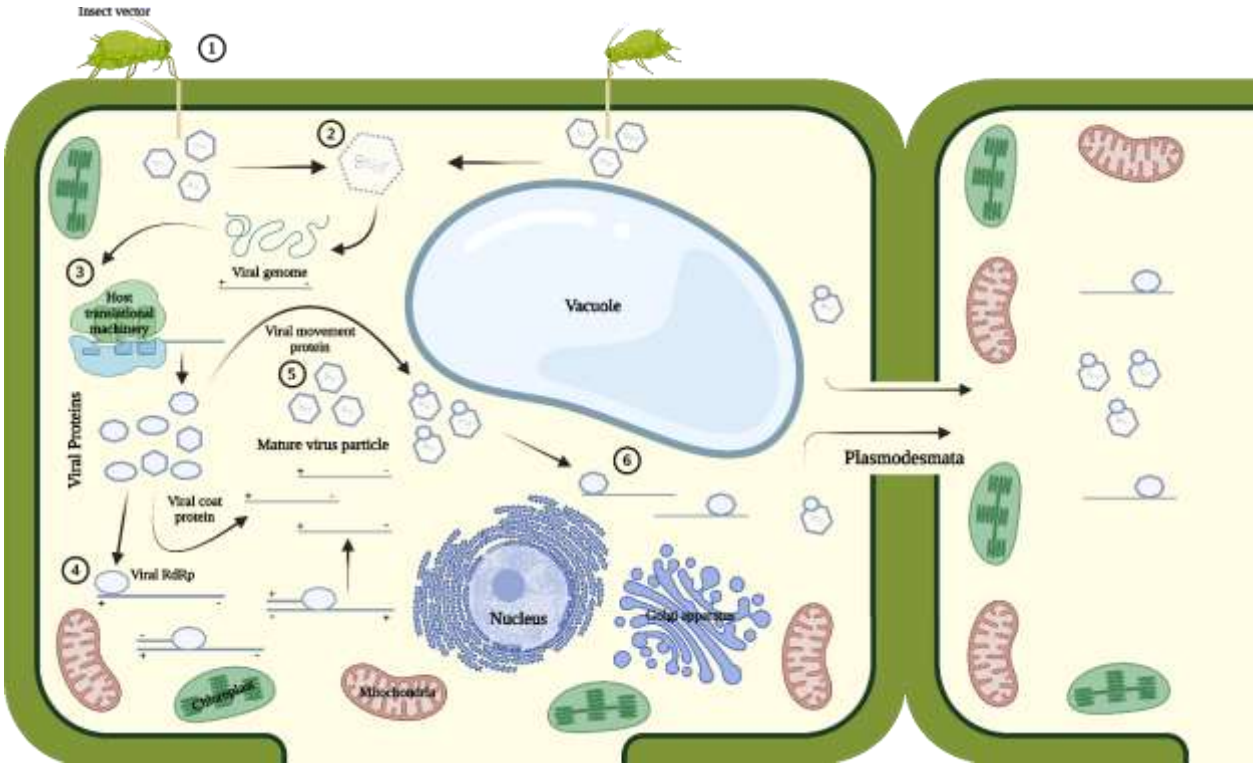

Figure 1.1: General replication cycle of a plant virus: (1) Entry inside the host via insect vector (2) Uncoating of genome (3) Translation of proteins required for genome replication, movement, and encapsidation (4) Genome replication (5) Encapsidaion (6) Movement to adjacent cell via plasmodesmata. (An image created by Bio Render).

Plant viruses interact with specific host factors (primarily defence proteins) during the invasion, and the outcome of these interactions determines whether the invaded plant will act as a host for the invader virus (Fraser., 1990). Initially, plant viruses were known as infectious agents that can only damage crops. However, with the advancements in molecular biology, plant viruses are now used as gene carriers, that can express a desirable protein particularly a vaccine inside a host plant. Thus, with the assistance of plant viruses, plants can act as bioreactors to produce desirable protein products (Pogue *et al.*, 2002).

1.2. Plant Viruses: A Grave Risk to the Agriculture

Plant viruses cause a profound threat to major crops in the form of yield loss. Several reports are showing the massive yield loss in chief crops such as Cotton, Rice, Wheat, Barley, and Maize. Key financial losses are estimated up to 1.5 billion USD in rice crop due to rice tungro disease (Teng *et al.*, 1990), an annual yield loss of 500,000 USD in wheat (McKirdy *et al.*, 2002), 300,000 USD in barley, and 15 to 94.4% yield loss in Oats owed to *Barley yellow dwarf virus* (BYDV), yield loss up to 40% in maize owed to *Maize dwarf mosaic virus* (MDMV), and 73% yield loss in sorghum in East Africa (Rao & Reddy 2020) and 10 - 20% in sugarcane in Pakistan (Qureshi & Afghan., 2005) caused by *Sugarcane mosaic virus* (SCMV). Cotton is another economically important crop which is rigorously affected by a viral disease known as Cotton leaf curl disease. The estimate of economic losses due to cotton leaf curl disease during the year 1992-97 was 5 billion USD in Pakistan (Briddon and Markham, 2001).

Legumes also serve as an important agricultural commodity in the Grass domestic production (GDP) of a country. Several reports showed a noticeable loss in the yield of legume crops due to the illness of plant viruses. A few major reports are 64 - 68% loss in the numeral of pods/plant in common beans triggered by *Bean yellow mosaic virus* (BYMV) (Hampton, 1975), 75%-90% losses during the flowering stage in Chickpea owed to chickpea chlorotic dwarf virus (Kanakala *et al.*, 2013), 11.6 - 43.8% and 24.2 - 66.7% in seed yield in cowpea as a result of *Cowpea banding mosaic virus* and *Cowpea chlorotic spot virus* infection respectively (Sharma and Varma, 1981), 10 - 100% yield loss in mung bean by *Mung bean yellow mosaic virus* (MYMV) infection in Republic of India, 71.5% reduction in black gram yield due to mosaic mottle disease (Nene., 1972), and an yearly loss of 205,000 tons in the yield of red gram grain as a consequence of sterility mosaic disease (SMD) infection (Kannaiyan *et al.*, 1984).

In vegetables the most prominent losses are in potato [40 - 85% average losses due to *Potato virus Y* (PVY), (Nagaich 1975)], tomato [17.6 - 99.7% due to *Tomato leaf curl New Delhi virus*

(ToLCNDV), a deadly virus prevalent in Asia, North Africa, and Southern Europe, (Kalloo., 1997)], Chili/pepper [20 - 80% loss due to *Chili leaf curl virus*, (Saha., 2005)], eggplant [65.2 - 70.3% decrease in yield due to Tobacco ring spot virus (TRSV) infection, (Sastry & Nayudu., 1978)], lady finger [63% reported yield loss due to tobacco stunt virus (Krishnareddy *et al.*, 2003) and 94 - 100% yield loss due to *Bhendi yellow vein mosaic virus* (BYVMV) (Pun., 1999)], cucurbits [60 - 100% yield loss in watermelon and other member of family *Cucurbitaceae* due *Watermelon bud necrosis virus* (WBNV) (Jain *et al.*, 1998 ; Mandal *et al.*, 2003)], carrot [60% loss in carrot roots owed to *Motley dwarf virus* infection (Watson & Serjeant., 1964)], sweet potatoes [80% storage production loss due to mix infection of *Crinivirus*, *Ipomovirus*, and *Potyvirus* (Mukasa *et al.*, 2006)], casava [annual loss of 6 billion USD in Uganda due to Cassava mosaic disease (CMD) (Thresh & Cooter., 2005)].

Among fruits the threatening viral disease *Citrus tristeza virus* (CTV) disease which is responsible for the loss of 500 million USD in citrus trees in Argentina (Nolla, & Valiela., 1976). Other prominent viral diseases in fruit crops are *Banana bunchy top virus* (BBTV) disease responsible for 50 million USD loss in banana production in several states of India (Selvarajan & Balasubramanian., 2013), Grapevine leaf roll disease (GLD) which causes 25,000 – 40,000 USD loss per hectare (Atallah *et al.*, 2012), and a disease of stone fruits, known as "sharka", caused by Plum pox virus (PPV) which is responsible for 83 - 100% yield loss in few susceptible cultivars (Kegler & Hartmann., 1998).

1.3. Family *Geminiviridae*

Geminiviridae is a plant virus family, which contain single-stranded circular genomes of approximately 2.5 – 5.2 kb. The genome is enclosed in an incomplete icosahedral, geminate (twinned) protein coat (Figure 1.2). The protein coat is non-enveloped and made up of single protein subunits. Members within the family *Geminiviridae* are transmitted via the insect carriers in a circulative, non-persistent manner. The most common vectors are leafhoppers, planthoppers, whitefly (*Bemicia tabaci*), and aphids. *Geminiviridae* has a expansive host range comprised of monocotyledonous and various dicotyledonous plant families. *Geminiviridae* is grouped into fourteen genera given that differences in the genome organization (Table 1.1, Figure 1.3) (Zerbini *et al.*, 2017). Each genus of the family is monopartite (consisting of a single genomic component) except *begomovirus*, which could be monopartite or bipartite (consisting of two genomic components). The genome usually consists of 4 – 6 open reading frames (ORFs) in both virion and opposite sense orientation. These ORFs in both orientations are usually parted

by a large intergenic region (LIR) and a small intergenic region (SIR). All members of the family shared a highly conserved stem-loop organization known as "nonanucletide" (TAATATT/AC). This highly conserved stem-loop region is present within the LIR and marks the origin of replication (Matthews & Hull., 2002). Few members of the family are associated with viral satellites namely, alpha or beta satellites. These satellites rely upon the host-virus for their proliferation or travel around the host plant, and in turn help the host-virus in symptom induction (Nawaz-ul-Rehman & Fauquet., 2009). Satellites associated with the genus *begomovirus* are discussed in detail later in this chapter.

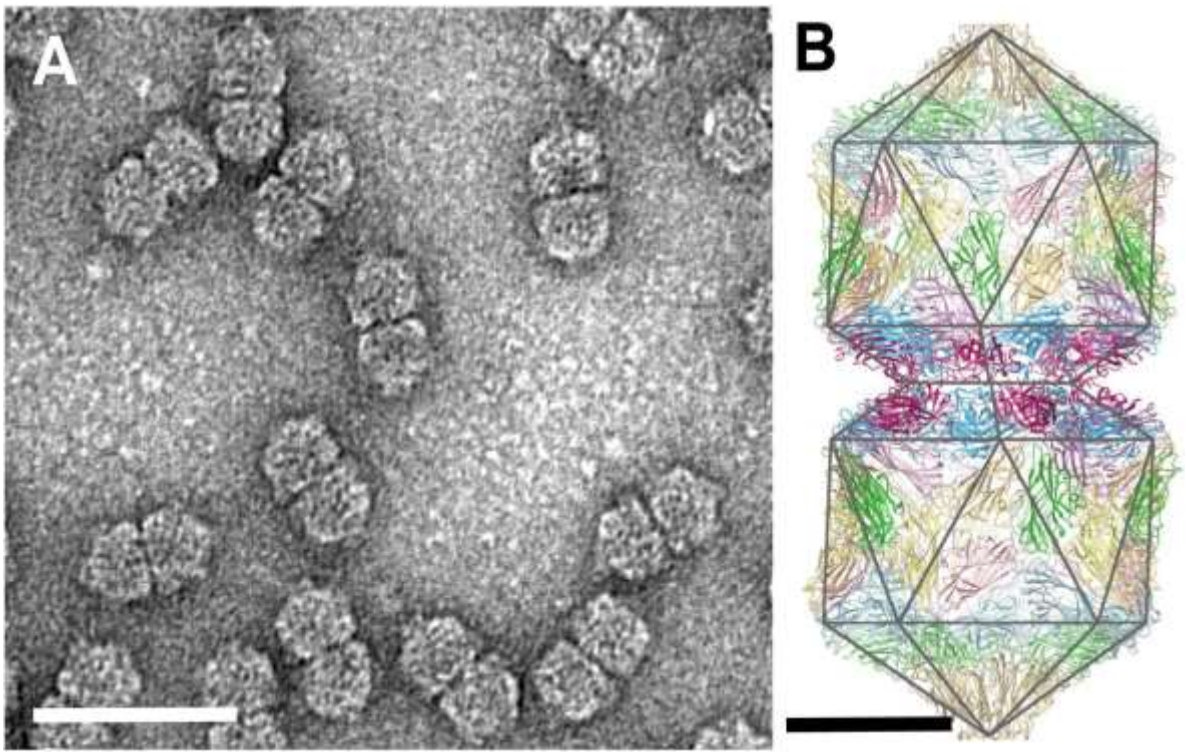

Figure 1.2: (A) Electron micrograph of a member from the genus *Mastrevirus*. (B) Structure of member from genus *Begomovirus*. Both structures depict the typical twin icosahedral protein capsids that enclose the Geminiviruses genome (Source: ICTV).

Table 1.1: Genera of the family *Geminiviridae*

Genus	Transmission	Host Range	Species	Accesion No.
Becurtovirus	Leafhopper	Dicots	Beet curly top Iran virus	[EU273818]
Begomovirus	Whitefly	Dicots	Bean golden mosaic virus	[JN419006]
Capulavirus.	Aphids	Dicots	Alfalfa leaf curl virus	[KP732474]
Citlodavirus	-	Dicots	Camellia chlorotic dwarf-associated virus	[MG452759]
Curtovirus	Leafhoppers	Dicots	Beet curly top virus	[M24597]
Eragrovirus	-	Monocots	Eragrostis curvula streak virus	[FJ665631]
Grablovirus	Alfalfa treehopper	Dicots	Prunus latent virus	[MF510408]
Maldovirus	-	Dicots and Monocots	Apple geminivirus 1	[KM386645]
Mastrevirus	Leafhoppers	Monocots	Chickpea chlorosis virus	[GU256531]
Mulcrilevirus	Leafhoppers	Dicots	Mulberry crinkle leaf virus	[KR131749]
Opunvirus	Cochineal insects	Cactaceae	Opuntia virus 1	[MN100000]
Topilevirus	-	Dicots	Tomato apical leaf curl virus	[MG491195]
Topocuvirus	Treehoppers	Dicots	Tomato pseudo-curly top virus	[X84735]
Turncurtovirus	Leafhoppers	Dicots	Turnip curly top virus	[KC108896]

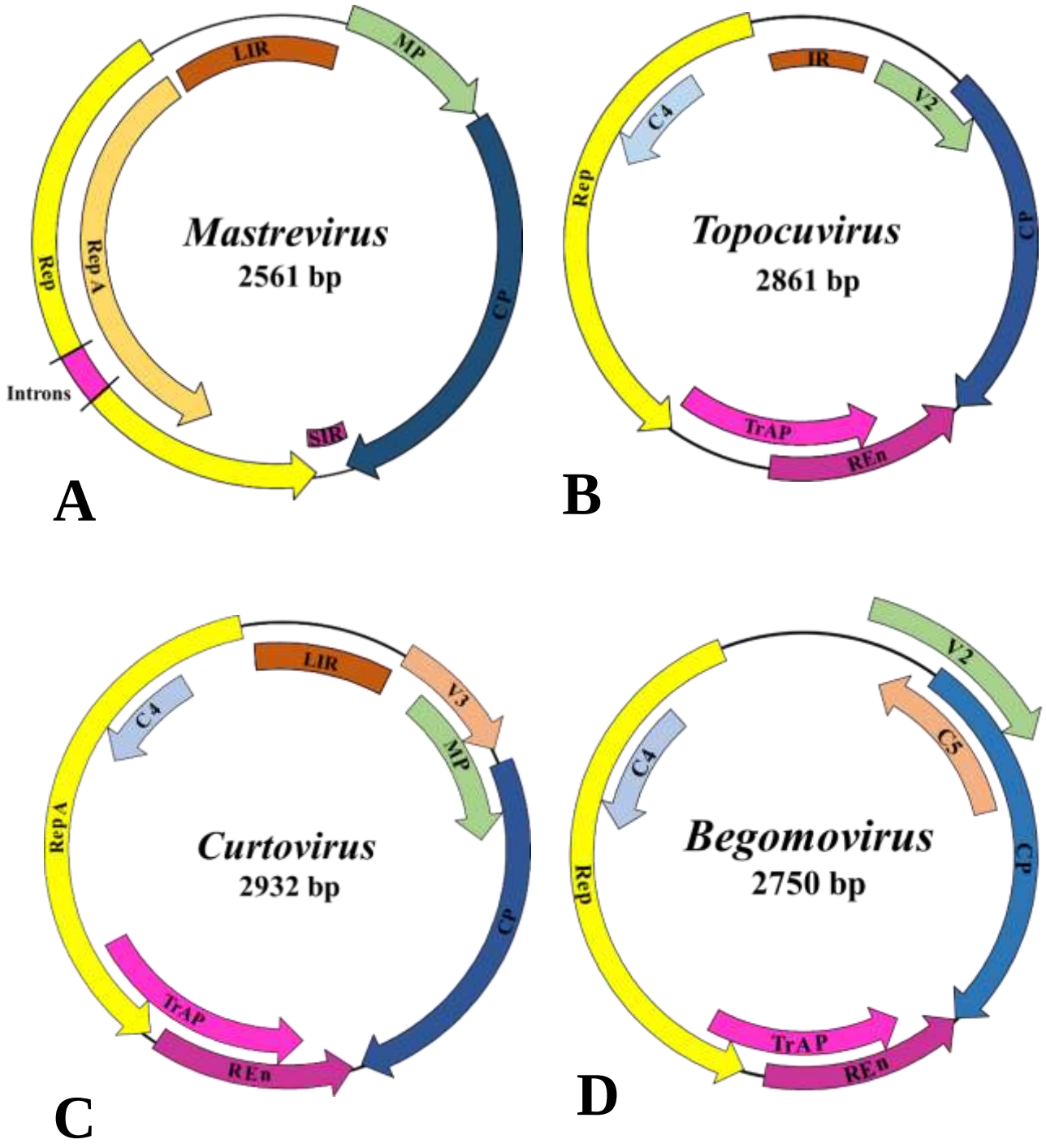

Figure 1.3: Genome maps of a few of the genera from the family *Geminiviridae*. **(A)** genus *Mastrevirus* **(B)** Genus *Topocuvirus* **(C)** Genus *Curtovirus* **(D)** Genus *begomovirus* monopartite. Each colour-coded arrow shows the position and direction of genes. SIR (Short Intergenic Region), LIR (Large Intergenic Region). (Genome maps recreated in pDRAW32).

1.4. Genus *Begomovirus*

Begomovirus is among the substantial genera of plant viruses, clustering nearly 450 species. This genus includes both monopartite and bipartite species which are frequently associated with ssDNA satellites namely, alphasatellite, betasatellite, and deltasatellite. These species are sometimes also distinguished as old world (OW) and new world (NW) begomoviruses. The genus *begomovirus* grouped several economically important viruses which are responsible for causing devastation in important crops. For example, Cotton leaf curl disease (CLCuD) responsible for million dollars loss in the cotton trade sector of Pakistan and India, is initiated by a group of 5 -6 begomoviruses (Briddon., 2003). Similarly, the *tomato yellow leaf curl virus* responsible for the massive loss in tomato fields (Moriones & Navas-Castillo., 2000), and the *Cassava mosaic virus*, a cause for the destruction of cassava fields in Africa (Thresh & Cooter., 2005), are also grouped in the genus *begomovirus*. Other economically important viruses from the genus *begomovirus* are, *Mungbean yellow mosaic virus*, *Squash leaf curl China virus*, and *Tomato leaf curl New Delhi virus* (Fiallo-Olivé *et al.*, 2021). In the light of these facts, the genus *begomovirus* is of great importance for plant pathologists and virologists.

1.5. Genome Organization

Begomoviruses consist of a ssDNA of approximately 2.8kb which is enclosed in a geminate protein coat. Typical monopartite *begomovirus* genome codes for 5-6 proteins with a known function. The bipartite *begomovirus* contains two related genomic components known as DNA-A (analogous to monopartite *begomovirus*) and DNA-B. DNA-A consists of two virion sense genes known as AV1 and AV2. AV1 is also known as coat protein (CP). The complementary sense genes are AC1 also known as Rep, AC2 also known as transcriptional activator protein (TrAp), AC3 also recognized as replication enhancer (REn), AC4, and AC5. DNA-B consists of two genes, BV1 identified as nuclear shuttle protein (NSP) and BC1 distinguished as Movement protein (MP). Each component consists of a conserved nonanucleotide (5′-TAATATT/AC-3′) sequence in the stem-loop region. The genomic components also consist of large intergenic (LIR) and small intergenic regions (SIR). The conserved section contains the virus-specific sequences known as “iterons” which are recognized by its Rep protein. Moreover, the cognate components of bipartite viruses also contain a 200 bp-long region within the LIR, which is known as the common region (CR) (Fiallo-Olivé et al., 2021). The typical genome organization of begomoviruses is illustrated in (Figures 1.3 & 1.4).

1.6. Satellite DNA Molecules

Satellites are distinct components of non-virus origin that depend on the host-virus for their proliferation, encapsidation, and movement around the host cells. There are several reports in the literature which indicate satellites are the essential component of single-stranded begomoviruses. Two satellites, namely the alphasatellite and betasatellite (Figure 1.5), are the most well-characterized satellites of begomoviruses. These satellites co-exist along with their helper viruses in the form of complexes and are thus essential for disease induction.

a) Alphasatellite initially known as DNA-1, is a self-replicating satellite, that was first discovered along with a cotton infecting begomovirus (Mansoor *et al.,* 1999). Like its helper virus, alphasatellie also has an ssDNA of approximately 1.3kb (Figure). ssDNA codes its specific replication-associated protein (Rep) and consequently, only dependent upon the helper virus for encapsidation and cell-cell movement (Briddon & Stanley 2006). Alphasatellites particularly have no direct involvement in the symptom induction of their partner virus. Nevertheless, the binding of alphasatellite encoded Rep protein to the helper virus Rep protein has been evident in previous studies. This binding event sometimes decreases the count of betasatellite, which rest on the partner virus Rep protein for their proliferation inside the host cell (Nawaz-ul-Rehman *et al.,* 2010).

b) Betasatellite first discovered along with *Ageratum vein yellowing virus* (AVYV) has been widely characterized begomovirus associated satellite. Betasatellite is almost 1.3kb, which is partial to the size of its partner virus and encodes a solo protein known as βC1 spanning almost 350 nucleotides in a complementary sense. The genome characterization of betasatellite has revealed two other features, the origin of replication containing the conserved nonanucleotide (TAATATTAC) sequence also known as Satellite conserved region (SCR), and a region with repeated adenine nucleotides recognized as A-rich region. (Briddon *et al.,* 2003). Betasatellite is now known to be the essential component for the disease induction of several begomoviruses. Betasatellite is also a part of cotton infecting begomovirus complex. The only betasatellite-encoded protein, βC1, plays an essential role in the reversion of the host cell cycle, by inhibiting the methylation content of host genes (Figure) (Bowdoin *et al.,* 2013).

c) Deltasatellite is a ss DNA molecule of approximately 600 bp, that do not code for any protein, often associated with begomoviruses (Lozano *et al.,* 2016).

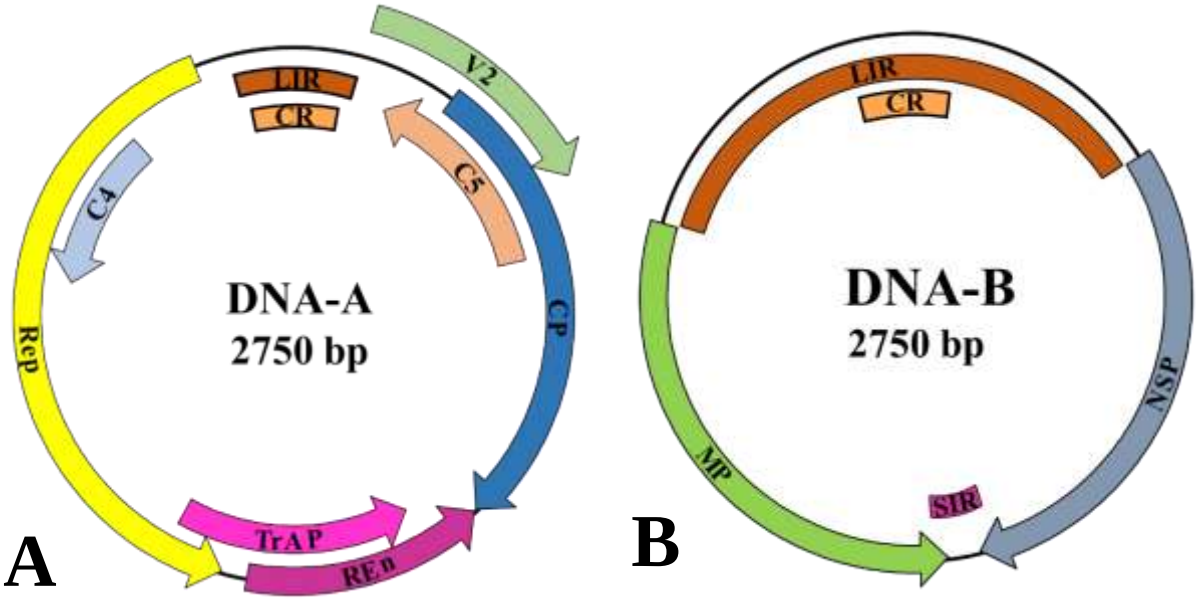

Figure 1.4: Genome map of bipartite *begomovirus*. (A) DNA-A (B) DNA-B. Each colour-coded arrow shows the position and direction of genes SIR (Short Intergenic Region), LIR (Large Intergenic Region), and CR (Common Region). (Genome maps recreated in pDRAW32).

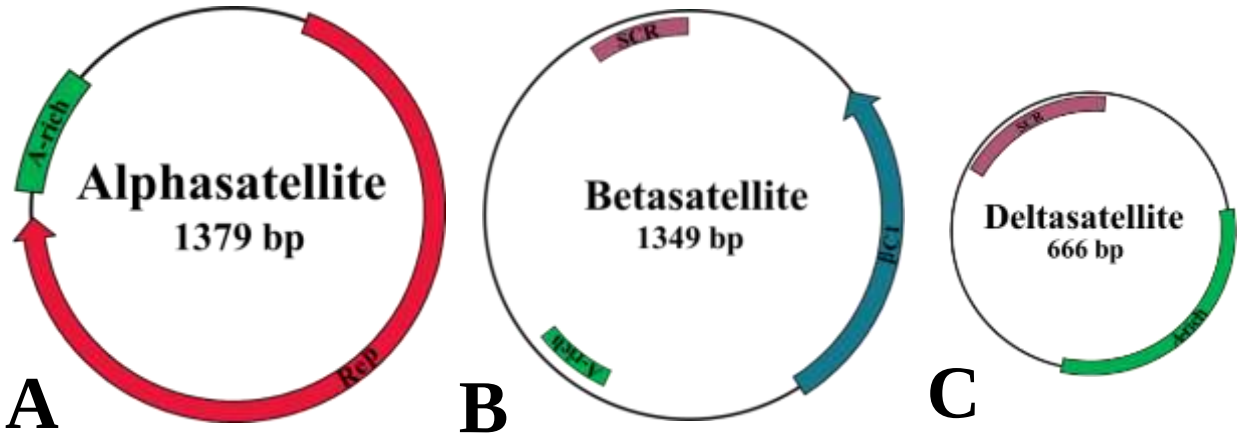

Figure 1.5: Genome maps of *begomovirus*-associated satellites **(A)** Alphasatellite **(B)** Betasatellite **(C)** Deltasatellite. Each colour-coded arrow shows the position and direction of genes. Deltasatellite does not code any protein. SCR (Satellite Conserved Regions). (Genome maps re-created in pDRAW32).

1.7. Transmission & Host Range of Begomovirus(es)

Just like other plant viruses, begomoviruses are also transmitted to the plant via an insect transmitter. *Bemicia tabaci* generally recognised as whitefly is the sole carrier of begomoviruses. Whitefly transmits begomoviruses in a non-persistent circulative manner. (Figure 1.7B) shows the movement of begomovirus throughout the insect. *B. tabaci* is a complex of 40 physically identical species that are "polyphagous" (De Barro *et al.*, 2011). Therefore, they can feed upon many plant host species. *B. tabaci* is abundantly available in an environment under suitable conditions. Moreover, few past reports have shown that different *B. tacabi,* species have different transmission capacities for different species of begomoviruses (Barman *et al.*, 2022). These facts contribute to the success of the whitefly being the supervector of begomoviruses, and thus it could be the possible factor behind the extraordinary spread of begomoviruses across the globe. Begomoviruses are currently reported from many plant species. The host range of begomoviruses might vary from species to species but, overall, it expands from the family; *Fabaceae, Malvaceae, Cucurbitaceae, Solanaceae, Euphorbiaceae,* and *Caricaceae.*

Figure 1.6: (A) Nymphs of whitefly on squash leaf, **(B)** Adult whiteflies on underside of squash leaf **(C)** Close view of an adult whitefly (Source: CABI digital library).

1.8. Replication of Begomoviruses Inside the Host Plant

Just like other viruses, begomoviruses utilize the host cell machinery for gene expression. The begomoviruses infect the resting cell, in which the cell cycle and DNA replication have already ceased. Hence, the viral genes accompanied several strategies to revive the process of the host cell cycle. After the entry inside the cell, begomoviruses shed their protein coat and the ssDNA travel towards the nucleus of the cell, where the ssDNA is first converted into dsDNA. Once the dsDNA intermediate has formed, the viral mRNA starts accumulating inside the nucleus of the cell. Viral mRNA thus travels towards the cytoplasm where it gets translated utilizing the host translational machinery. The synthesized viral proteins thus aid in the replication and

packaging of the virus genome (Figure 1.7) (Hanley-Bowdoin *et al.,* 2013). The function of begomovirus proteins is tabulated in (Table 1.2).

1.9. Begomovirus Proteins and Host Factor Interactions

DNA-A encoded Rep initiates the process of replication by creating a nick in the nonanucletide region (Figure 1.7). While REn acts as the enhancer of the replication process and TrAp increases the transcription of viral proteins. Apart from their direct role in virus replication, these three proteins perform a role in the restoration of the host cell cycle machinery. Normally the host cell is in the resting phase upon the invasion of the virus. *Begomovirus* proteins undergo several interactions with different host factors and thus restore the machinery necessary for DNA replication. "E2F" is a transcriptional factor which is known to control the expression of genetic factors associated with the host DNA duplication and cell cycle control. The expression of E2F is regulated by another protein known as Retinoblastoma related protein (RBR). RBR, a plant homolog of the animal tumour suppression gene "pRB", regulates plant cell multiplication and specification (Gutzat *et al.,* 2012). Begomovirus Rep and Ren disrupt the RBR, which causes the release of E2F. E2F thus increases the expression of genes which are essential for DNA replication during the S and G phases of the cell cycle (Figure 1.8) (Arguello-Astorga *et al.,* 2004). Moreover, different studies prove the interaction of begomovirus Rep and Ren proteins with host proliferating cell nuclear antigen (PCNA) (Castillo *et al.,* 2013), replication factor C (Luque et al., 2002), replication protein A (Singh et al., 2007), RAD5 (Kaliappan et al., 2015), and MCM2 (Suyal et al., 2013).

1.10. Host, Virus, and Vector Interaction

Plant viruses are distinct from animal or human viruses in a way that they require a vector (mostly an insect) to migrate from one host to another. Hence, a plant virus encounters a vector as a temporary host before entering the actual host (Plant). Vectors for plant viruses are mostly insect species which are polyphagous. Thus, they can feed upon different types of food material, in other words, different plant species. Such distinct host species play a crucial task in establishing the virus population (Simmonds *et al.,* 2019). Each interaction of a host and virus is individual, which leads to the adaptation of distinct viruses, strain or variant, to a new host species and thus widen the host range. Furthermore, expansion in agricultural and horticultural growth with a greater number of vector species available in the environment favored the development and rapid expansion of new virus species, and strains. Such facts demand in-depth study to fully grasp the tripartite interaction between virus, vector, and host plant.

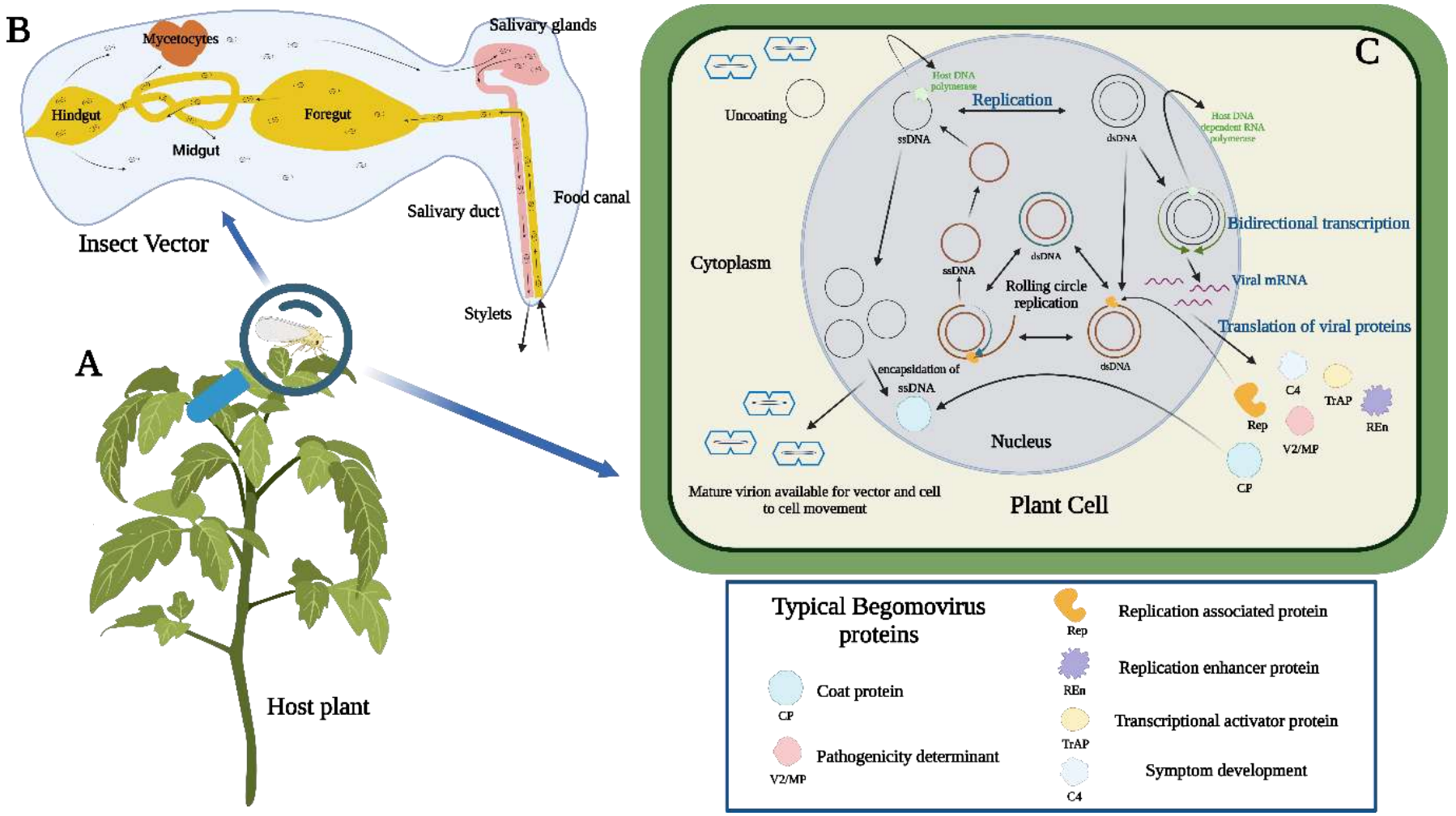

Figure 1.7: Replication of *begomovirus* inside a host plant. A Host plant B Movement of virus inside a whitefly C Replication model of ssDNA inside the host cell nucleus. (An image created by Bio Render).

Table 1.2: Function of proteins encoded by both components of genus *begomovirus* and βC1.

Gene/Protein	Function	Reference
CP	Coat Protein/Virus particle formation Cell to cell movement/Systemic spread Viral DNA accumulation Insect transmission	Briddon *et al.,* 1990 Wartig *et al.,* 1997 Boultan *et al.,* 1993
V2	Pathogenicity determinants Systemic movement Suppressor of RNA silencing	Wartig *et al.*, 1997 Chowda-Reddy *et al.,* 2008
Rep	Replication Modulation of viral gene expression Activation of host cell cycle	Etessami *et al.,* 1991 Egelkrout *et al.,* 2001
TrAp	Gene activation Viral pathogenicity Suppression of gene silencing	Haley *et al.,* 1992 Hong *et al.,* 1996 Chowda-Reddy *et al.,* 2009
REn	Enhances viral DNA accumulation Symptom development	Sunter *at al.,* 1990 Sung & Coutts. 1995
C4	Silencing suppressor Inhibit other plant defense responses Symptom determinant	Ismayil *et al.,* 2018 Mei *et al.,* 2021 Zhan *et al.,* 2018
C5	Virulence factor Silencing suppressor	Li *et al.,* 2015 Li *et al.,* 2021
NSP	ssDNA shuttling between nucleus and cytoplasm Avirulance determinant	Noueiry *et al.,* 1994 Garrido-Ramirez *et al.,* 2000
MP	ssDNA long distance movement	Noueiry *et al.,* 1994

βC1	Symptom induction Interference with host methyl cycle	Briddon & Stanley, 2006 Bowdoin *et al.*, 2013

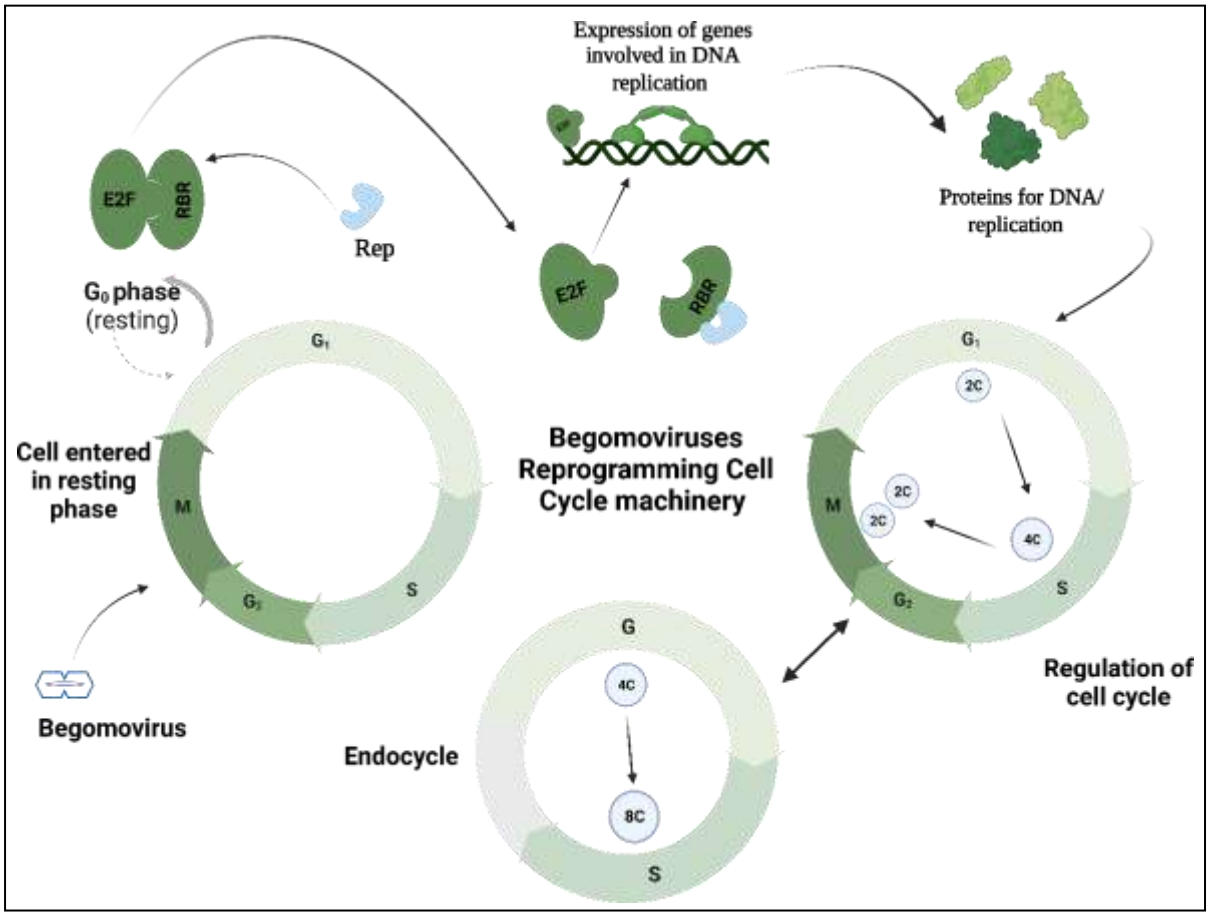

Figure 1.8: Reprogramming of host cell cycle via *begomovirus* encoded Rep protein. Rep interacts with "RBR" and causes it to release host transcriptional factor "E2F". E2F triggers the transcription of genes within the cell cycle regulation (Proteins required for the G and S phases of cell cycle). Once the cell cycle is regulated, the cell may decide to enter the Endocycle. In endocycle cells do not undergo "mitosis". However, DNA continues replicating. (An image adapted from Bowdoin *et al.*, 2013 and recreated in Bio Render).

1.11. Recombination, & Reassortment

Viruses do mutate frequently to survive in the ever-evolving immune system of host species. Thus, mutations such as indels or point mutations, in the genetic fragments could be the possible reason for the rapid evolution of plant viruses. Apart from mutation, other strategies adopted by viruses are recombination and reassortment of genetic fragments. Viruses adapt such strategies to fit with the new environment or new host. (Fiallo-Olivé & Navas-Castillo, 2023). Many reports from the past reinforce the fact that recombination is a source for the emergence of new groups in the genus *begomovirus*. The most prominent is the resistance breakdown against

CLCuD in the Burewala district of Pakistan due to the emergence of a new strain of Cotton leaf curl Multan virus (CLCuMV) (Mansoor *et al.,* 2003). Resistance breaking strain is presently known as the Burewala strain of *Cotton leaf curl Kokhran virus* (CLCuKoV). Cassava mosaic disease (CMD) is a devastating disease of cassava crops in Africa. The diversity of CMD-causing begomoviruses is mainly the consequence of recombination among these species (Crespo-Bellido *et al.,* 2021). Another example is the emergence of *Tomato yellow leaf curl Malaga virus* in Portugal due to the recombination event between the two different strains of ToYLCVs (Fiallo-Olivé *et al.,* 2019). Thus, recombination is the major factor that contributes to the increasing variety of begomoviruses across the globe.

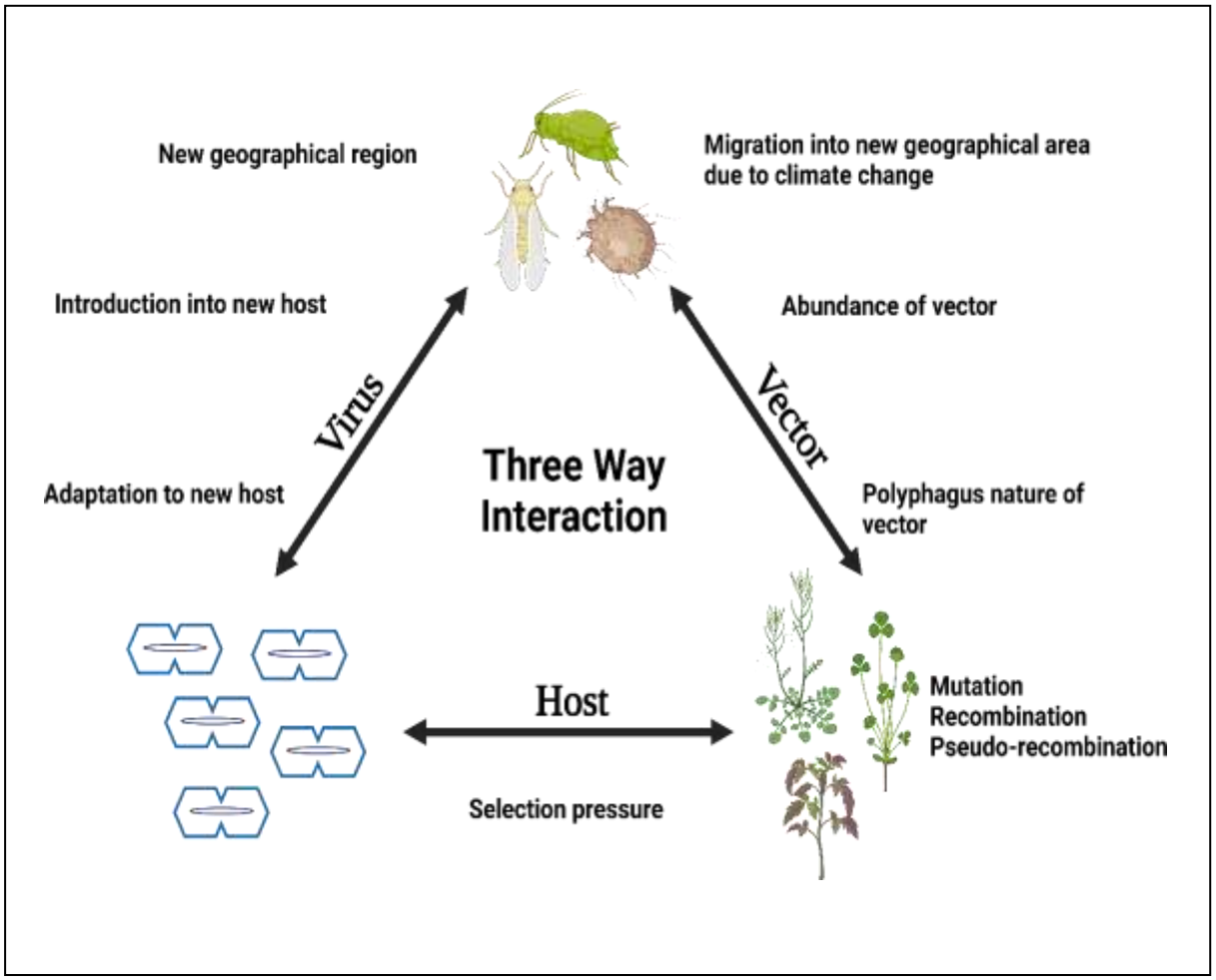

Figure 1.9: Schematic representation of tripartite interaction between Virus, Vector and Host plant. (An image created in Bio Render).

Objectives

Considering the challenges imposed by begomoviruses on our agricultural sector, this study is designed to detect and explore the diversity of these viruses in different plant species.

The objectives of the present study are:

1. Diagnosis of begomovirus(es) using universal primers.
2. Amplification of potential begomovirus(es)
3. DNA sequencing and sequence analysis of potential amplicons
4. Recombination and phylogenetic analysis

CHAPTER 2: LITERATURE REVIEW

Dhobale., ***et al*** **(2023)** reported begomoviruses encompassing a bipartite genome being the key threat to the cultivation of mungbean (*Vigna radiata L.*) in different regions of India. They detected two different strains named MYMV (*Mungbean yellow mosaic virus*) and MYMIV (*Mungbean yellow mosaic India virus*). The former strain was detected in Bihar while the latter was confirmed in Assam and Orissa. Further, the authors claimed to clone and sequence the full-length genomic sequences of both strains by RCA (Rolling circle amplification). Phylogenetic analysis of resultant sequences showed that MYMV isolate of Bihar was most strictly related to a MYMV-Urdbean isolate of New Delhi while the MYMIV isolate of Assam and Odisha had the nucleotide sequence identity with MYMIV mungbean isolate of Meghalaya. DNA-A and DNA-B of detected viruses showed a distinct ancestral lineage, which was owing to the higher recombination events and genetic variability among begomoviruses. They further claimed to construct infectious clones for the infectivity of analysis of respective viruses. Infectivity of clones was assessed via agroinfiltration in eight mungbean genotypes, cowpea (*Vigna unguiculata L.*) and tobacco (*Nicotiana benthamiana*). Agroinfiltrated plants showed varying degrees of symptoms that were signs of Yellow Mosaic Disease (YMD).

García-Rodríguez ***et al.,*** **(2023)** investigated the weed from genera *Sida* and *Malvastrum* for the prospective presence of begomoviruses. Authors used the Illumina platform followed by contig annotation that was accomplished using the BLASTn tool of GenBank, for the identification of begomoviruses. The occurrence of begomoviruses was further confirmed via PCR and full-length viruses were sequenced employing sanger sequencing. The identified virus was asserted to be a new species that was named *Sida chlorotic leaf virus* (SCLV). Phylogenetic and recombination studies revealed that SCLV is a product of recombination between a member of the *Abutilon mosaic virus* (AbMV) clade and another from the SLCV (Squash leaf curl virus) clade. This study indicates that weeds serve as a probable reservoir and recombination site of begomoviruses.

Rahman ***et al.,*** **(2023)** published an investigation conducted on the soybean plants showing the typical symptoms of begomoviruses in Faisalabad, Pakistan. They collected infected samples, amplified targeted viruses, and sequenced the resultant products via sanger sequencing. The authors claimed to identify multiple monopartite and bipartite viruses along with associated DNA satellites. Among the identified species were *Mungbean yellow mosaic India virus*

(MYMIV), *Mungbean yellow mosaic virus* (MYMV), *Pedilanthus leaf curl virus* (PLCuV), *Cherry tomato leaf curl virus* (CTLCuV), *Tomato leaf curl Kerala virus* (TLCuKV), and *Cotton leaf curl Kokhran virus* (CLCuKoV). Among the identified species MYMV was the most widespread species followed by PLCuV. In the present study, virus-associated DNA satellites were also identified. Among them the most prominent were, AYVI *Ageratum yellow vein India alphasatellite Papaya leaf curl betasatellite*, *Tomato leaf curl betasatellite*, and *Cotton leaf curl Multan betasatellite* (CLCuMB). Such a colossal number of begomoviruses associated with symptom-enhancing satellites is a major danger to soybean cultivation in Pakistan.

Gu *et al.*, (2023) observed severe stunting and chlorotic symptoms along with deformation of fruits in muskmelon, pumpkin, luffa, and squash at three different places in China. According to the authors, these symptoms were associated with some disease transmitted probably via whitefly and it was responsible for a huge loss in the economy. To detect the causal agent a total of 40 symptomatic samples were gathered from all three provinces. Total DNA was obtained, and PCR was executed via the primers specific for *Cucurbits chlorotic yellows virus* (CCYV), *Cucurbit yellow stunting disorder virus* (CYSDV), and Begomovirus. Results showed that 29 out of 40 samples were positive for begomoviruses. The resulting sequences showed the maximum nucleotide sequence resemblance with *Tomato leaf curl New Delhi virus* (ToLCNDV). All positive samples were further tested using the primers specific for ToLCNDV, which showed the same results. As samples were collected from three different regions, however, all amplified sequences showed maximum (97 – 99%) nucleotide identity with ToLCNDV. Previously ToLCNDV was reported from oriental melons and tomatoes from China. The authors claimed that ToLCNDV infecting muskmelon, pumpkin, luffa, and squash has been reported first time in China.

Kwak *et al.*, (2022) investigated tomato and pepper plants in Myanmar for the potential occurrence of begomoviruses. They collected a total of fifty-nine samples and diagnosed them using the polymerase chain reaction PCR technique. Authors claimed that 59.3% of samples contained the begomovirus infection. A complete genome sequence of amplicons revealed a mixed infection of five different begomovirus species in tomato and pepper fields of Myanmar. Among the identified species were *Tomato yellow leaf curl Kanchanaburi virus* (TYLCKaV), *Tomato yellow leaf curl Thailand virus* (TYLCTHV), , *Tobacco leaf curl Yunnan virus* (TbLCYnV), *Chili leaf curl Pakistan virus* (ChiLCV/PK), and *Tobacco curly shoot Myanmar*

virus (TbCSV-[Myanmar]). Based on 91% specie demarcation criteria for begomoviruses, the TbCSV-[Myanmar] was affirmed as a new species. The study concluded the most widespread species in Myanmar were ChiLCV/PK and TbCSV-[Myanmar].

Akram *et al.*, (2022) noticed typical symptoms (leaf curling and vein swelling symptoms) of begomoviruses on an ornamental plant *Codiaeum variegatum* commonly known as garden cotton in Faisalabad, Pakistan. Such obvious symptoms directed the authors to perform a molecular-level examination of the plant. PCR amplification was done using specific primers designed for begomoviruses and the amplified products were cloned in a T/A cloning vector pTZ57R/T. Upon sequencing the potential clones using specific primers, two different monopartite begomoviruses were revealed. Begomoviruses associated beta satellite was also amplified from both symptomatic plants. Phylogenetic analysis revealed that the begomoviruses infecting garden cotton were *Pedilanthus leaf curl virus* (PeLCV) and *Papaya leaf curl virus* (PaLCuV). However, the associated satellite was found to be newly discovered and named as "Codiaeum leaf curl betasatellite" (CoLCuB). Authors claimed that this is the first report of begomoviruses infecting *Codiaeum variegatum* in Pakistan.

AlHudaib *et al.*, (2022) investigated tomato and muskmelon plants showing the characteristic symptoms of begomoviruses in Saudi Arabia. Infection was primarily confirmed via core coat protein sequences and then processed using Illumina MiSeq sequencing. Authors claimed to discover twelve full-length (2.7–2.8 kb) begomovirus sequences along with eight sequences equivalent to the length (~1.3) of begomovirus-associated beta satellites. Among the identified sequences were *Tomato yellow leaf curl virus* (TYLCV), *Cotton leaf curl Gezira virus* (CLCuGeV), Okra leaf curl Oman betasatellite (OLCuOMB), and Okra yellow crinkle Cameroon alphasatellite (OYCrCMA) from tomato. Sequences identified from muskmelon were bipartite showing largest sequence identity with *Tomato leaf curl Palampur virus* (ToLCPalV). Symptomatic tomato plants also showed mixed infection. Recombination analysis was performed for all identified sequences. According to the authors, TYLCV and OLCuOMB were the recombinant isolates. Authors further claimed that all identified sequences grouped with their respective Iranian isolates in a phylogenetic tree. This is the first report showing the incidence of begomoviruses in tomato and muskmelon in Saudi Arabia.

Khan *et al.*, (2021) observed mosaic accompanied by yellow vein and leaf crumple symptoms in two different Hollyhock plants. The severity of the symptoms led to the testing of plants via

Polymerase Chain reaction (PCR) technique. PCR amplification using universal primers confirmed the incidence of begomoviruses in the nucleic acid extract of both plants. However, begomoviruses-associated betasatellite was only detected in the nucleic acid extract of the plant showing yellow vein mosaic symptoms. By determining the sequence of amplified viruses, it was revealed that both viral products shared 99.9% similarity. However, the similarity index with all other previously characterized begomoviruses was less than 91%. These results indicated that the virus amplified from hollyhock is a new species and it was named as *Hollyhock vein yellowing virus*" (HoVYV). HoYYV-associated betasatellite showed identity with cotton leaf curl Multan betasatellite (CLCuMuB). This study is a clear indication that ornamental plants are potential reservoirs of begomoviruses and thus pose a serious threat to the begomoviruses' vulnerable commercial crops.

Sanchez-Chavez *et al.*, (2020) investigated the cucumber plant showing the typical symptoms of begomoviruses in the North American region. The authors mentioned that previously there is only one report from the North American region indicating the begomovirus infection in cucumbers. The authors collected the symptomatic leaves and subjected them to Illumina sequencing. Results indicated the presence of begomoviruses which led to the PCR diagnosis of the infected plants. PCR using the specific primers give the products equivalent to the typical size length of begomoviruses. PCR products were cloned and sequenced. Authors claimed to identify five different begomovirus species none of which were previously reported from cucumber. A resemblance of these sequences with the already available sequences in the database showed that one of the species is a newly identified species as it showed a utmost of 82% nucleotide sequence identity with already available sequences. This newly identified species grouped with the *Squash leaf curl virus* (SLCuV) group in the phylogenetic tree. These results indicate the mixed infection in a cucumber plant. Mixed infection of begomoviruses is a clear indication of recombination. Therefore, the authors mentioned that cucumber might serve as a centre of begomoviral recombination.

Fiallo-Olivé *et al.*, (2015) reported the incidence of begomoviruses in a wild plant species native to Brazil. The authors investigated the plant named Melochia which belongs to the Malvaceae family. The first step was the collection of symptomatic leaves from two different plants followed by total DNA extraction. PCR amplification using the begomovirus specific primers gave full-length amplified products from both samples. PCR amplification was positive for both DNA-A and DNA-B components which confirmed that the potential virus was bipartite.

Resultant PCR fragments were cloned in pBluescript II and sequenced to explore the genomic makeup of the viruses. Upon sequencing, it was revealed that the DNA-A parts of amplified begomoviruses showed a maximum of (82 %) sequence *Centrosema yellow spot virus* and (81 %) *Tomato yellow spot virus*. While one of the DNA-B showed a maximum of (80 %) nucleotide sequence similarity to *Solanum mosaic Bolivia virus* and the other showed (79 %) identity to two different isolates of *Sida micrantha mosaic virus*. The results presented that the amplified viruses are new species of begomoviruses and therefore, named *Melochia mosaic virus* (MelMV) along with and isolate descriptor; [Brazil-Corumbá B25-2014] and *Melochia yellow mosaic virus* (MelYMV) with an isolate descriptor; [Brazil-Corumbá B26-2014], respectively. The authors claimed that this is the very first report showing the infection of begomoviruses in the genus *Melochia*.

Bandaranayake *et al.*, (2014) investigated several plants from the *Cucurbitaceae* showing the typical symptoms of begomoviruses in Sri Lanka. According to authors previously there was no report of begomoviruses infecting plants in Siri Lanka. However, *begomovirus*-specific symptoms were observed in Bitter gourd along with the presence of whitefly, which is the most popular vector of the suspected virus. The authors performed the PCR-based assay using CP specific primers for the confirmation of disease. Almost 550 bp product was amplified from all symptomatic plants which include Ridge gourd (*Luffa acutangula*), Pumpkin (*Cucurbita maxima*), Cucumber (*Cucumis sativus*) and Snake gourd (*Trichosanthes cucumerina*). The amplified CP showed 97% nucleotide sequence identity with ToLCNDV isolate from Pakistan. The authors claimed that this is the first report showing begomovirus occurrence in Siri Lanka.

Marwal *et al.*, (2013) investigated two distinct plants exhibiting the typical symptoms of begomovirus in Rajasthan, India. Plants were *Jasminum sambac* from the family *Oleaceae* and *Millingtonia hortensis* from the family *Bignoniaceae*. Nucleic acid extracts from both plants were tested via the PCR-based amplification using the begomovirus-specific primer pair set. The resultant coat protein sequence of about 550 bp was cloned and sequenced. The sequences from *J. sambac* showed the maximum nucleotide similarity of 96% with the *Sonchus yellow mosaic virus* isolate MP 2 coat protein gene, and the *Ageratum enation virus* - Lucknow CP gene. While the sequence from *M. hortensis* gave a maximum of 94% sequence identity with the *Rose leaf curl virus* and 93% similarity with *Catharanthus yellow mosaic virus* complete genome. According to the authors, it was the first report presenting the begomovirus infection in *J. sambac* and *M. hortensis*.

CHAPTER 3: MATERIALS & METHODS

3.1. Sample Collection:

Sample showing the usual symptoms of begomoviruses were sorted from two different areas of Islamabad at the end of summer 2022. A total of five samples were collected including Cestrum, Hibiscus, Duranta, and Jasmine showing a range of typical begomovirus symptoms. Symptomatic plants were photographed for the sake of comparison with asymptomatic plants (Figure 4.1). Details of sample location and symptoms are provided in (Table 4.1).

3.2. Plant Total DNA Extraction:

Total DNA was obtained using the CTAB protocol published by (Doyle & Doyle., 1987) with some modifications. 1g of plant leaf sample was ground in a mortar and pestle using liquid nitrogen. CTAB buffer (2% Cetyl Tri methyl Ammonium Bromide (CTAB), 100mM Tris-HCl, 20mM EDTA, 1.4M NaCl, 2% CTAB powder, 02% mercaptoethanol) pre-heated at 65 ℃, was then mixed with the leaves powder. The resultant slurry was again incubated at 65 ℃ for 30 minutes with continuous shaking. After the incubation, the mixture was allowed to cool for 3-5 minutes. An equal amount of Chloroform Isoamyl-alcohol in the ration of (24:1) was then added and mixed gradually. The mixture was then centrifuged (Eppendorf Centrifuge 5804R) to separate the phases. After centrifugation tubes showed three different phases. The lower organic phase contains all plant proteins, the upper aqueous phase contains nucleic acids, and the middle phase separates the upper and lower phases. The superior phase was then transmitted into a fresh tube and ice-cold isopropanol was added in 2/3 ratio. This step allowed the DNA to precipitate. The thread-like structure of DNA was pulled out with the help of a glass rod and washed twice using a wash buffer (70% ethanol & 10mM Ammonium Acetate). Washing was done via centrifugation for 5 minutes (Eppendorf centrifuge, 5415R) at 10,000 rpm. After washing the upper phase was removed carefully keeping the pellet intact. The DNA pellet was then air-dried for 10-15 minutes. Pellet was then eluted in 1X TE (10mM Tris-HCl, 1mM EDTA) buffer and stored at -20℃.

3.3. DNA Quality Confirmation:

The quality of the extracted DNA was evaluated by running the DNA over agarose gel. DNA stained with ethidium bromide and run over 1% agarose gel prepared in TAE (10mM Tris-HCl, 1mM EDTA) buffer. Gel was then viewed in a UV transilluminator. The genomic DNA band was visualized via comparison with a 1kb gene ruler, which was run as a marker. To further

quantify the DNA more precisely, absorption ratios at 260/280 were determined using the Nanodrop (Berthold Colibri – Microvolume Spectrophotometer). The concentration of DNA determined in nanograms per microliter was then employed for further processing.

3.4. Dilution of Genomic DNA:

Genomic DNA was diluted 10 folds to make its concentration workable for further processing. Dilutions were made using the below-mentioned formula:

$$M1V1 = M2V2$$

3.5. PCR Amplification:

The polymerase chain reaction (PCR) was first carried out using diagnostic primers followed by gene-specific and then full-length primers to obtain the full-length genomic components. PCR mixture was prepared containing the following reagents: 10X Taq buffer with KCl (5µl), 1.5 mM MgCl2 (3µl), 2mM dNTPs (2µl), 2 µl of each 50µM primers (both Forward and Reverse), 100 ng of total DNA, 0.5U Taq Polymerase (Thermo Fisher Scientific) (1µl), sterilized water up to the volume of 50 µl. The PCR reaction was set in a thermal cycler (Applied Biosystems). Reaction conditions were set as; preheating at 94℃ for 1 min, 30 cycles of (denaturation at 94℃ for 1 min, primer annealing at 55℃ for 1 min, extension at 72℃ for 3 mins) followed by a final extension at 72℃ for 10 mins (Figure 3.1 & 3.2). Details of primers used in this study are presented in a tabular form below (Table 3.1).

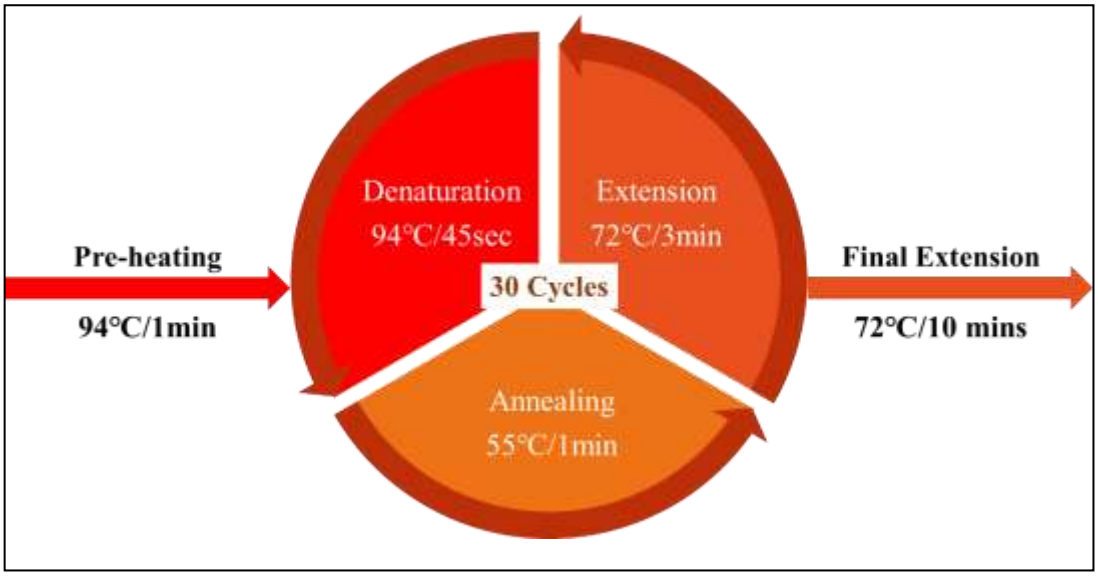

Figure 3.1: PCR amplification cycle for DNA-A

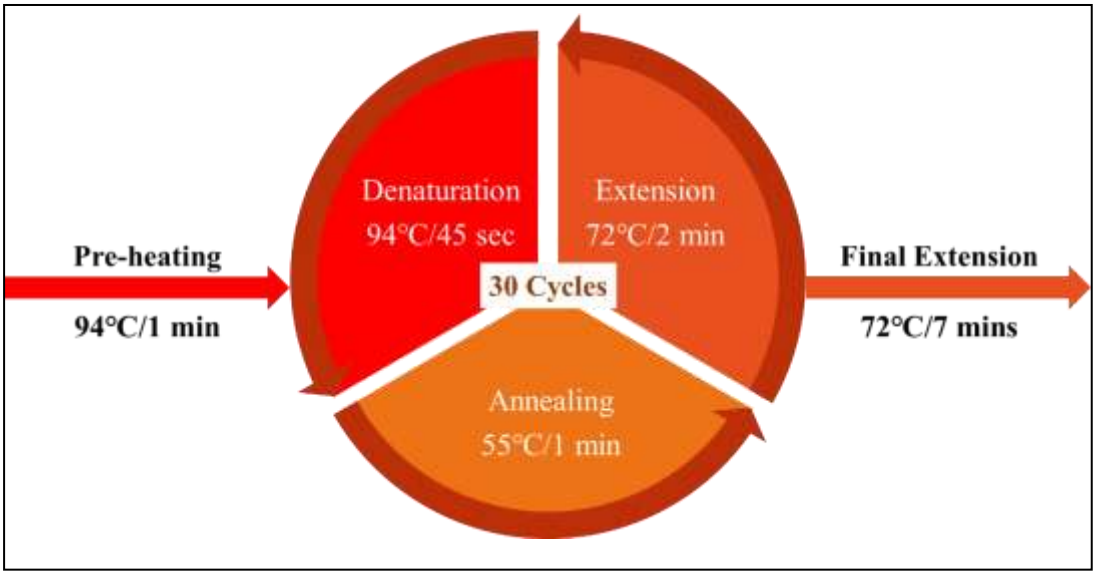

Figure 3.2: PCR amplification cycle for betasatellite

Table 3.1: Names, sequences and targeted component of primers used in this study.

Primer	Sequence (5'……3')	Targeted Component
CLCV1 **CLCV2**	5'-CCGTGCTGCTGCCCCCATTGTCCGCGTCAC-3' 5'-CTGCCACAACCATGGATTCACGCACAGGG-3'	DNA A, partial
SON-F **SON-R**	5' -GGGCCCCCATGAACTCTTTAAAGTG- 3' 5'-GGGCCCAAAGGGACTGGCAATC-3'	DNA-A
BURN-F **BURN-R**	5'-CCATGGTTGTGGCAGTTGATTGACAGATAC-3' 5'-CCATGGATTCACGCACAGGGGAACCC-3'	DNA-A
BURX-F **BURX-R**	5'-CTCGAGAGTGTCCCCGTCCTTGTCG-3' 5'-CTCGAGTGGGGAGAGTTTCAGATCG-3'	DNA-A
KTB-F **KTB-R**	5'-CTGCAGCATCATTTGTGAGCGCATATTC-3' 5-CTGCAGAGGTCACCTTGTCATTTCCTTC-3'	DNA-B

Beta01 **Beta02**	5'-GGTACCACTACGCTACGCAGCAGCC-3' 5'-GGTACCTACCCTCCCAGGGGTACAC-3'	Beta Satellites

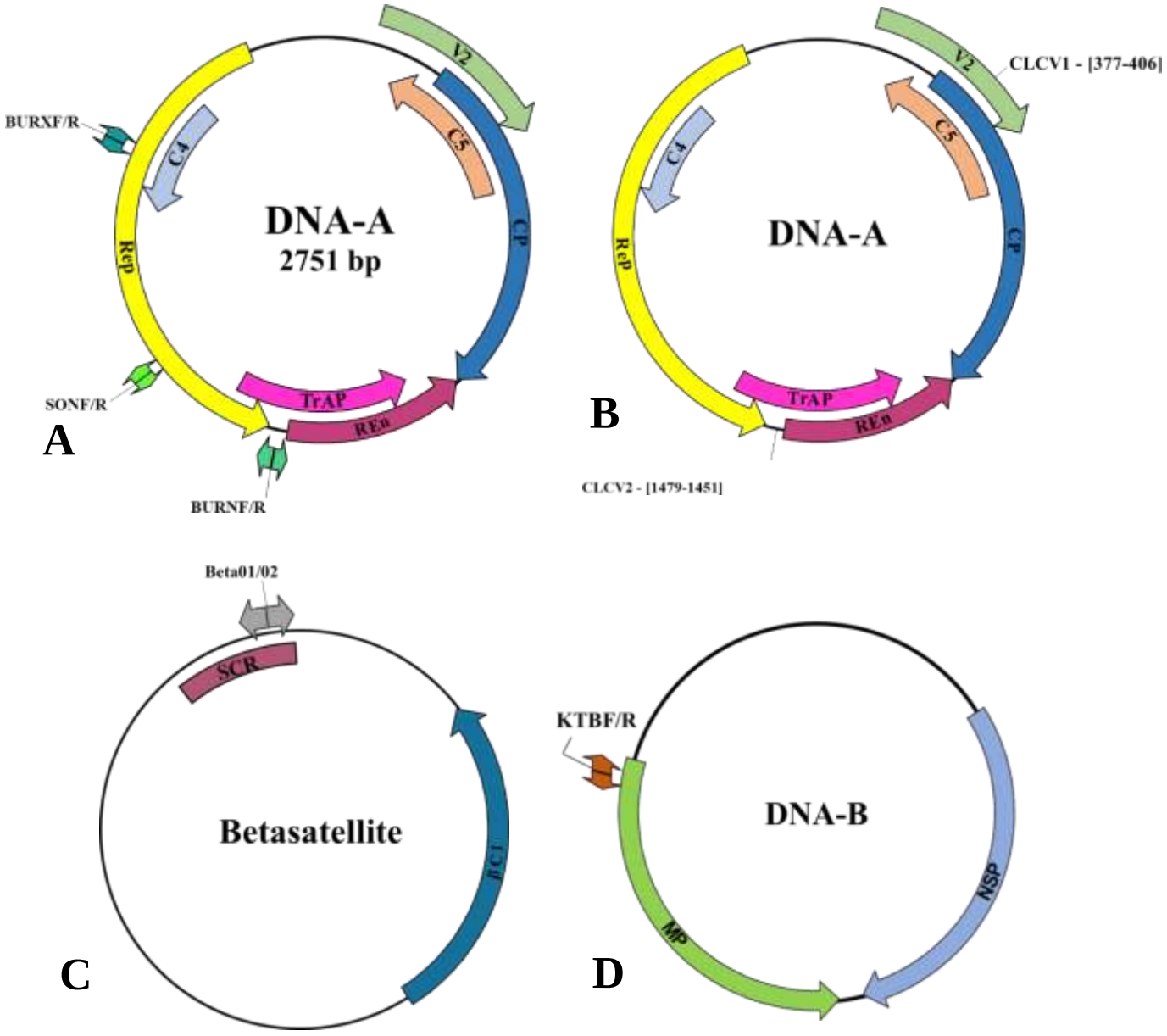

Figure 3.3: DNA-A, DNA-B, and betasatellite genome maps showing the primer location and orientation of all primer sets used in this study. **(A)** Genome map of DNA-A showing the primers locations of abutting primers, SONF/R, BURNF/R, and BURXF/R, **(B)** DNA-A showing the primer sites of diagnostic primers CLCV1/2 **(C)** Genome map of betasatellite showing the primer sites for universal primers Beta01/02 **(D)** Genome map of DNA-B showing the primer sites for abutting primers KTBF/R

3.6. Agarose Gel Electrophoresis

PCR products were run over 1% agarose gel prepared in TAE (25 mM Tris, Glacial acetic acid, and 1 mM EDTA) buffer. For the preparation of 1% agarose gel in 50ml volume of TAE buffer, 0.5g agarose powder was mixed into a 250ml flask confining TAE buffer. The flask was placed in a microwave oven for 1 min until it boiled, and the solution became clear. Gel was then allowed to cool for at least 10 mins. Once the gel got cool enough that no vapours were shown, 5µl of ethidium bromide 0.1% (w/v) was added to it. The gel was then mixed gently while keeping the flask covered with aluminium foil and transferred into a gel casting with a suitable comb. Gel was allowed to solidify for 30 mins. PCR products were then mixed with DNA loading dye [0.25% bromophenol blue dye (w/v), 30% glycerol (v/v), and 0.25% xylene cyanol FF dye (w/v)] loaded into the wells. 1kb gene ruler was run as a marker. The gel was run in a TAE buffer at 100 volts for 45 mins. Gel was then viewed in a UV transilluminator, and desired size bands were excised for further processing.

3.7. PCR Product Purification

For the purification of PCR products from agarose gel two different kits were used throughout the study. First, the purification was done via **Silica Bead DNA Gel Extraction Kit** (Thermo Fisher Scientific) as per the producer's instructions. The PCR product was weighed, and binding buffer was added into it, 3 times the weight of the gel. The gel was the allowed to melt at 55℃. The colour of the solution was noted at this step. Once the gel melted, 5µl silica bead powder suspension was added to it and vortexed thoroughly to allow the DNA to bind with silica beads. Centrifugation of mixture was then carried out for 1 min at 10,000 rpm. After centrifugation, silica beads formed a pellet. The upper phase was removed, and the silica beads pellet was washed with a silica bead wash buffer twice. After washing, the pellet was allowed to and resuspended in an elution buffer (TE). Centrifugation was carried out at 10,000 rpm for 1 min, which allowed silica beads to settle down while leaving the DNA in the elution buffer. The supernatant carrying the DNA was then picked carefully without disturbing the pellet and transferred into a fresh tube. Another method used for PCR product purification employed **GeneJET Gel Extraction Kit** (Thermo Fisher Scientific). A gel containing the PCR product was weighed and 1 volume of binding buffer was added. The tube was then incubated at 55°C until the gel melted. Gel bands melted in binding buffer were then transferred into the Gene jet purification column and centrifuged at 10,000 rpm for 1 min. Flow through was discarded and an extra 100µl of binding buffer was added. The column was centrifuged again at previously mentioned conditions and flow through was disposed of. Now 700 µl of wash buffer was added

and centrifuged. After this, an empty spin was given to remove any entangled ethanol. Finally, the column was then transferred into a fresh tube and eluted into 25-30 µl of elution buffer. The eluted DNA was stored at -20℃ for further processing.

3.8. DNA Sequencing

PCR-purified products were shipped to Macrogen, Korea for DNA sequencing via the Sanger sequencing method. Product-specific primers were used for initial sequencing. Internal primers were designed to acquire full-length sequences. Sequence contigs were assembled and read in DNASTAR (https://www.dnastar.com/software/lasergene/).

3.9. Sequence Analysis

DNA sequences were initially analyzed to find out the sequence similarity with begomoviruses DNA sequences already submitted in the database, via the NCBI nBLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequence similarities were checked and compared. Sequences were assembled via DNAStar software and full-length sequences were also compared via the BLASTn tool. Full-length sequences were then analyzed via the NCBI ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/) tool to determine the number of open reading frames present in them. Each ORF was then individually BLAST to check its similarity with the ORF sequences already submitted in the database.

3.10. Phylogenetic Analysis

Phylogenetic analysis was done using the software MEGA 11 (https://www.megasoftware.net/). Sequences were first aligned using the MUSCLE algorithm and a phylogenetic tree was contructed employing the Maximum likelihood method using a 1000 bootstrap value. The resulting phylograms were edited and labelled in iTOL (interactive Tress of Life) (https://itol.embl.de/upload.cgi).

3.11. Sequence Demarcation Tool (SDT) Analysis

A sequence demarcation tool (http://web.cbio.uct.ac.za/~brejnev/) was used for the SDT analysis of CLCuMuV. All sequences of CLCuMuV were downloaded from the public database NCBI. The sequence file was aligned in SDT v1.4 via the MUSCLE algorithm. The resulting files were saved and interpreted in M.S Excel. A percentage identity score table of all six strains of CLCuMuV is generated.

CHAPTER 4: RESULTS

4.1. Sample Collection:

Symptomatic plant samples of four different plant species were collected during a routine survey from two different locations in Islamabad. Names and other details are mentioned in (Table 4.1). These plants are ornamental and are available throughout the year. However, symptom severity may enhance during the summer season. Asymptomatic leaves were also collected as a negative control. Samples were packed in a zipper bag and placed at -80 for long-term use.

Figure 4.1: Apparently asymptomatic leaves of **(A)** *Hibiscus rosa sinensis* **(D)** *Duranta repens* **(G)** *Jasminum officinale* **(B)** & **(C)** Symptomatic leaves of *Hibiscus rosa sinensis* **(E)** & **(F)** Symptomatic leaves of *Duranta repens* **(H)** & **(I)** Symptomatic leaves of *Jasminum officinale*.

Table 4.1: Sample names, exhibited symptoms, location, and date of collection.

Sample No.	Plant Name	Common Name	Symptoms	Location	Collection Date
1	*Hibiscus rosa sinensis*	China rose	Vein thickening	NUST, Islamabad	23-09-2022
2	*Hibiscus rosa sinensis*	China rose	Vein thickening	NUST, Islamabad	26-09-2022
3	*Duranta repens*	Pigeon berry	Leaf curling	F-8, Islamabad	24-09-2022
4	*Jasminum officinale*	Jasmine	Leaf Mottling	NUST, Islamabad	25-09-2022

4.2. DNA Extraction

Total DNA was extracted via Cetyl Tri methyl Ammonium Bromide (CTAB) method and quantified by running it over 1% agarose gel. Gel was visualized in UV transilluminator. Good quality DNA was obtained (Figure 4.2) and diluted further for PCR.

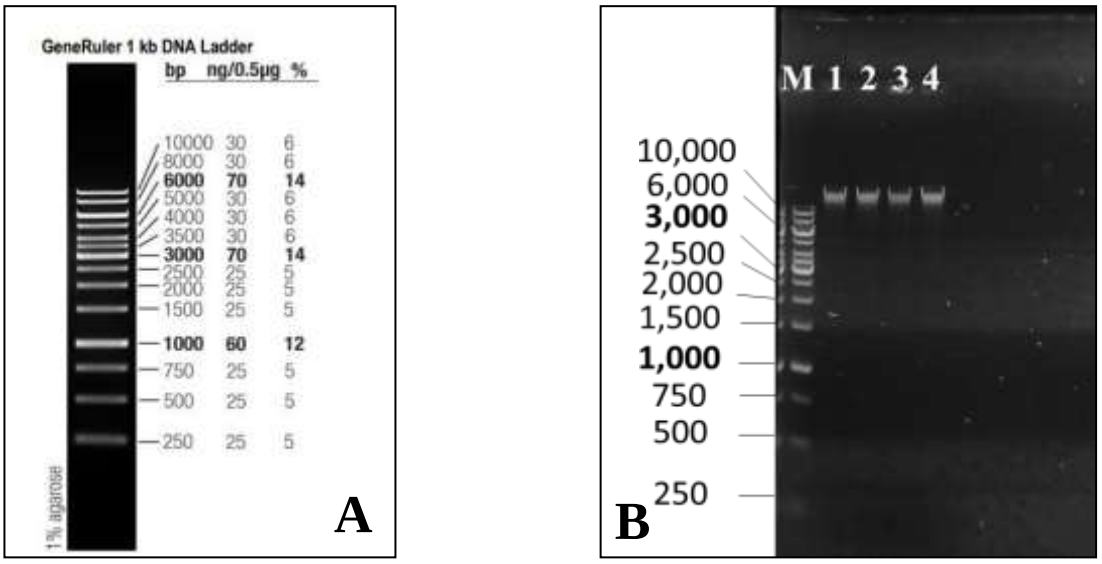

Figure 4.2: (A) 1kb gene ruler (0.5µg/lane) **(B)** 1% Agarose gel illustrating total DNA of different plant samples; Lane (M) is showing the 1kb gene ruler, Lane (1) *D. repens*, Lane (2 & 3) *H. sinensis*, Lane (4) *J. officinale*.

4.3. Detection of Begomovirus through Diagnostic Primer Pair

Initially, to confirm the infection of potential begomoviruses in the apparently infected leaf samples, PCR reactions were run employing the diagnostic/universal primer sets. Diagnostic primer sets are designed in a way that they can amplify the maximum number of concerned sequences. Usually, such primers are designed from a highly conserved region. In this study one set of diagnostic primer pair, (CLCV1/2; Table 3.1) was used to initially confirm the presence of begomovirus DNA-A in the sample Hib-1 and Hib-2. CLCV1/2 has the potential to amplify cotton infecting begomoviruses and give a product of ~ 1kb, starting from the region of V2 overhanging CP till the mid of REn. (Figure 3.3**)** shows the anticipated amplification region of CLCV1/2. CLCV1/2 showed positive results with both samples, Hib-1, and Hib-2. A desired band of approximately 1.1kb was obtained with both samples. Results of CLCV1/2 on 1% agarose gel are shown below (Figure 4.3) in the form of clear bands of almost 1kb.

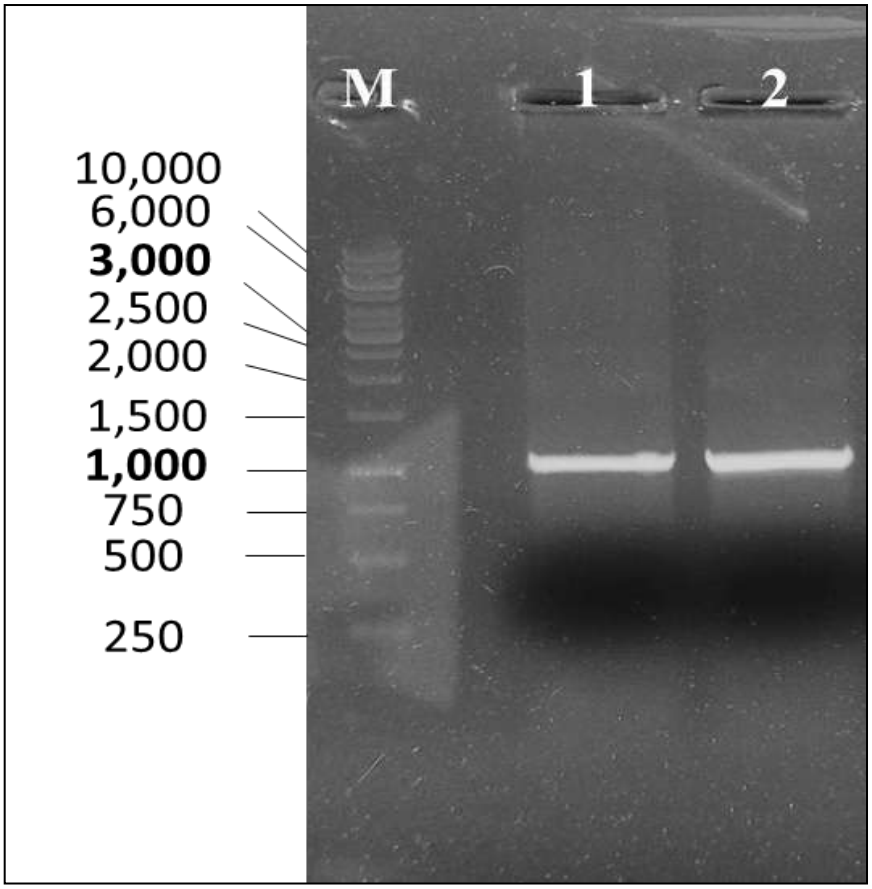

Figure 4.3: 1% Agarose gel illustrating PCR amplified product; Lane (M) is showing the 1kb gene ruler, Lane (1) & Lane (2) showing 1kb bands amplified via CLCV1/2.

4.4. Amplification of Full-Length Begomovirus(es) using Abutting Primer Pairs

Following the confirmation using diagnostic primers, PCR was run using full-length primer sets. Abutting primer pairs used in this study were SonF/SonR, BurxF/BurxR, BurnF/BurnR for DNA-A and KTBF/KTBR for DNA-B, as mentioned earlier. Primer pair SonF/SonR are specific for several begomoviruses and thus can amplify a wide range of begomoviruses. The primer sets BurxF/BurxR and BurnF/BurnR are more specific for cotton-infecting begomoviruses. Sample Hib-1 and Hib-2 gave positive results with all the above-mentioned primer sets. Duranta sample gave positive results with SonF/SonR and BurxF/BurxR. While Jasmine sample gave a positive result with primer set SonF/SonR.

A universal primer pair, Beta01/02 (Briddon *et al.*, 2002) was used for the detection of begomovirus-associated betasatellite (Table 3.1). As described in Chapter 1, monopartite begomoviruses exist in the form of the complex along with ssDNA satellite molecules. Therefore, the presence of *begomovirus*-associated betasatellite is a clear indication of begomovirus infection. Beta01/02 is a universal set of primers that can amplify most of the begomovirus-associated betasatelites. The primer site is mentioned in (Figure 3.3). Sample Hib-1, Hib-2 and Duranta showed positive results with primer set Beta01/02, while jasmine sample showed negative results. For samples Hib-1 and Hib-2, clear 1.3kb bands were obtained each time, while for Duranta sample multiple bands or sometimes a good band of 1kb was obtained. These results are a clear indication of the presence of begomovirus-betasatellite complex in samples Hib-1, Hib-2, and Duranta. A full-length betasatellite of ~1.3kb from sample Hib-1, Hib-2, and Duranta was amplified using the primer set Beta01/02.

Bipartite begomoviruses are usually not associated with satellite molecules. As Jasmine sample did not show any amplification with Beta01/02 primers, therefore it was tested using another set of primers, KTBF/R. KTBF/R are designed for the amplification of full-length DNA-B component of begomoviruses and give an amplification product of approximately 2.8kb. The nucleotide position site for KTBF/R is shown in (Figure 3.3**).** The Jasmine sample showed positive results with primer set KTBF/R. This indicates the presence of bipartite begomoviruses in Jasmines. All samples except Jasmine, showed negative results with the primer KTBF/R. A full-length amplification product of ~ 2.8 was amplified from Jasmine sample via the primer set KTBF/R. Each amplification product was named separately, and a summary table of amplification products is given below (Table 4.2). All PCR amplified products were run over 1% agarose gel and displayed along with proper labelling below (Figures 4.4 & 4.5)**.**

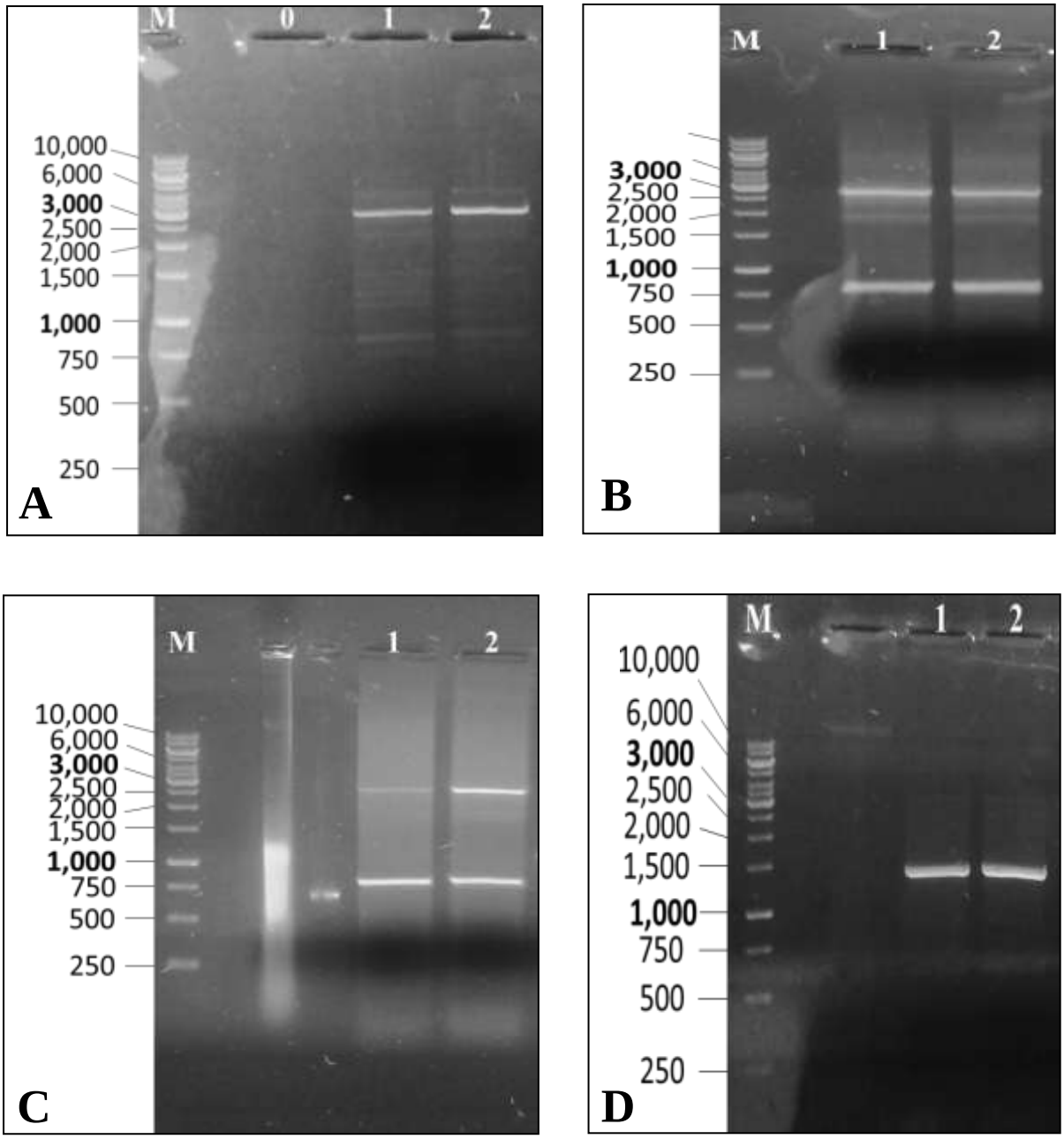

Figure 4.4: 1% Agarose gel illustrating **(A)** Lane (M) is showing 1kb gene ruler and Lane (1 & 2) showing ~2.8kb amplification product **(HAHN)** of primers BurxF/BurxR from the sample Hib-1 **(B)** Lane (M) is showing 1kb gene ruler and Lane (1 & 3) showing ~2.8kb amplification product **(HAS)** of the primers SonF/SonR from the sample Hib-1. **(C)** Lane M is showing 1kb gene ruler and Lane (1 & 2) showing ~2.8kb amplification product **(HAH)** of primers BurnF/BurnR from the sample Hib-2 **(D)** Lane M is showing 1kb gene ruler and Lane (1 & 2) showing ~1.3kb amplification products **(HAB and HAB1)** of the primers Beta01/02 from the sample Hib-1 & Hib-2 respectively.

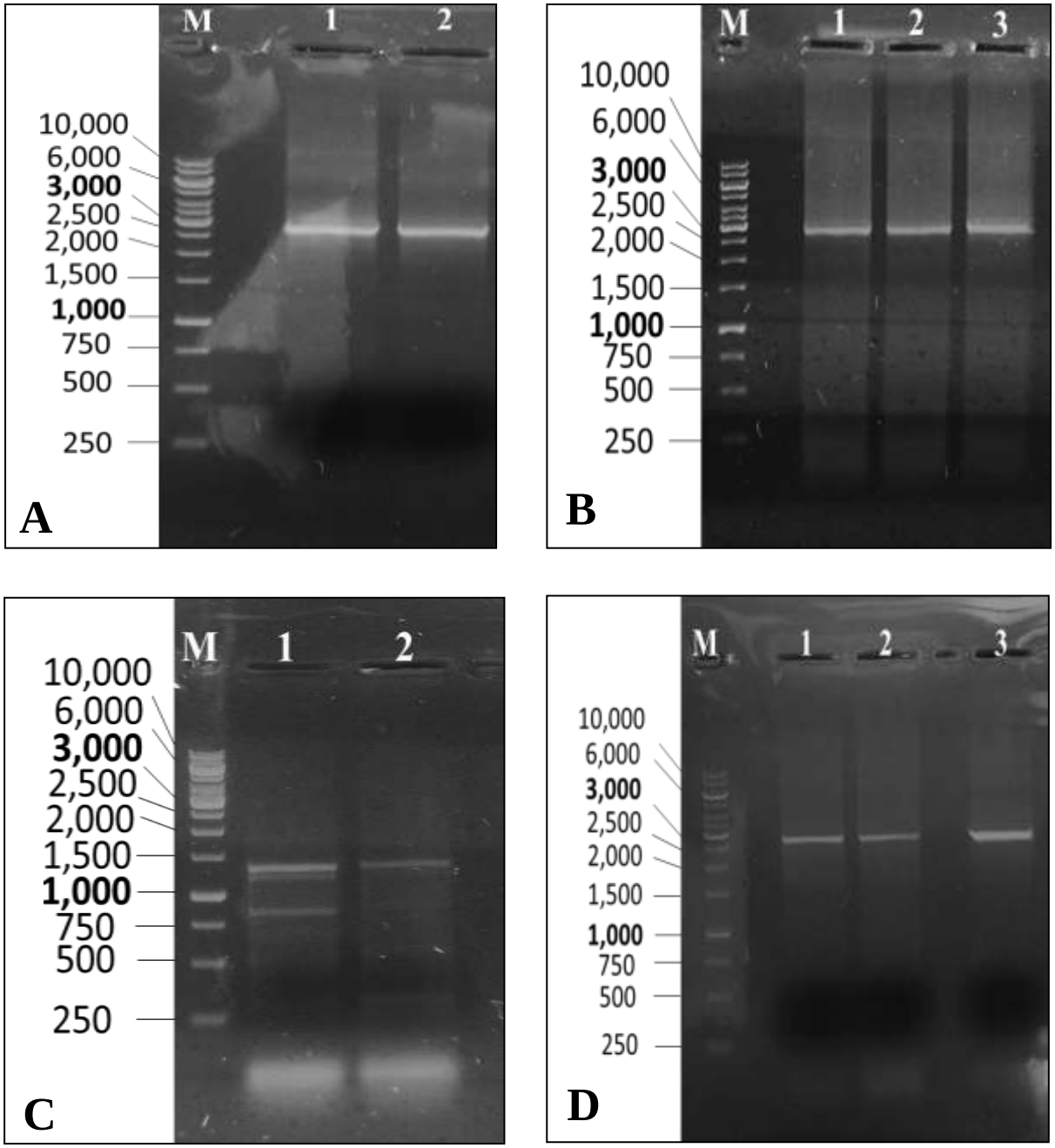

Figure 4.5: 1% Agarose gel illustrating **(A)** Lane M showing 1kb gene ruler and Lane (1 & 2) showing ~2.8kb amplification product **(DA)** of primers SonF/SonR from the Duranta sample **(B)** Lane M showing 1kb gene ruler, Lane (1,2 & 3) showing ~2.8kb amplification product **(DAX)** of the primers BurxF/BurxR from the Duranta sample **(C)** Lane M showing 1kb gene ruler, Lane (1 & 2) showing ~1.3kb amplification product **(DB)** of primers Beta01/02 from Duranta sample. **(D)** Lane M showing 1kb gene ruler, Lane (1 & 2) showing ~2.8kb amplification product **(JA)** of the primers SonF/SonR, Lane 3 is showing the amplification product **(JB)** ~2.8kb of primers KTBF/R from Jasmine sample.

Table 4.2: PCR products summary table, showing the details of positive results for each sample.

Sample	Positive Result Primers	Diagnostic/Full-length	Isolate Code	Component	Size
Hibiscus-1	CLCV1/2	Diagnostic	**HD2**	DNA-A	1.1kb
	SonF/SonR	Full-length	**HAS**	DNA-A	2.8kb
	BurxF/BurxR	Full-length	**HAHN**	DNA-A	2.8kb
	Beta01/02	Full-length	**HAB**	Betasatellite	1.3kb
Hibiscus-2	CLCV1/2	Diagnostic	**HD1**	DNA-A	1.1kb
	BurnF/BurnR	Full-length	**HAH**	DNA-A	2.8kb
	Beta01/02	Full-length	**HAB1**	Betasatellite	1.3kb
Duranta	SonF/SonR	Full-length	**DA**	DNA-A	2.8kb
	BurxF/BurxR	Full-length	**DAX**	DNA-A	2.8kb
	Beta01/02	Full-length	**DB**	Betasatellite	1.3kb
Jasmine	SonF/SonR	Full-length	**JA**	DNA-A	2.8kb
	KTBF/R	Full-length	**JB**	DNA-B	2.8kb

4.5. Potential Begomovirus Elution from Agarose Gel

PCR bands were excised and cleaned via PCR purification kit (GeneJET Gel Extraction Kit, Thermo Scientific). For quality confirmation, all cleaned products were run over 1% gel (Figure 4.6) for quality confirmation and presented below.

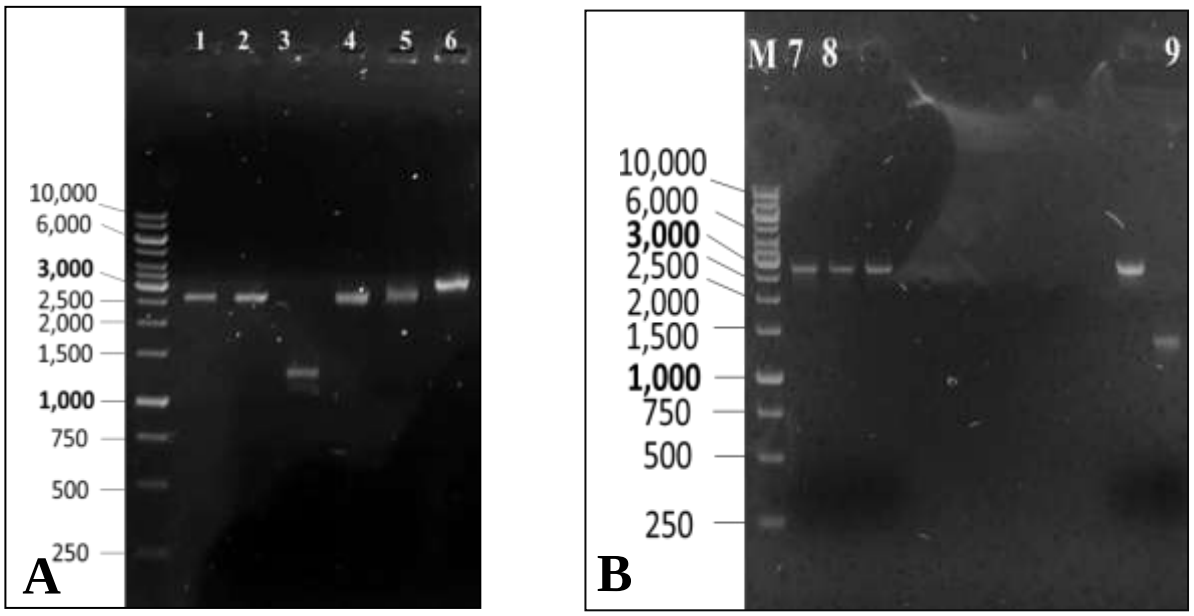

Figure 4.6: 1% Agarose gel illustrating (A) Lane M is showing 1kb gene ruler and Lane 1 – 6 are showing gene clean products of isolates DA, DAX, DB, HAHN, JA, and JB (B) Lane M is showing 1kb gene ruler, Lanes 7 & 8 are showing isolate HAH and HAS, and Lane 7 is showing gene clean product of isolate HAB.

4.6. DNA Sequencing

PCR products were sequenced from both orientation and then using internal primer to get the complete sequence. The resulting sequences were analyzed via NCBI BLASTn, which revealed that all 2.8 kb amplicons showed similarity with DNA-A of distinct *begomovirus* species while 1.3kb sequences showed similarity with Cotton leaf curl disease associated betasatellites.

4.7.1 Sequence Analysis of HAHN Isolate

Upon assembling the sequence contigs for the isolate, HAHN, a complete sequence was obtained. The length of complete sequence length was 2751 bases. Initially the sequence was analyzed via the NCBI nBLAST tool. BLAST analysis reveals that the sequence for isolate, HAHN shared the maximum of 98% sequence identity with *Cotton leaf curl Multan virus* (CLCuMuV). This result implied that HAHN is a variant of CLCuMuV. Further the sequence was then analyzed via the NCBI ORF finder tool. ORF finder showed several ORFs at both complementary and virion strands of the isolate HAHN. Each ORF was then individually

analyzed via the NCBI pBLAST tool. Seven overlapping ORFs were then characterized based upon the nucleotide sequence identity with the sequences already present in the database. Out of seven potential coding ORF two, V2 and coat protein (CP), were in virion sense, while five others, Replication enhancer (Ren), Transcriptional activator (TrAP), Replication associated protein (Rep), C4, and C5, were in complementary sense. Genome characterization of HAHN also revealed the conserved hairpin loop containing the nonanucleotide region (TAATATTAC). Start and end site of each gene, number of nucleotide and amino acids were analyzed and mentioned in Table 4.3. A graphic illustration of for the HAHN isolate was drawn with the help pDRAW32 and presented below Figure 4.7.

Table 4.3: Nucleotide start and end position of seven overlapping genes from HAHN Isolate, encoded protein sizes, and highest percentage identity with the sequences already present in the database.

Isolate	ORF	Start (nt)	End (nt)	Length (nt/aa)	Protein Size (kDa)	Highest percentage identity (%)
HAHN	**Rep**	1495	2586	1092 | 363	40.7	98.62% CLCuMV - Rep
	CP	276	1046	771 | 256	29.66	99.2% CLCuMV - CP
	TrAP	1146	1598	453 | 150	17.28	96% CLCuAV - TrAP
	REn	1049	1453	405 | 134	15.68	97% CLCuMV -REn
	V2	116	481	366 | 121	13.98	100% CLCuMV - V2
	C4	2127	2429	303 | 100	11.261	97% CLCuKoV- C4
	C5	180	584	405 | 134	15.68	99% CLCuMV - C5

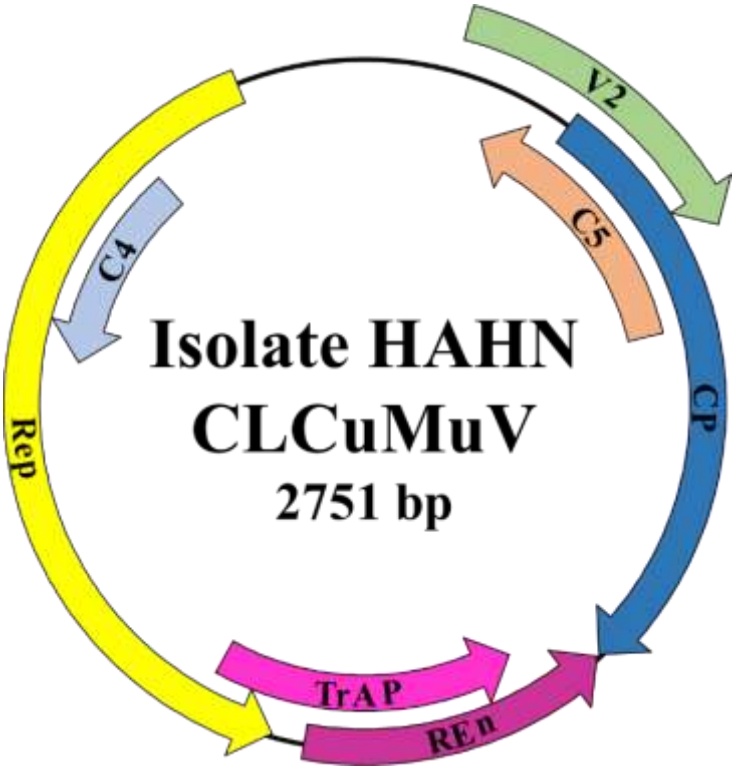

Figure 4.7: Complete genome map for isolate, HAHN showing organization of seven overlapping genes. (Created in pDRAW32).

4.7.1.1. Analysis of Intergenic Region and Rep Conserved domains

Begomovirus' intergenic region includes stem-loop region containing the highly conserved nonanucleotide and virus specific repeated sequences called iterons. Iterons are known to be the binding sites for beogomovirus encoded Rep, which is an important component of virus replisome. The intergenic region of the isolate, HAHN was thus analyzed to determine the stem-loop region and potential presence of iterons. Analysis revealed the existence of stem-loop region with the highly conserved nonanucleotide, TAATATTAC, TATA boxes, CA motif, and iterons spread around the TATA boxes Figure 4.8. Iterons correspond to the replication associated protein, Rep, which also contain conserved domain and motifs. Therefore, conserved region of Rep undergoes modifications as do the iterons modify from species to species. The analysis for conserved region of Rep was done via the NCBI CDD (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Analysis showed the Rep of HAHN isolate contains two conserved domains, Geminivirus Rep catalytic domain (pfam00799)**,** and geminivirus Rep protein central domain (pfam08283) Figure 4.10 (A). Conserved domains are mentioned in the Figure 4.10 (A) & (B).

```
HAHN        :  TATTTTACTCTGGCAATTGGAGACAGGAGACAATATATAGTGTCTCCAAATGGCATATTC    : 2518
MK482365.1  :  C-TTTTACTCTGGCAATTGGAGACTGGAGACAATATATAGTGTCTCTAAATGGCATATTC    : 2517
NC          :  --TTTTACTCTGGGAATTGGAGACTAGATACAATTTATAGTGTCTCCAAATGGCATATTC    : 2518
AJ002447.1  :  --TTTTACTCTGGGAATTGGAGACTAGATACAATTTATAGTGTCTCCAAATGGCATATTC    : 2518
AJ002458.1  :  --TTTTACTCTGGGAATTGGAGACTGGATACAATTTATAGTGTCTCCAAATGGCATATTC    : 2518
AJ496287.1  :  --TTTTACTCTGGGAATTGGAGACTGGATACAATTTATAGTGTCTCCAAATGGCATATTC    : 2517
KF413618.1  :  --TTTTACTCTGGCAATTGGAGACAGGAGACAATATATAGTGTCTCCAAATGGCATTATC    : 2517
MN698836.1  :  --TTTTACTCTGGCAATTGGAGACAGGAGACAATATATAGTGTCTCCAAATGGCATTATC    : 2517


HAHN        :  TGTAAATAACTCGA-AGTTCGTTTGAAATTCAAATTCCCCTTTGGGCTCCAAAAGCGGCC   : 2577
MK482365.1  :  TGTAAATAACTAGA-AGTTCGTTTGAAATTCAAATTCCCCTTTGGGCTCCAAAAGCGGCC   : 2576
NC          :  TGTAAATAACTAGA-AGTTCGTTTGAAATTTAAATTCCCCTTTGGGCTCCAAAAGCGGCC   : 2577
AJ002447.1  :  TGTAAATAACTAGA-AGTTCGTTTGAAATTTAAATTCCCCTTTGGGCTCCAAAAGCGGCC   : 2577
AJ002458.1  :  TGTAAATAACTAGA-AGTTCGTTTGAAATTTAAATTCCCCTTTGGGCTCCAAAAGCGGCC   : 2577
AJ496287.1  :  TGTAAATAACTAGA-AGTTCGTTTGAAATTTAAATTCCCCTTTGGACTCCAAAAGCGGCC   : 2576
KF413618.1  :  -----GTAATTTGAGAAATCATTTCAAAAT ---------CCTCACGATCCAAAAAGCGGCC   : 2564
MN698836.1  :  -----GTAATTTGAGAAATCATTTCAAAAT ---------CCTCACGCTCCAAAAAGCGGCC   : 2564


HAHN        :  ATCCGTATAATATTACCGGATGGCCGCGCGATTTTTT-TGTGGGCCCCCGATTTATGAGA    : 2636
MK482365.1  :  ATCCGTATAATATTACCGGATGGCCGCGCGATTTTTT-TGTGGGCCCCCGATTTATGAGA    : 2635
NC          :  ATCCGTTTAATATTACCGGATGGCCGCGCGATTTTTTATGTGGGCCCCCGAGTTATGAGA    : 2637
AJ002447.1  :  ATCCGTTTAATATTACCGGATGGCCGCGCGATTTTTTATGTGGGCCCCCGAGTTATGAGA    : 2637
AJ002458.1  :  ATCCGTTTAATATTACCGGATGGCCGCGCGATTTTTTATGTGGGCCCCCGAGTTATGAGA    : 2637
AJ496287.1  :  ATCCGTTTAATATTACCGGATGGCCGCGCGATTTTTTATGTGGGCCCCCGAGTTATGAGA    : 2636
KF413618.1  :  ATCCGTATAATATTACCGGATGGCCGCGCGATTTTTTTTGTGGGCCCCCTATTTATGAGA    : 2624
MN698836.1  :  ATCCGTATAATATTACCGGATGGCCGCGCGATTTTTTTTATGGGCCCCCTATTTACGAGA    : 2624
```

Figure 4.8: Stem-loop region containing nonanucloetide, Iterons, cis-acting elements of the HAHN isolate.

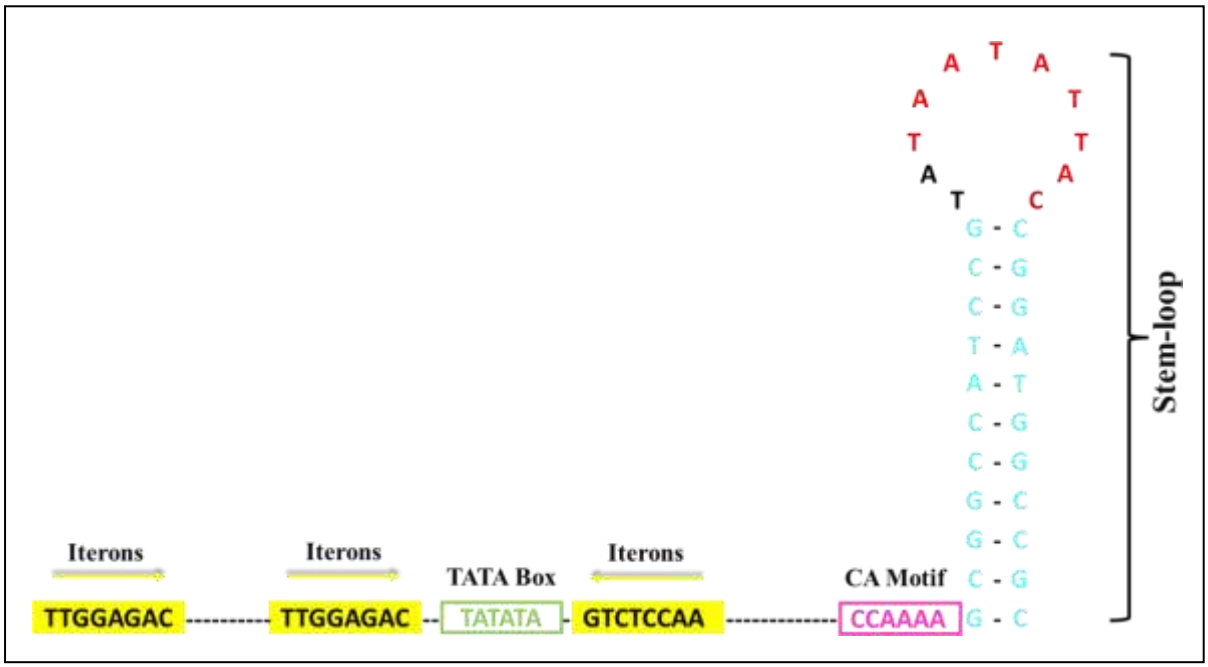

Figure 4.9: Graphical representation of intergenic region for the isolate, HAHN, showing the stem-loop region containing the nonanucleotide (shown in red), and other conserved regions.

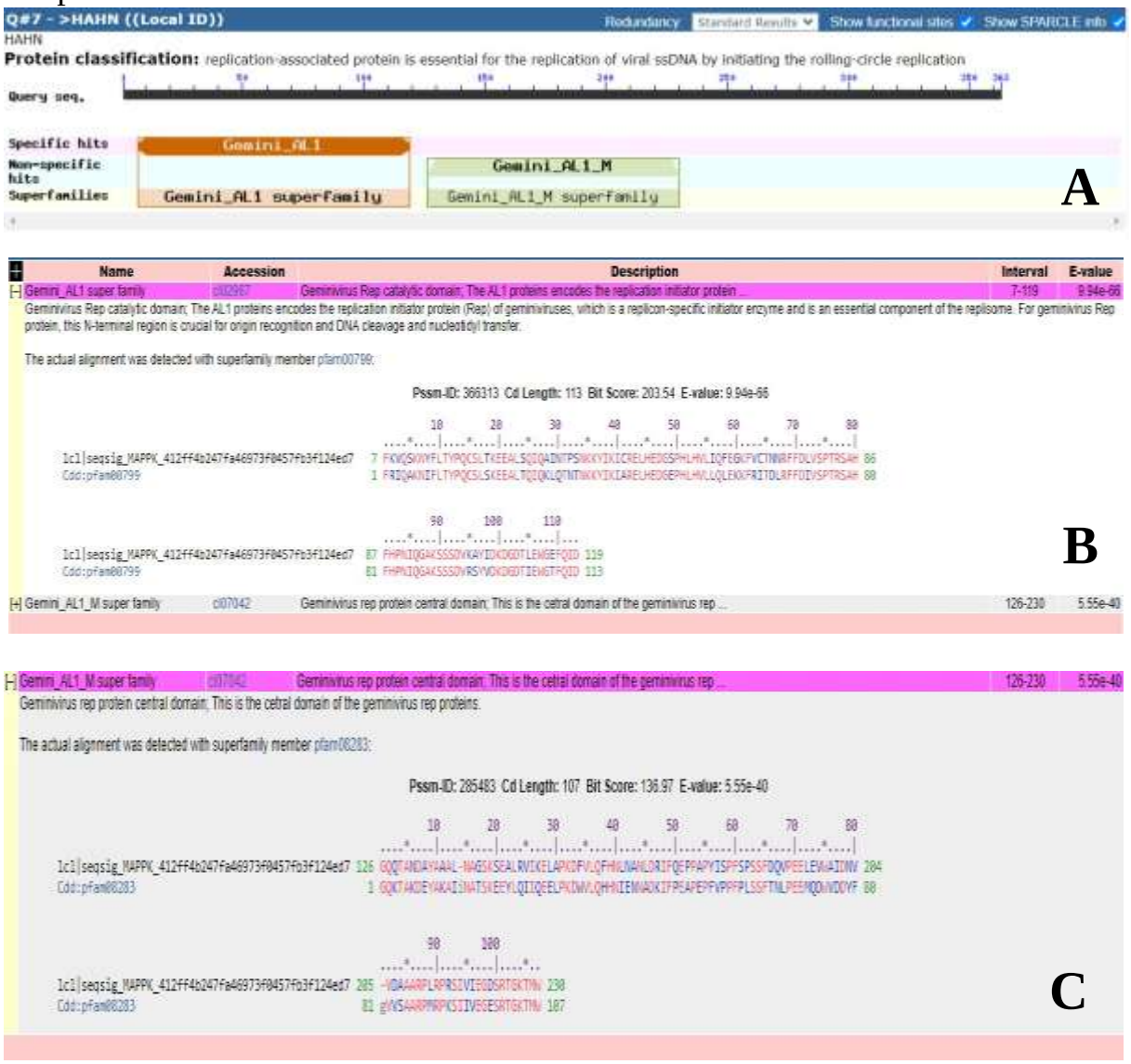

Figure 4.10: (A) NCBI CDD analysis showing the conserved domains present in HAHN isolate Rep protein (B) Graphical logo for Geminivirus_AL1 (C) Graphical logo for Geminivirus_AL1_M

4.7.1.2. Phylogenetic Analysis

Previous analysis revealed that HAHN isolate is a variant of CLCuMuV. Hence, phylogenetic analysis for HAHN isolate was performed by comparing the HAHN isolate of CLCuMuV with other four species of *Cotton leaf curl virus*, *Cotton leaf curl Alabad virus* (CLCuAlV), *Cotton leaf curl Bangalore virus* (CLCuBaV), *Cotton leaf curl Kokhran virus* (CLCuKoV), and *Cotton leaf curl Gezira virus* (CLCuGeV). Sequences were aligned in the MEGA11 employing the UPGMA algorithm. The aligned file is thus utilized for constructing the phylogenetic tree via the maximum-likelihood algorithm built in MEGA11 software. The phylogenetic relation displayed indicates the relationship of HAHN isolate with CLCuMuV clade Figure 4.8. The analysis was run with a bootstrap value of 1000 to obtain maximum certainty in the resulting tree. The CLCuMuV clade clustering the HAHN isolate showed a maximum bootstrap value of

1, which corresponds to 100% confidence. Thus, it is now more evident that HAHN isolate is a variant of Cotton leaf curl disease causing species of *begomovirus*, CLCuMuV.

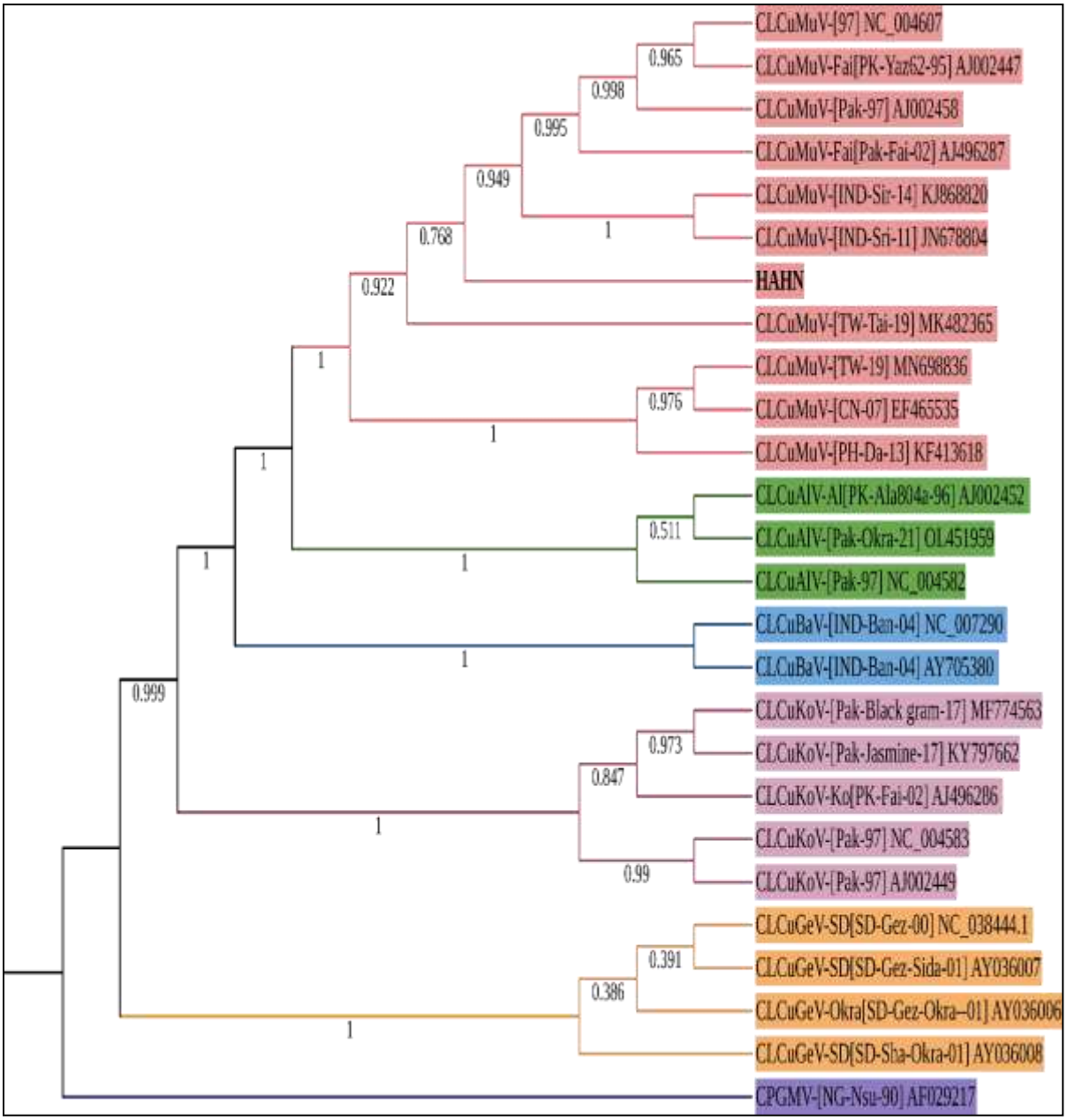

Figure 4.11: Cladogram constructed via Maximum likelihood method presenting **HAHN** clustered in CLCuMuV clade. Each color presents a different clade. Bootstrap value is mentioned at each clade. Other than CLCuMuV, clades of *Cotton leaf curl Alabad virus* (CLCuAlV), *Cotton leaf curl Bangalore virus* (CLCuBaV), *Cotton leaf curl Kokhran virus* (CLCuKoV), *Cotton leaf curl Gezira virus* (CLCuGeV) are shown in different colors. *Cowpea golden mosaic virus* (CPGMV) was used as an outgroup and presented in Purple.

4.7.1.3. Strain Demarcation Analysis of HAHN Isolate

Isolate HAHN belongs to the CLCuMuV, one of the species responsible for causing Cotton leaf curl disease. Presently there are six strains of CLCuMuV (Brown *et al.*, 2015). Therefore, SDT analysis was run to designate a strain of CLCuMuV for isolate, HAHN. SDT is a tool that can perform pairwise sequence comparison of sequences and thus recognized as standard for specie and strain demarcation. All sequences of CLCuMuV available in public database NCBI were collected. The percentage identity threshold for data collection was set at 90%. The analysis was run in SDT v1.4 employing the MUSCLE algorithm. The resultant matrix was then visualized and interpreted in M.S. Excel. The names and accession numbers of all six strains as mentioned in (Brown *et al.*, 2015), were used as a standard representation of each strain. Upon analyzing the matrix scores, it was interpreted that HAHN isolate, belongs to the Faisalabad strain of CLCuMuV. The final percentage identity table for all six strains of CLCuMuV is presented below.

Table 4.4: Percentage identities of all six strains of CLCuMuV. Matrix generated, interpreting the identity score matrix of SDT analysis.

Strains	CLCuMuV-Dawinii (9)	CLCuMu-Faislabad (82)	CLCuMuV-Hibiscus (2)	CLCuMuV-Hisar (8)	CLCuMuV-Pakistan (63)	CLCuMuV-Rajasthan (39)
CLCuMuV-Rajasthan (39)	81 - 89%	85 - 93%	83 - 87%	84 - 94%	87 - 97%	95-100%
CLCuMuV-Pakistan (63)	87 - 93%	87 - 94%	88 - 92%	87 - 94%	95-100%	
CLCuMuV-Hisar (8)	88 - 95%	87 - 95%	88 - 91%	95-100%		
CLCuMuV-Hibiscus (2)	89 - 90%	88 - 91%	99 - 100%			
CLCuMuV-Faislabad (82)	89 - 93%	95 - 100%				
CLCuMuV-Dawinii (9)	95 - 100%					

4.7.1.4. Phylogenetic Analysis of CLCuMuV Strains

A phylogenetic analysis was conducted for the further confirmation of HAHN isolate, with all six strains of CLCuMuV. A total of 225 full length sequences present in the database were employed for the phylogenetic analysis in MEGA X software. Sequences were aligned via the MUSCLE alignment using the default parameters. The aligned sequences were then used for construction of phylogenetic tree via the maximum likelihood method setting the bootstrap parameter to 1000. The resultant tree clearly indicates the relationship of HAHN isolate with CLCuMuV-Faislabad strain Figure 4.12. The other five strains are shown in different colors. This analysis clearly shows that the isolates from similar strains are clustered together regardless of their geographical region. The only overlap was observed between the few sequences of Faisalabad and Hisar strain.

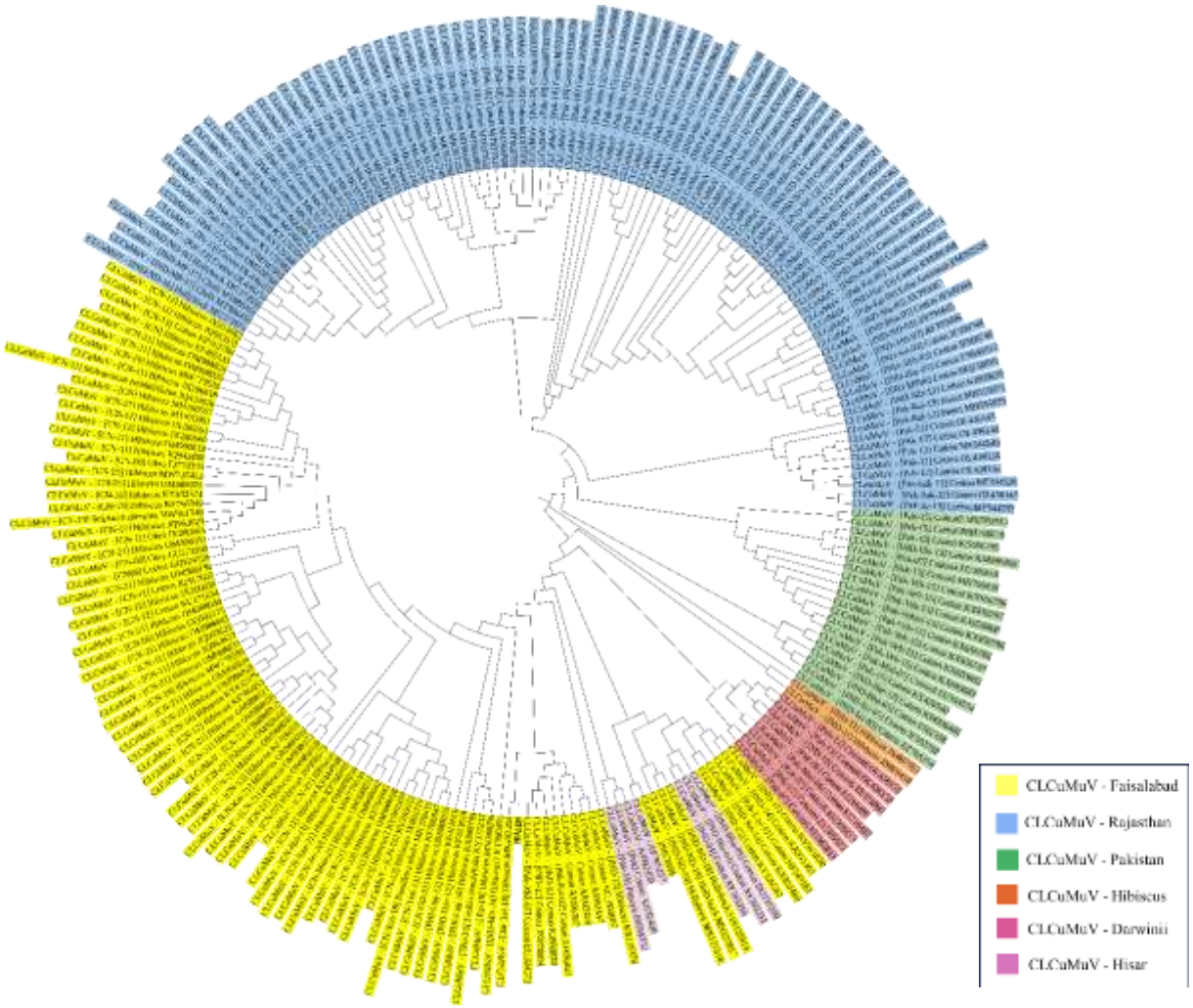

Figure 4.12: Cladogram of all 225 full-length sequences of CLCuMuV, constructed via Maximum likelihood method presenting **HAHN** clustered in CLCuMuV-Faislabad clade. Each color represents a clade for different strain.

4.7.2. Sequence Analysis of Isolate HAB

A total of 1346 bp were retrieved after assembling the contigs of HAB isolate. The sequence was analyzed in nBLAST, which revealed that the HAB isolate shared the highest nucleotide sequence identity of 96% with Cotton leaf curl Multan betasatellite (CLCuMB). This result confirmed the presence of begomovirus-betasatellite complex in Hibiscus plant. Isolate HAB was then further characterized via NCBI ORF finder. A single betasatellite gene, βC1, was found in the complementary sense encompassing a total of 357 nucleotides. Start and end sites for the ORF βC1 are given in Table 4.5. A conserved stem loop region containing the nonanucleotide (TAATATTAC) sequence was also characterized. A graphic illustration of the HAB isolate was drawn with the help pDRAW32 and presented below Figure 4.9.

Table 4.5: Nucleotide start and end position of seven overlapping genes from HAHN Isolate, encoded protein sizes, and highest percentage identity with the sequences already present in the database.

Isolate	ORF	Start (nt)	End (nt)	Length (nt/aa)	Protein Size (kDa)	Highest percentage identity (%)
HAB	**βC1**	551	195	357 | 118	13.74	100% CLCuMB- βC1

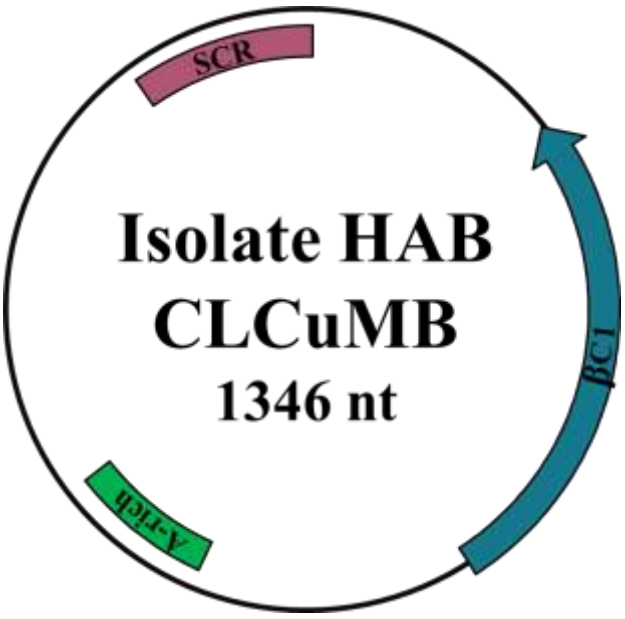

Figure 4.13: Complete genome map for isolate, HAB showing organization of single gene βC1, SCR, and A rich region. (Created in pDRAW32).

4.7.2.1. Intergenic Region Analysis for HAB, Isolate

Begomovirus associated betasatellite do not code their own Rep protein. Therefore, they relied upon the helper virus DNA-A component for their replication. Rep binds to the specific regions, known as iterons which are usually present in the LIR of the DNA-A. Therefore, the intergenic regions of HAS isolate were explored in the search of similar iterons, as demonstrated in case of HAHN. Analysis showed that the betasatellite also harbors similar Iterons (highlighted in yellow in **Figure 4.10).**

```
HAB  :  GTGATTGAAGAATAGTTAAAAAATGGAGACACCGATAGGTAAATTGTCTCCAATATATTGGAGAACA

HAB  :  GGGGACTCATAAATTGCCTTTCCTAAAATACCCCCGCTTTTGTGTCTAAG

HAB  :  ATAAAGTTAACATTCTCTCTCCTCTTTTGATCTC

HAB  : gacacgcgcggcggtgt

HAB  : G
```

A

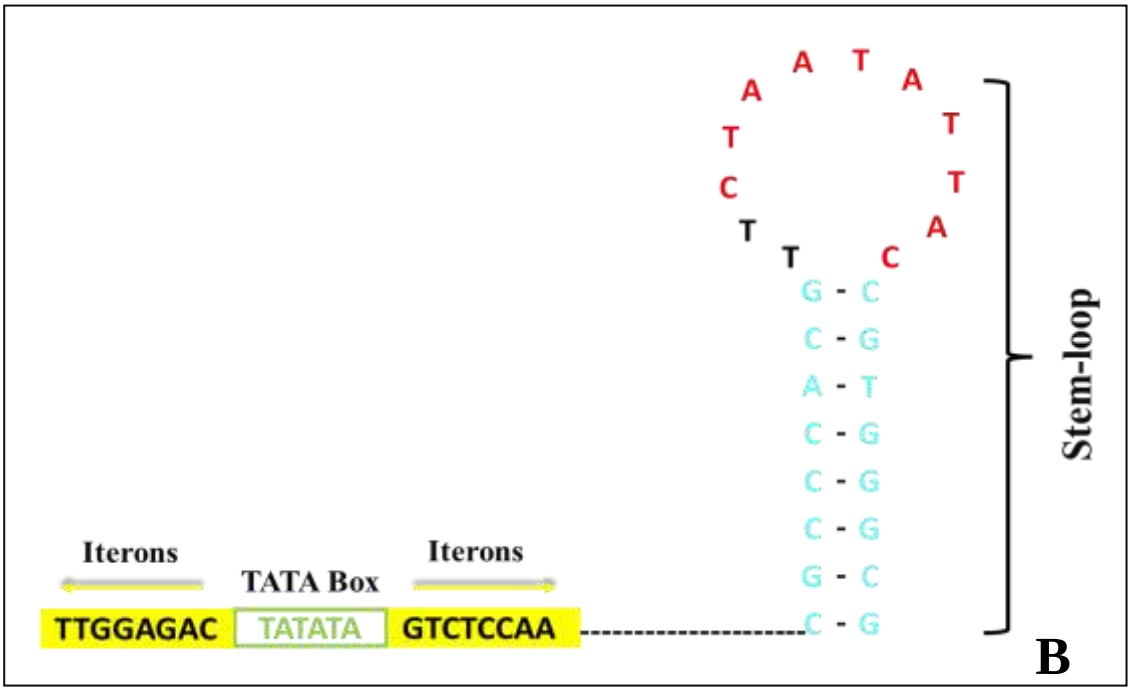

Figure 4.14: (A) Intergenic region of HAB isolate. Stem region is highlighted in blue, while nonanucloetide is shown is Red, sequence highlighted in yellow is demonstrating Iterons. **(B)** Graphical representation of HAB intergenic region, showing the stem-loop region containing the nonanucleotide (shown in red), and other conserved regions.

4.7.2.2. Phylogenetic Analysis

To further confirm the relation of HAB isolate with CLCuMB, a phylogenetic analysis was performed. For phylogenetic analysis, most closely related sequences were downloaded from public database, NCBI. A few distant sequences, Kroton leaf curl betasatellite were also pooled in the analysis. Phylogram was generated employing the maximum likelihood algorithm built in the MEGA11 software. Phylogram shows that HAB isolate clusters in the CLCuB clade Figure 4.10. Thus, analysis results re-confirmed that HAB isolate belongs to the Cotton leaf curl betasatellite.

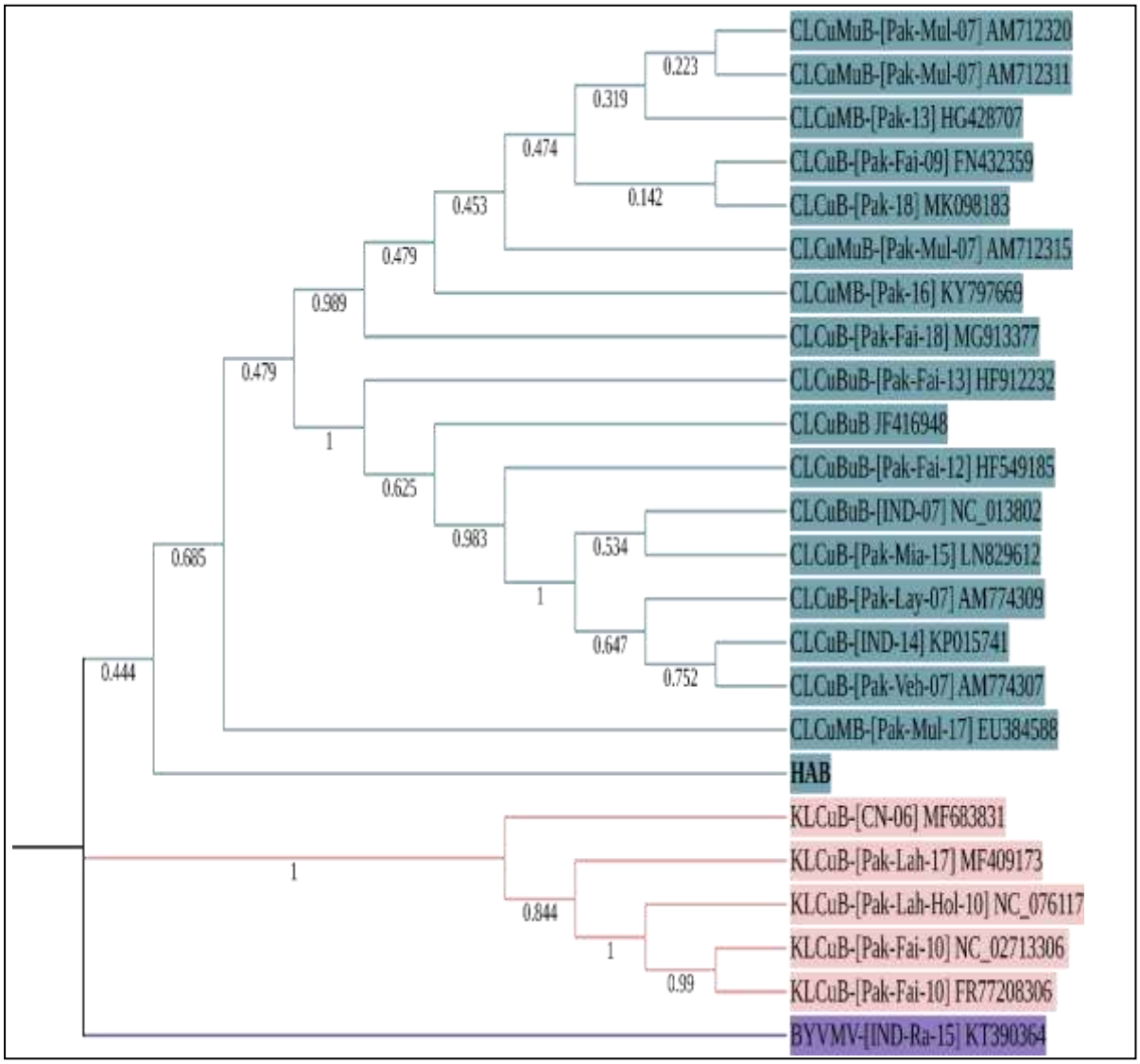

Figure 4.15: Phylogram constructed via Maximum likelihood method presenting **HAB** clustered in CLCuB clade. Each color presents a different clade. Bootstrap value is mentioned at each clade. Other than CLCuB, clade of Kroton leaf curl betasatellite (KLCuB). Bhendi yellow vein mosaic betasatellite (BYVMB) was used as an outgroup and presented in Purple.

4.7.3. Sequence Analysis for Isolate HAH

HAH isolate was sequenced with both forward and reverse (BurxF/BurxR) primers. Upon assembling the contigs for Isolate HAH a total of 1892 bp sequence was revealed thus presenting a partial sequence for DNA-A component. Initially the sequence analysis was performed via the NCBI nBLAST tool. The analysis revealed that the isolate, HAH, shared a maximum of 97% sequence identity with CLCuMuV, thus presenting the variant of CLCuMuV. Further characterization was then performed in NCBI ORF finder tool, which revealed that few ORFs with a typical begomovirus orientation. As the sequence was partial, thus only a few of the genes were characterized from the sequence. Due to the location of primer site in between the ORFs, a portion of few amino acids is lost in between Rep and C4. Gene start and end sites for the partial of HAH are presented in a tabulated form below Table 4.6.

Table 4.6: Nucleotide start and end position of partial genes from HAH Isolate, highest percentage identity with the sequences already present in the database.

Isolate	Isolate length	ORF	Start (nt)	End (nt)	Partial/Complete	Highest percentage identity (%)
HAH	1892 bp	Rep	1278	319	Partial	86% - CLCuMuV-Rep
		TrAP	422	75	Partial	96% - CLCuAlV-TrAP
		C4	1121	951	Partial	53% - CLCuMuV-C4
		REn	277	>2	Partial	98% - CLCuMuV-REn

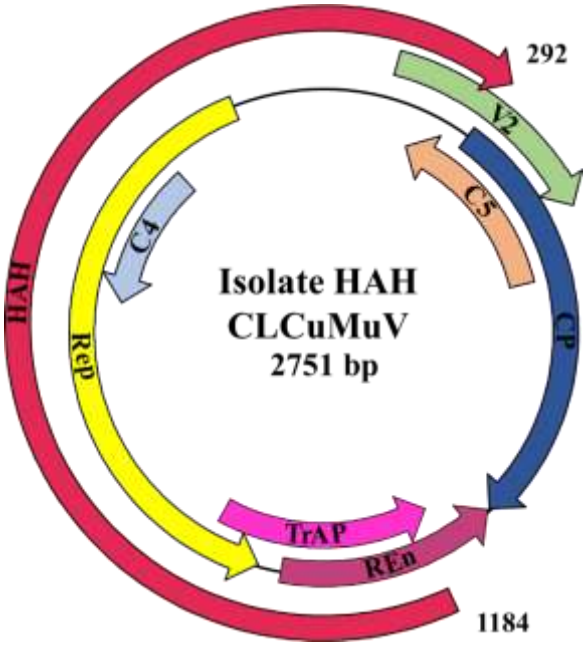

Figure 4.16: Genome map of CLCuMuV presenting the partial sequence for isolate, HAH, expanding from REn to the mid of V2, including nonanucleotide sequence. (Created in pDRAW32).

4.7.4. Sequence Analysis of HAB1 Isolate

A total of 1013 bp sequence was revealed upon the sequencing of HAB1 isolate via the forward primer (Beta01). Initial analysis via then nBLAST tool revealed that the HAB1 shares 96% sequence identity with CLCuMB. Further the sequence was then characterized via the NCBI ORF finder tool. An ORF in the complementary sense was characterized encoding the complete βC1 protein of typical CLCuMB origin. The start and the end sites for the βC1 are mentioned in the tabulated form below. These results revealed Hibiscus being the alternate host for economically important viruses.

Table 4.7: Nucleotide start and end position of partial genes from HAH Isolate, highest percentage identity with the sequences already present in the database.

Isolate	Isolate length	ORF	Start (nt)	End (nt)	Partial/Complete	Highest percentage identity (%)
HAB1	βC1	593	237	357 | 118	Complete	100% with CLCuMB

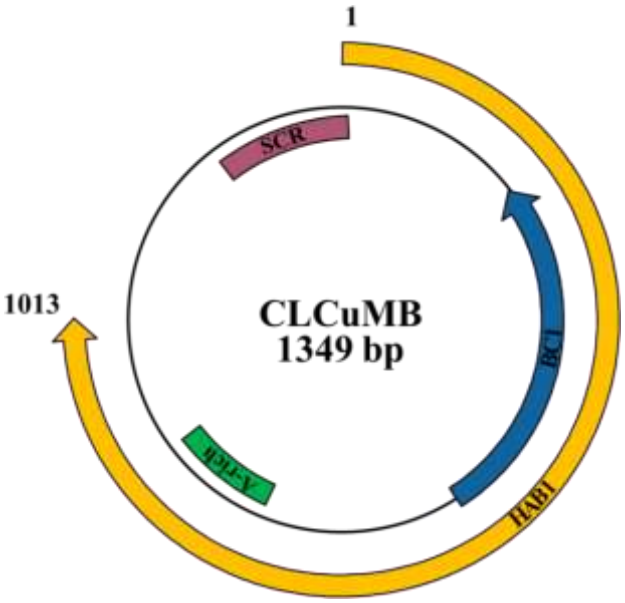

Figure 4.17: Genome map of CLCuMB presenting the partial sequence for isolate, HAB1, covering βC1 and A-rich regions, typical for CLCuMB. (Created in pDRAW32).

4.7.5. Sequence Analysis of DA Isolate

DA isolate was sequenced partially, up to 1608 nucleotides. The sequence was analyzed via the nBLAST tool, which revealed 95% sequence identity with *Chili leaf curl India virus* (ChiLCINV). The sequence was further characterized via the NCBI ORF finder tool, for the hunt of potential coding regions. This analysis revealed a complete TrAP and REn ORFs of ChiLCINV origin, while a partial sequences for Rep and C4 that are also of ChiLCINV origin. This represents Duranta being the alternate host for begomoviruses. Start and end site for each ORF, for DA isolate are tabulated below.

Table 4.8: Nucleotide start and end position of partial and complete genes from DA Isolate, highest percentage identity with the sequences already submitted in the database.

Isolate	Isolate length	ORF	Start (nt)	End (nt)	Partial/Complete	Highest percentage identity (%)
DA	1608 nucleotides	TrAP	859	452	Complete	97.78% ChiLCINV-TrAP
		REn	711	307	Complete	98.5% ChiLCINV-REn
		Rep	1773	1600	Partial	57% ChiLCINV-Rep
		C4	1643	1386	Partial	81% ChiLCINV-C4

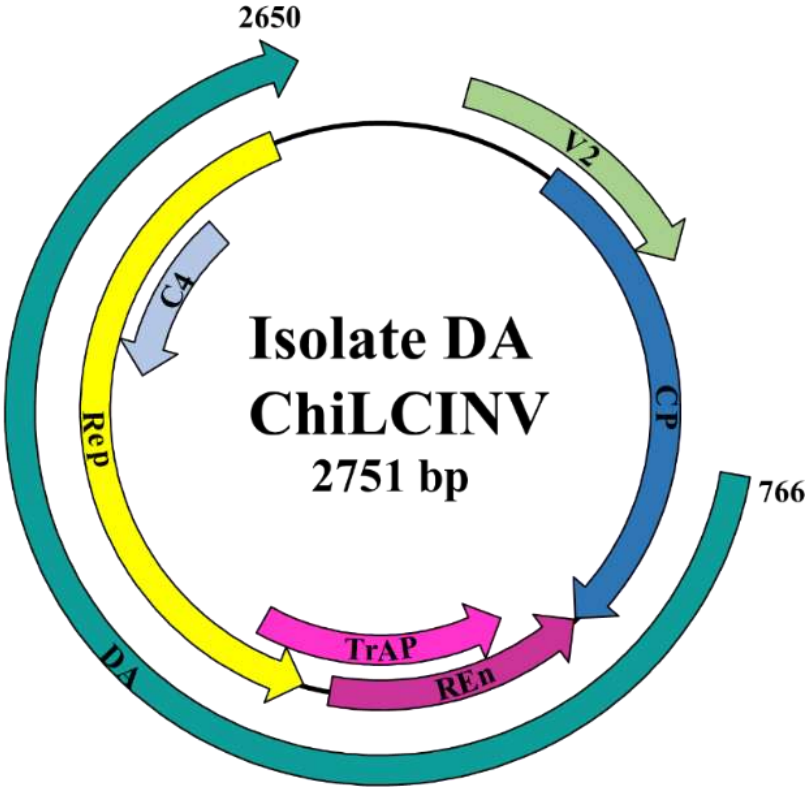

Figure 4.18: Genome map of ChiLCINV presenting the partial sequence for isolate, DA, covering 1608 nucleotides, from mid of CP to the region just before nonanucleotide, of a typical ChiLCINV origin. (Created in pDRAW32).

4.7.6. Sequence Analysis of DAX Isolate

DAX was amplified from duranta, using a different set of primers from that utilized for DA. Initially it was thought that both viruses could be of the same origin. Upon assembling the sequence contigs, a total of 1251 bp sequence was revealed. The sequence was analyzed in the NCBI nBLAST tool, which shows highest percentage identity of 94% with Rose leaf curl *virus* (RoLCuV), and *Papaya leaf crumple virus* (PaLCrV), which is not of Pakistan origin. Thus, it became clear that the *D. repens* has a mix infection of two distinct begomoviruses. The DAX isolate was then further characterized via the NCBI ORF finder tool. This analysis revealed a complete TrAP, while a partial REn and Rep ORFs. The protein coding regions were further analyzed via the pBLAST tool. Each ORF gave the highest sequence identity with three different begomoviruses. These results are indicative of recombination which is common among the viruses which belong to the genus *begomovirus*. The nucleotide starts and end region of each ORF characterized are mentioned in the table below.

Table 4.9: Nucleotide start and end position of partial and complete genes from DAX Isolate, highest percentage identity with the sequences already submitted in the database.

Isolate	Isolate length	ORF	Start (nt)	End (nt)	Partial/Complete	Highest percentage identity (%)
DAX	1251 nucleotides	REn	489	217	Partial	98% ChLCV-REn 97% RoLCV-Ren 93% PaLCrV-REn
		TrAP	637	230	Complete	100% ChLCV-TrAP 98% RoLCV-TrAP 94% PaLCrV-TrAP
		Rep	1097	561	Partial	99% RoLCV-Rep 98 % ChLCV-Rep 97% PaLCrV-Rep

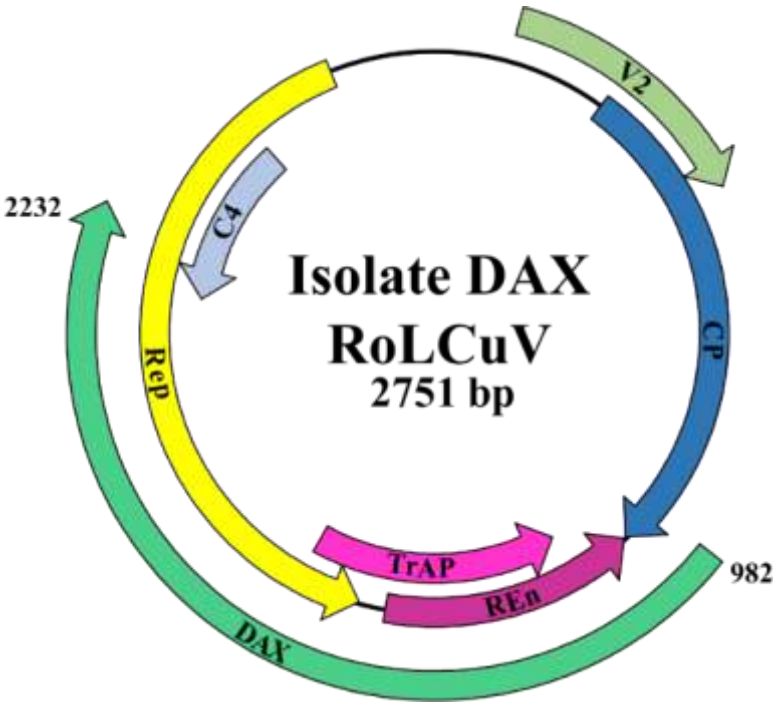

Figure 4.19: Genome map of PaLCrV presenting the partial sequence for isolate, DAX, covering 1251 nucleotides, from end of CP to the region to the mid of Rep, of a typical ChiLCINV origin. (Created in pDRAW32).

4.7.7. Sequence Analysis of DB Isolate

DB isolate was amplified from Duranta via the universal primer set beta01/02. Upon sequencing only 358 nucleotides were revealed, which showed the maximum nucleotide sequence identity of 91% with CLCuMB. However, this result is not enough to claim the association of CLCuMB with two, other distinct begomoviruses, ChiLCINV, and PaLCrV. Moreover, as only 358 nucleotides were sequenced so the βC1 ORF cannot be characterized, as its start codon relies somewhere at of 550 bp.

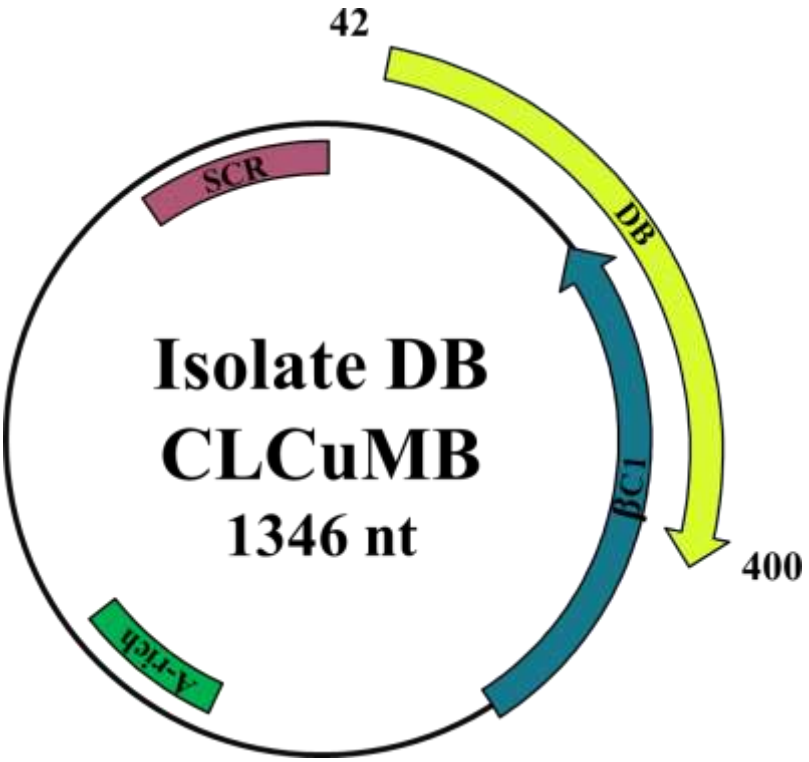

Figure 4.20: Genome map of CLcuMB presenting the partial sequence for isolate, DB, covering 358 nucleotides. (Created in pDRAW32).

4.7.8. Sequence Analysis of JA isolate

JA isolate was sequenced up to 1164 nucleotides. The nBLAST analysis showed the highest percentage similarity of 98% for JA isolate with ChiLCINV. The sequence was then further characterized in the NCBI ORF finder tool. Genome characterization revealed a complete ORFs coding TrAP and REn, while a partial Rep ORF of typical ChiLCINV origin. Nucleotide start and end site for each ORF mentioned in tabulated form below **Table**. This result shows the widened host range for ChiLCINV. Moreover, this is the first report of ChiLCINV from Jasmine, a common ornamental plant.

Table 4.10: Nucleotide start and end position of partial and complete genes from JA Isolate, highest percentage identity with the sequences already submitted in the database.

Isolate	Isolate length	ORF	Start (nt)	End (nt)	Partial/Complete	Highest percentage identity (%)
JA	1164 nucleotide	REn	447	25	Complete	97% ChiLCINV-REn
		TrAP	595	188	Complete	100% ChiLCINV-TrAP
		Rep	746	525	Partial	100% ChiLCINV-Rep

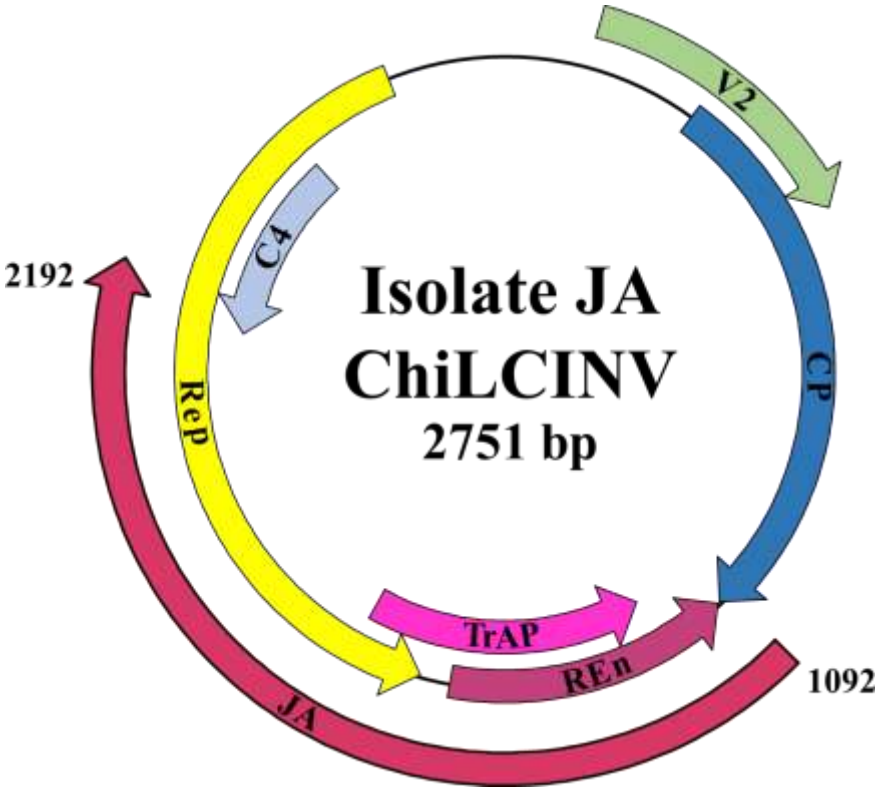

Figure 4.21: Genome map of ChiLCINV presenting the partial sequence for isolate, JA, covering 1164 nucleotides, just from TrAP to the mid of Rep, of a typical ChiLCINV origin. (Created in pDRAW32).

Table 4.11: Summary table for all isolates from Hibiscus, Duranta, and Jasmine, along with the accession numbers of best matched sequences in the database.

Isolate	Sequence length (nucleotides)	Partial/Complete	Highest percentage similarity	Accession No. of best match in database (NCBI)
HAHN	2751	Complete	98% with CLCuMuV	MK482365
HAB	1346	Complete	96% with CLCuMB	FN432359
HAHS	1048	Partial	97% CLCuMuV	MK482365
HAH	1892	Partial	97% CLCuMuV	AJ002458
HAB1	1013	Partial	96% with CLCuMB	JQ317604
DA	1608	Partial	95% with ChiLCINV	MW759565
DAX	1251	Partial	94% RoLCuV & 94% PaLCrV	MN746285
DB	358	Partial	91% CLCuMB	KY432575
JA	1164	Partial	98% ChiLCINV	MW759565

CHAPTER 5: DISCUSSION

Begomovirus is the largest genus of plant viruses that differs from other plant viruses due to its rapid evolution rate and enlarged host range. Several economically important viruses belong to the genus *begomovirus* such as CLCuD causing begomoviruses, ToLCNDV, CMD causing begomoviruses, ChiLCINV and many others. Owing to the economic importance of begomoviruses, measuring their diversity and exploring the host range is very important. The present study was conducted with an aim to explore the diversity of begomoviruses in three different plant species via PCR based diagnostic technique. The study was designed to detect the potential occurrence of begomoviruses followed by their phylogenetic analysis. Tracing the phylogenetic relations are important for epidemiological studies. Except for a few reports, there is currently no solid evidence for seed transmission of begomoviruses. These viruses then potentially reside in the alternative host plants, which act as source of inoculum during crop season. Alternative hosts are usually the weeds or the ornamental plants, which act as a safe house for viruses even under unfavorable conditions. Current trends in begomovirus research are toward epidemiological studies with a great stress over tracing the travel history of virus and defining the host range. The results of this study reveal a great diversity of begomoviruses in the area around Islamabad.

Hibiscus rosa sinensis, commonly known as China rose, was observed to be showing the typical symptoms of begomoviruses. Two different Hibiscus samples, both collected from the different locations inside the National University of Sciences & Technology (NUST), were examined in this study, for the potential occurrence of begomoviruses. A begomovirus betasatellite complex (HAHN and HAB) was identified from Hib-1 while another begomovirus betasatellite complex (HAH and HAB1) was identified from Hib-2. DNA-A showed highest percentage similarity with Taiwan isolate of CLCuMuV, while betasatellite showed highest percentage identity with Pakistani isolate of CLCuMB. Further, for DNA-A, a phylogenetic analysis was performed employing all full-length sequences of CLCuMuV available in the database till June 2023. Analysis revealed that the HAHN isolate clusters with the Faisalabad strain of CLCuMuV. SDT analysis also reconfirmed that the HAHN isolate is a variant of CLCuMuV-Faisalabad strain, which is the most abundant and diverse strain of CLCuMuV.

Phylogenetic analysis of all 225 sequences showed that similar strains clustered together regardless of their geographical origin. It has been observed from the full-length sequence phylogenetic analysis, that the host range pattern is strain specific. Faisalabad strain is

frequently reported from hibiscus, as reconfirmed from this study as well, while there is not even a single report of other strains such as Pakistan and Rajasthan from Hibiscus. CLCuMuV and its betasatellite complex belongs to the CLCuD causing begomoviruses. CLCuMuV has been reported to be infecting the Hibiscus from China (Arif *et al.,* 2018) and India (Srivastava *et al.,* 2016). This is the first report of CLCuMuV infecting *Hibiscus rosa sinensis* from Pakistan and thus re-confirm the potential of *Hibiscus rosa sinensis* to act as an alternative host of CLCuD causing begomoviruses. Although the CLCuMuV is not dominating in the region, its presence in the form of chunks is alarming. This could potentially lead to a new recombination event as happened back in 2001 (Mansoor *et al.,* 2003).

Upon the genome characterization of isolate, HAHN, ORF C5 was also revealed. C5 is among the recently reported proteins from genus *begomovirus* and established to have a potential role as host silencing suppressor (Li *et al.,* 2021). This shows that the virus has multiple proteins to carry out a similar function. These facts reinforce that for resistance development against begomoviruses via CRISPR-Cas9 or RNAi techniques, it is imperative to employ different proteins of a virus in the form of single construct.

Isolates (DA, DAX, and DB) from *Duranta repens* indicates the incidence of **mixed infection** as well as **recombination**. Three different amplicons were sequenced. All of them showed similarity with different species of begomoviruses. DA showed similarity with ChiLCINV, DAX showed highest percentage similarity with RoLCV and PaLCV. The most surprising sequence was DB, which showed highest percentage similarity with CLCuMuB. DAX isolate showed similarity with different species of begomoviruses including RoLCV, PaLCV, and ChLCV, which is a clear indication of recombination. However, recombination analysis cannot be performed, as the isolate is not sequenced completely. *D. repens* has previously reported to host diverse begomoviruses (Anwar & Tahir., 2018; Tahir *et al.,* 2006; Mustujab *et al.,* 2015). This indicates the potential of a common ornamental plant to act as a reservoir host, thus a continuous source of inoculum for begomoviruses.

Jasminum officinale is also a common ornamental plant which can be seen extensively. Sample from *J. officinale* was processed and gave positive results with two different primers indicating the presence of bipartite *begomovirus*. However, only the DNA-A amplicon (JA) was partially sequenced which showed the highest percentage similarity with ChiLCINV. According to our best knowledge, this is the first report of ChiLCIN infecting *J. officinale*, and thus added to the existing host range of the virus.

Results of this study added to the current diversity of begomoviruses and raise the question, **what contributes to the extraordinary success of begomoviruses?** There are several justifications for that. Begomoviruses have a broadened host range, which is expanding even more. The polyphagous nature of the vector species could be the possible reason for such a broad host range. Moreover, the massive availability of vector species in the environment also contributes to the existing diversity of begomoviruses. Adapting to the new host or environment could potentially force the virus to undergo mutations. Such point mutation owing to selection pressure leads to the emergence of new strains and variants. As mentioned earlier, begomoviruses have a few hidden ORFs with a potential role as silencing suppressor and pathogenicity enhancer (Li *et al.,* 2021) (Gong *et al.,* 2022). Such an increasing number of proteins is a possible reason for the exceptional success of begomoviruses.

Apart from these facts, the role of begomovirus associated satellites cannot be ignored. Single encoded gene of betasatellite, βC1, is essential for symptom induction. Few recent studies have shown the occurrence of new ORFs with a potential role in flattening the way for helper virus inside the host (Gui *et al.,* 2022). Betasatellite thus is decisive weapon of begomoviruses which help them survive and cause infectivity in the host plant. Another remarkable fact in the case of plant viruses, which has been investigated in this study, is alternative host. Usually, wild plants such as weeds, and ornamental plants act as an alternative host for begomoviruses. As re-confirmed in this study, *Hibiscus rosa sinensis* acts as an alternative host for CLCuMuV. These alternative plants act as a site of recombination and consistent source of inoculum for other plants.

CONCLUSION

Results of the present study clearly confirmed the diversity of begomoviruses in the area around Islamabad. All plant samples that were processed in this study are common ornamental plants that can be seen widely in our surroundings. The study showed an incidence of mixed infection and reinforced the fact that alternative host plants could act as site of recombination for begomoviruses. The presence of economically important viruses in the common ornamental plant, *Hibiscus rosa sinensis* is elusive. This study thus highlights the need for a country wide survey to evaluate other ornamental plants and immediate removal of these plants around the agricultural sites. Moreover, it added to the previous knowledge of begomoviruses diversity, and paves a new prospect for development of resistance against the largest genus of plant viruses.

REFERENCES

Akram, A., Khan, A. H., Rasool, G., Mansoor, S., Moffett, P., Briddon, R. W., & Saeed, M. (2020). Codiaeum variegatum in Pakistan harbours pedilanthus leaf curl virus and papaya leaf curl virus as well as a newly identified betasatellite. *Archives of virology*, *165*(8), 1877–1881.

AlHudaib, K. A., Almaghasla, M. I., El-Ganainy, S. M., Arshad, M., Drou, N., & Sattar, M. N. (2022). High-Throughput Sequencing Identified Distinct Bipartite and Monopartite Begomovirus Variants Associated with DNA-Satellites from Tomato and Muskmelon Plants in Saudi Arabia. *Plants*, *12*(1), 6.

Anwar, S., & Tahir, M. (2018). Identification of a new begomovirus infecting Duranta repens in Pakistan. *Archives of virology*, *163*(3), 809-813.

Arguello-Astorga, G., Lopez-Ochoa, L., Kong, L. J., Orozco, B. M., Settlage, S. B., & Hanley-Bowdoin, L. (2004). A novel motif in geminivirus replication proteins interacts with the plant retinoblastoma-related protein. *Journal of Virology*, *78*(9), 4817-4826.

Arif, M., Lin, W., Lin, L., Islam, W., Jie, Z., He, Z., ... & Wu, Z. (2018). Cotton leaf curl Multan virus infecting Hibiscus sabdariffa in China. *Canadian Journal of Plant Pathology*, *40*(1), 128-131.

Atallah, S. S., Gómez, M. I., Fuchs, M. F., & Martinson, T. E. (2012). Economic impact of grapevine leafroll disease on Vitis vinifera cv. Cabernet franc in Finger Lakes vineyards of New York. *American Journal of Enology and Viticulture*, *63*(1), 73-79.

Bandaranayake, W. M. E. K., Wickramarachchi, W. A. R. T., Wickramasinghe, H. A. M., Rajapakshe, R. G. A. S., & Dissanayake, D. M. K. K. (2014). Molecular detection and characterization of begomoviruses associated with cucurbitaceae vegetables in Sri Lanka. *Journal of the National Science Foundation of Sri Lanka*, *42*(3).

Barman, M., Samanta, S., Chakraborty, S., Samanta, A., & Tarafdar, J. (2022). Copy number variation of two begomovirus acquired and inoculated by different cryptic species of whitefly, Bemisia tabaci in Okra. *PloS one*, *17*(3), e0265991.

References

Boulton, M. I., Pallaghy, C. K., Chatani, M., MacFarlane, S., & Davies, J. W. (1993). Replication of maize streak virus mutants in maize protoplasts: evidence for a movement protein. *Virology*, *192*(1), 85-93.

Briddon, R. W. (2003). Cotton leaf curl disease, a multicomponent begomovirus complex. *Molecular Plant Pathology*, *4*(6), 427-434.

Briddon, R. W., & Stanley, J. (2006). Subviral agents associated with plant single-stranded DNA viruses. *Virology*, *344*(1), 198-210.

Briddon, R. W., Bull, S. E., Amin, I., Idris, A. M., Mansoor, S., Bedford, I. D., ... & Markham, P. G. (2003). Diversity of DNA β, a satellite molecule associated with some monopartite begomoviruses. *Virology*, *312*(1), 106-121.

Briddon, R. W., Bull, S. E., Mansoor, S., Amin, I., & Markham, P. G. (2002). Universal primers for the PCR-mediated amplification of DNA β: a molecule associated with some monopartite begomoviruses. *Molecular biotechnology*, *20*, 315-318.

Briddon, R. W., Mansoor, S., Bedford, I. D., Pinner, M. S., Saunders, K., Stanley, J., ... & Markham, P. G. (2001). Identification of DNA components required for induction of cotton leaf curl disease. *Virology*, *285*(2), 234-243.

Briddon, R. W., Pinner, M. S., Stanley, J., & Markham, P. G. (1990). Geminivirus coat protein gene replacement alters insect specificity. *Virology*, *177*(1), 85-94.

Brown, J. K., Zerbini, F. M., Navas-Castillo, J., Moriones, E., Ramos-Sobrinho, R., Silva, J. C., ... & Varsani, A. (2015). Revision of Begomovirus taxonomy based on pairwise sequence comparisons.

Castillo, A. G., Collinet, D., Deret, S., Kashoggi, A., & Bejarano, E. R. (2003). Dual interaction of plant PCNA with geminivirus replication accessory protein (Ren) and viral replication protein (Rep). *Virology*, *312*(2), 381-394.

Chowda-Reddy, R. V., Achenjang, F., Felton, C., Etarock, M. T., Anangfac, M. T., Nugent, P., & Fondong, V. N. (2008). Role of a geminivirus AV2 protein putative protein kinase C motif on subcellular localization and pathogenicity. *Virus research*, *135*(1), 115-124.

References

Chowda-Reddy, R.V., Dong, W., Felton, C., Ryman, D., Ballard, K. and Fondong, V.N. (2009) Characterization of the cassava geminivirus transcription activation protein putative nuclear localization signal. Virus Res. 145, 270–278.

Crespo-Bellido, A., Hoyer, J. S., Dubey, D., Jeannot, R. B., & Duffy, S. (2021). Interspecies recombination has driven the macroevolution of cassava mosaic begomoviruses. *Journal of Virology*, *95*(17), 10-1128.

De Barro, P. J., Liu, S. S., Boykin, L. M., & Dinsdale, A. B. (2011). Bemisia tabaci: a statement of species status. *Annual review of entomology*, *56*, 1-19.

Dhobale, K. V., Murugan, B., Deb, R., Kumar, S., & Sahoo, L. (2023). Molecular Epidemiology of Begomoviruses Infecting Mungbean from Yellow Mosaic Disease Hotspot Regions of India. *Applied Biochemistry and Biotechnology*, 1-22.

Egelkrout, E. M., Robertson, D., & Hanley-Bowdoin, L. (2001). Proliferating cell nuclear antigen transcription is repressed through an E2F consensus element and activated by geminivirus infection in mature leaves. *The Plant Cell*, *13*(6), 1437-1452.

Etessami, P., Saunders, K., Watts, J., & Stanley, J. (1991). Mutational analysis of complementary-sense genes of African cassava mosaic virus DNA A. *Journal of General Virology*, *72*(5), 1005-1012.

Fiallo-Olivé, E., & Navas-Castillo, J. (2023). Begomoviruses: what is the secret (s) of their success?. *Trends in Plant Science*.

Fiallo-Olivé, E., Lett, J. M., Martin, D. P., Roumagnac, P., Varsani, A., Zerbini, F. M., ... & Consortium, I. R. (2021). ICTV virus taxonomy profile: Geminiviridae 2021. *The Journal of General Virology*, *102*(12).

Fiallo-Olivé, E., Trenado, H. P., Louro, D., & Navas-Castillo, J. (2019). Recurrent speciation of a tomato yellow leaf curl geminivirus in Portugal by recombination. *Scientific reports*, *9*(1), 1332.

Fiallo-Olivé, E., Zerbini, F. M., & Navas-Castillo, J. (2015). Complete nucleotide sequences of two new begomoviruses infecting the wild malvaceous plant Melochia sp. in Brazil. *Archives of virology*, *160*, 3161-3164.

Fraser, R. S. S. (1990). The genetics of resistance to plant viruses. *Annual review of phytopathology*, *28*(1), 179-200.

García-Rodríguez, D. A., Partida-Palacios, B. L., Regla-Márquez, C. F., Centeno-Leija, S., Serrano-Posada, H., Bañuelos-Hernández, B., & Cárdenas-Conejo, Y. (2023). Sida chlorotic leaf virus: a new recombinant begomovirus found in non-cultivated plants and Cucumis sativus L. *PeerJ*, *11*, e15047.

Garrido-Ramirez, E. R., Sudarshana, M. R., Lucas, W. J., & Gilbertson, R. L. (2000). Bean dwarf mosaic virus BV1 protein is a determinant of the hypersensitive response and avirulence in Phaseolus vulgaris. Molecular plant-microbe interactions, 13(11), 1184-1194.

Gergerich, R. C., & Dolja, V. V. (2011). Introduction to plant viruses, the invisible foe. *Plant Health Instr. Available online: https://www. apsnet. org/edcenter/disandpath/viral/introduction/Pages/PlantViruses. aspx (accessed on 29 May 2020)*.

Gong, P., Zhao, S., Liu, H., Chang, Z., Li, F., & Zhou, X. (2022). Tomato yellow leaf curl virus V3 protein traffics along microfilaments to plasmodesmata to promote virus cell-to-cell movement. *Science China Life Sciences*, *65*(5), 1046-1049.

Gu, Q., Yan, L., Liu, L., Bao, W., Fang, H., Xu, J., ... & Peng, B. (2023). First report of tomato leaf curl New Delhi virus infecting several cucurbit plants in China. *Plant Disease*, (ja).

Gui, X., Liu, C., Qi, Y., & Zhou, X. (2022). Geminiviruses employ host DNA glycosylases to subvert DNA methylation-mediated defense. *Nature Communications*, *13*(1), 575.

Gutzat, R., Borghi, L., & Gruissem, W. (2012). Emerging roles of RETINOBLASTOMA-RELATED proteins in evolution and plant development. *Trends in plant science*, *17*(3), 139-148.

Haley, A., Zhan, X., Richardson, K., Head, K., & Morris, B. (1992). Regulation of the activities of African cassava mosaic virus promoters by the AC1, AC2, and AC3 gene products. *Virology*, *188*(2), 905-909.

Hampton, R. O. (1975). The nature of bean yield reduction by bean yellow and bean common mosaic virus. *Phytopathology*, *65*(12), 1342-1346.

Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D., & Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Reviews Microbiology*, *11*(11), 777-788.

HONG, Y., SAUNDERS, K., HARTLEY, M. R., & STANLEY, J. (1996). Resistance to geminivirus infection by virus-induced expression of dianthin in transgenic plants. *Virology*, *220*(1), 119-127.

Ismayil, A., Haxim, Y., Wang, Y., Li, H., Qian, L., Han, T., ... & Liu, Y. (2018). Cotton Leaf Curl Multan virus C4 protein suppresses both transcriptional and post-transcriptional gene silencing by interacting with SAM synthetase. *PLoS pathogens*, *14*(8), e1007282.

Jain, R. K., Pappu, H. R., Pappu, S. S., Krishna Reddy, M., & Vani, A. (1998). Watermelon bud necrosis tospovirus is a distinct virus species belonging to serogroup IV. *Archives of virology*, *143*, 1637-1644.

Kaliappan, K., Choudhury, N. R., Suyal, G., & Mukherjee, S. K. (2012). A novel role for RAD54: this host protein modulates geminiviral DNA replication. *The FASEB Journal*, *26*(3), 1142-1160.

Kalloo, G. (1997). *Leaf curl virus of tomato and chili in India* (No. RESEARCH). AVRDC.

Kanakala, S., Verma, H. N., Vijay, P., Saxena, D. R., & Malathi, V. G. (2013). Response of chickpea genotypes to Agrobacterium-mediated delivery of Chickpea chlorotic dwarf virus (CpCDV) genome and identification of resistance source. *Applied Microbiology and Biotechnology*, *97*, 9491-9501.

Kannaiyan, J., Nene, Y. L., Reddy, M. V., Ryan, J. G., & Raju, T. N. (1984). Prevalence of pigeonpea diseases and associated crop losses in Asia, Africa, and the Americas. *International Journal of Pest Management*, *30*(1), 62-72.

Kegler, H., & Hartmann, W. (1998). Present status of controlling conventional strains of plum pox virus. *Plant virus disease control*, 616-628.

Khan, A., Imtiaz, Y., Tahir, M., & Briddon, R. W. (2021). Complete genome sequence of hollyhock vein yellowing virus, a novel monopartite begomovirus infecting hollyhock in Pakistan. *Archives of Virology*, *166*, 2607-2610.

Krishnareddy, M., Jalali, S., & Samuel, D. K. (2003). Fruit distortion mosaic disease of okra in India. *Plant disease*, *87*(11), 1395-1395.

Kwak, H. R., Hong, S. B., Byun, H. S., Park, B., Choi, H. S., Myint, S. S., & Kyaw, M. M. (2022). Incidence and Molecular Identification of Begomoviruses Infecting Tomato and Pepper in Myanmar. *Plants*, *11*(8), 1031.

Li, F., Xu, X., Huang, C., Gu, Z., Cao, L., Hu, T., ... & Zhou, X. (2015). The AC 5 protein encoded by Mungbean yellow mosaic India virus is a pathogenicity determinant that suppresses RNA silencing- based antiviral defenses. *New Phytologist*, *208*(2), 555-569.

Li, P., Su, F., Meng, Q., Yu, H., Wu, G., Li, M., & Qing, L. (2021). The C5 protein encoded by Ageratum leaf curl Sichuan virus is a virulence factor and contributes to the virus infection. *Molecular Plant Pathology*, *22*(9), 1149-1158.

Lozano, G., Trenado, H. P., Fiallo-Olivé, E., Chirinos, D., Geraud-Pouey, F., Briddon, R. W., & Navas-Castillo, J. (2016). Characterization of non-coding DNA satellites associated with sweepoviruses (genus Begomovirus, Geminiviridae)–definition of a distinct class of begomovirus-associated satellites. *Frontiers in Microbiology*, *7*, 162.

Luque, A., Sanz-Burgos, A. P., Ramirez-Parra, E., Castellano, M. M., & Gutierrez, C. (2002). Interaction of geminivirus Rep protein with replication factor C and its potential role during geminivirus DNA replication. *Virology*, *302*(1), 83-94.

Mandal, B., Jain, R. K., Chaudhary, V., & Varma, A. (2003). First report of natural infection of Luffa acutangula by Watermelon bud necrosis virus in India. *Plant Disease*, *87*(5), 598-598.

Mansoor, S., Amin, I., Iram, S., Hussain, M., Zafar, Y., Malik, K. A., & Briddon, R. W. (2003). Breakdown of resistance in cotton-to-cotton leaf curl disease in Pakistan. *Plant pathology*, *52*(6), 784-784.

Marwal, A., Sahu, A., Prajapat, R., & Gaur, R. K. (2013). First report of Begomovirus infecting two ornamental plants: Jasminum sambac and Millingtonia hortensis. *Indian Phytopathol*, *66*(1), 115-116.

Matthews, R. E. F., & Hull, R. (2002). *Matthews' plant virology*. Gulf professional publishing.

McKirdy, S. J., Jones, R. A. C., & Nutter Jr, F. W. (2002). Quantification of yield losses caused by barley yellow dwarf virus in wheat and oats. *Plant Disease*, *86*(7), 769-773.

Mei, Y., Wang, Y., Hu, T., He, Z., & Zhou, X. (2021). The C4 protein encoded by Tomato leaf curl Yunnan virus interferes with mitogen-activated protein kinase cascade-related defense responses through inhibiting the dissociation of the ERECTA/BKI1 complex. *The New Phytologist*, *231*(2), 747-762.

Moriones, E., & Navas-Castillo, J. (2000). Tomato yellow leaf curl virus, an emerging virus complex causing epidemics worldwide. *Virus research*, *71*(1-2), 123-134.

Mukasa, S. B., Rubaihayo, P. R., & Valkonen, J. P. T. (2006). Interactions between a crinivirus, an ipomovirus and a potyvirus in coinfected sweetpotato plants. *Plant pathology*, *55*(3), 458-467.

Mustujab, A., Briddon, R. W., & Tahir, M. (2015). First report of Catharanthus yellow mosaic virus infecting Duranta repens in Pakistan. *Journal of Plant Pathology*, *97*(1).

Nawaz-ul-Rehman, M. S., & Fauquet, C. M. (2009). Evolution of geminiviruses and their satellites. *FEBS letters*, *583*(12), 1825-1832.

Nawaz-ul-Rehman, M. S., Nahid, N., Mansoor, S., Briddon, R. W., & Fauquet, C. M. (2010). Post-transcriptional gene silencing suppressor activity of two non-pathogenic alphasatellites associated with a begomovirus. *Virology*, *405*(2), 300-308.

Nene, Y. L. (1972). A survey of viral diseases of pulse crops in Uttar Pradesh. *A survey of viral diseases of pulse crops in Uttar Pradesh.*

Nene, Y. L. (1972). A survey of viral diseases of pulse crops in Uttar Pradesh. *A survey of viral diseases of pulse crops in Uttar Pradesh.*

Nolla, J. A. B., & Valiela, M. V. F. (1976). Contributions to the history of plant pathology in South America, Central America, and Mexico. *Annual Review of Phytopathology*, *14*(1), 11-29.

Noueiry, A. O., Lucas, W. J., & Gilbertson, R. L. (1994). Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. *Cell*, *76*(5), 925-932.

Pogue, G. P., Lindbo, J. A., Garger, S. J., & Fitzmaurice, W. P. (2002). Making an ally from an enemy: plant virology and the new agriculture. *Annual review of Phytopathology*, *40*(1), 45-74.

Pun, K. B. (1999). Effect of age of okra plants on susceptibility to Okra yellow vein mosaic virus. *Indian J Virol*, *15*, 57-58.

Qureshi, M. A., & Afghan, S. (2005). Sugarcane cultivation in Pakistan. *Sugar Book Pub. Pakistan Society of Sugar Technologist*.

Rahman, S. U., Raza, G., Zubair, M., Ahmed, N., Domier, L. L., Jamil, N., ... & Amin, I. (2023). Multiple begomoviruses infecting soybean; a case study in Faisalabad, Pakistan. *Biologia*, *78*(2), 609-620.

Rao, G. P., & Reddy, M. G. (2020). Overview of yield losses due to plant viruses. In *Applied plant virology* (pp. 531-562). Academic Press.

Saha, A. K., Nath, P. S., & De, B. K. (2005). Assessment of yield loss in chilli due to chilli leaf curl virus (CLCV) disease. *RESEARCH ON CROPS*, *6*(2), 259.

Sanchez-Chavez, S., Regla-Marquez, C. F., Cardenas-Conejo, Z. E., Garcia-Rodriguez, D. A., Centeno-Leija, S., Serrano-Posada, H., ... & Cardenas-Conejo, Y. (2020). First report of begomoviruses infecting Cucumis sativus L. in North America and identification of a proposed new begomovirus species. *PeerJ*, *8*, e9245.

Selvarajan, R., & Balasubramanian, V. (2013). Natural occurrence of banana bunchy top virus in Ensete superbum in India. *Indian Journal of Virology*, *24*, 97-98.

Sharma, S. R., & Varma, A. (1981). Assessment of losses caused by cowpea banding mosaic (CpBMV) and cowpea chlorotic spot (CpCSV) virus in cowpea. In *Proceedings of the IIIrd international symposium on plant pathology (IPS), New Delhi, p 95 (Abstr)*.

Simmonds, P., Aiewsakun, P., & Katzourakis, A. (2019). Prisoners of war—host adaptation and its constraints on virus evolution. *Nature Reviews Microbiology*, *17*(5), 321-328.

Singh, D. K., Islam, M. N., Choudhury, N. R., Karjee, S., & Mukherjee, S. K. (2007). The 32 kDa subunit of replication protein A (RPA) participates in the DNA replication of Mung bean yellow mosaic India virus (MYMIV) by interacting with the viral Rep protein. *Nucleic acids research*, *35*(3), 755-770.

Srivastava, A., Kumar, S., Jaidi, M., & Raj, S. K. (2016). Association of Cotton leaf curl Multan virus and its associated betasatellite with leaf curl disease of Hibiscus rosa-sinensis in India. *New Disease Reports*, *33*(4), 2044-0588.

Sung, Y.K. and Coutts, R.H. (1995) Mutational analysis of potato yellow mosaic geminivirus. J. Gen. Virol. 76, 1773–1780.

Sunter, G., Hartitz, M. D., Hormuzdi, S. G., Brough, C. L., & 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*(1), 69-77.

Suyal, G., Mukherjee, S. K., Srivastava, P. S., & Choudhury, N. R. (2013). Arabidopsis thaliana MCM2 plays role (s) in mungbean yellow mosaic India virus (MYMIV) DNA replication. *Archives of virology*, *158*, 981-992.

Tahir, M., Haider, M. S., Shah, A. H., Rashid, N., & Saleem, F. (2006). First report of bipartite begomovirus associated with leaf curl disease of Duranta repens in Pakistan. *Journal of Plant Pathology*, *88*(3), 339-342.

Teng, P. S., Torres, C. Q., Nuque, F. L., & Calvero, S. B. (1990). Current knowledge on crop losses in tropical rice. *Crop loss assessment in rice*, 39-53.

Thresh, J. M., & Cooter, R. J. (2005). Strategies for controlling cassava mosaic virus disease in Africa. *Plant pathology*, *54*(5), 587-614.

Thresh, J. M., & Cooter, R. J. (2005). Strategies for controlling cassava mosaic virus disease in Africa. *Plant pathology*, *54*(5), 587-614.

Wartig, L., Kheyr-Pour, A., Noris, E., De Kouchkovsky, F., Jouanneau, F., Gronenborn, B., & Jupin, I. (1997). Genetic analysis of the monopartite tomato yellow leaf

curl geminivirus: roles of V1, V2, and C2 ORFs in viral pathogenesis. *Virology*, *228*(2), 132-140.

Watson, M., & Serjeant, E. P. (1964). The effect of motley dwarf virus on yield of carrots and its transmission in the field by Cavariella aegopodiae Scop. *Annals of Applied Biology*, *53*(1), 77-93.

Zerbini, F. M., Briddon, R. W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., ... & Consortium, I. R. (2017). ICTV virus taxonomy profile: Geminiviridae. *The Journal of general virology*, *98*(2), 131.

Zhan, B., Zhao, W., Li, S., Yang, X., & Zhou, X. (2018). Functional scanning of apple geminivirus proteins as symptom determinants and suppressors of posttranscriptional gene silencing. *Viruses*, *10*(9), 488.

Appendices

Appendix I

Sequence of HAHN Isolate:

>HAHNACCGGATGGCCGCGCGATTTTTTTGTGGGCCCCCGATTTATGAGAGTGCTCCCTGAAAGCTAAAT
AACGCTCCCGCC
CACTATAAGTACTTGCGCACTAAGTTTCAAATTCAAACATGTGGGATCCATTGTTAAACGAATTTCCTGA
TACGGTGCACGGG
TTTCGGTGTATGCTTTCTGTCAAATATTTGCAACTTTTATCGCAGGATTATTCACCGGATACCCTAGGTTAC
GATTTAATACGGG
ATTTAATCTGTATTCTACGTTCCCGTAATTATGTCGAAGCGAGCTGCAGATATCGTCATTTCTACGCCCGC
GTCGAAAGTACGT
CGGCGTCTGAACTTCGGCAGCCCATACACCAACCGTGTTGTTGCCCCCATTGTCCGCGTCACAAAACAA
CAGGCATGGACA
AACAGGCCTATGAACAGGAAGCCCAGAATGTATCGGATGTACAGAAGTCCGGATGTTCCAAAGGGTTGT
GAAGGCCCATGT
AAGGTACAGTCTTTTGAGTCCAGACATGATGTTGTTCATATTGGTAAGGTAATGTGTATTTCTGATGTTAC
TCGTGGTGTCGGT
TTGACCCATCGTATTGGTAAACGTTTTTGTGTCAAGTCAGTTTATGTTTTAGGTAaGATATGGATGGATGA
AAATATAAAGACC
AGGAATCACACGAATTCGGTCATGTTCTTTTTAGTTCGCGATCGACGACCTGTTGACAAACCTCAGGATT
TTGGTGAGGTATT
CAATATGTTTGACAACGAACCCAGTACAGCAACTGTGAAGAATAGTCATAGGGACCGTTATCAGGTGTT
GAGGAAATGGCAT
GCAACCGTTACGGGTGGTCAATATGCGAGTAAGGAACAGGCTTTGGTCAAGAAGTTTGTCAGAGTGAAC
AATTATGTTGTTT
ACAATCAACAGGAAGCAGGAAAATACGAGAATCACACCGAGAATGCGTTAATGCTTTATATGGCTTGTAC
CCATGCTAGTAAC
CCAGTTTATGCTACGCTGAAGATTAGGATATATTTTTATGACTCTGTAACGAGTTGATATTAATAAAGTTTG
AATTTTATTTCTGA
ATATTGTTCTACATACATAGTTTGTTGGATTACATTGTACAATACATGTTCTACAGCTTTAATAACTAAATT
AATTGAAATTACACC
GAGATTGTTCAGATATTTGAGGACTTGGGTTTTGAATACCCTTAAGAAAAGACCAGTCTGAGGGTGTAAG
GTCGTCCAGATT
CGGAAGGTTAGAAAACACTTGTGCAGTCCCAGAGCTTTCCGAGTGTTGTAGTTGAAATGGATCCTGATCG
TTAGTATGTCCA
TGTTCGTCGTGAATGGACGGTTGTCGTGGCTGAGGATCTTGAAATAGAGGGGATTTGGAACCTCCCAGAT
ATATGCGCCATT
CCCTGCTTGAGCTGCAGTGATGGGTTCCCCTGTGCGTGAATCCATGGTTGTGGCAGTTGATTGACAGATA
ATAAGAACACCC
GCATTCaAGATCTACTCTCCTCCTCCTGATGCGCCTCTTCGCTTCCCTGTGCTGTACTTTGATTGGTACCTG
AGTACAGGGGTC
CTTCAAGTGTGATGAAGATCGCATTCTTTACTGCCCAGTTCTTTAGTGCGGTGTTCTTTTCCTCGTCTAGG
AATTCTTTATAACT
GCTGTTGGGACCAGGATTGCAGAGGAAGATTGTTGGTATCCCGCCTTTAATTTGAACTGGCTTCCCGTAC
TTTGTGTTTGATT
GCCAGTCCCTTTGGGCCCCCATGAACTCTTTAAAGTGTTTGAGGAAATGCGGGTCGACGTCATCAATGA
CGTTATACCAGGC
GTCGTTACTGTAGACCTTGGGACTCAGGTCCAGATGTCCACATAAATAGTTATGTGGTCCCAGTGATCTA
GCCCACATCGTCT
TCCCCGTTCGACTATCTCCCTCAATTACTATACTACGAGGTCTAAGGGGCCGCGCAGCGGCATCGACAAC
GTTATCGATGGCC
CAAACTTCAAGTTCTTCTGGAACTTGATCGAAAGAAGAAGGAGAAAAAGGAGAAATATAGGGAGCCGG
TGGCTCCTGAAA
GATTCTGTCTAGATTTGCATTTAAATTATGAAATTGTAGTACAAAATCTTTAGGAGCTAGTTCCTTAATGA
CTCTAAGAGCCTCT

GACTTACTGCCTGCGTTAAGTGCTGCGGCGTAAGCGTCGTTGGCTGTCTGCTGTCCTCCTCTTGCTGATC
TTCCATCGATCTG
AAACTCTCCCCACTCGAGAGTGTCTCCGTCCTTGTCGATGTAGGCCTTGACGTCTGAGCTTGATTTAGCT
CCCTGAATGTTCG
GATGGAAATGTGCTGATCTGGTTGGGGATACCAGGTCGAAGAATCTGTTATTCGTGCAGACGAATTTGCC
CTCGAACTGGAT
GAGCACATGGAGATGAGGGCTCCCATCTTCGTGTAACTCTCTGCAGATTTTGATGTATTTTTTATTCGAGG
GTGTGTTGATGG
CTTGAATTTGGGAAAGTGCTTCCTCTTTAGTGAGTGAGCACTGTGGATAAGTGAGGAAATAATTTTTGGA
TTGTACTTTAAAA
CGTTTGGGGGGAGCCATTGACTTTGGTCAATTGGAGACAAGTGAGGGGATATTTTACTCTGGCAATTGG
AGACAGGAGAC
AATATATAGTGTCTCCAAATGGCATATTCTGTAAATAACTCGAAGTTCGTTTGAAATTCAAATTCCCCTTT
GGGCTCCAAAAGC GGCCATCCGTATAATATT

Appendix II

Sequence of HAB Isolate: >HAB

ACCGTGGGCGAGCGGTGTCCGATGGTTTCTTGGTGGGTCCCACTGCTGGTATTGACTTGATTTGACTTATATTGGGCCAATTT
AATGGGTTGAAAATGTTTGGGCCTTTGGAAGAGGGtCTTTTATGGAATAGGGCTCTGTATATTGTTGGTGATATGTGTGTTAAA
TATGCATTGCTGGTTTGTGTTTGGAATTTAAAcggTGAACTTTTtATTGAATacGTACGGTTcgTTTACATCCATTCCCAATATCTCT
GGGTTTTCAAGTACAAGTATATCAAGTCTGTGAACTATATCTTCTATCTCGATCTCTTCTATCTTTGCCCCGTTGTATGCGAATAG
GAAATTCGCTATGATGTTCCCTTCAAAGCCGTTGAAGTCGAATGGAACGTGAATGTCTTCGTACGTGTACTGGACGATCCCTTC
ATATTTGATGAGTGATGGTGACTTGGTGGATACTAGCTTCATGTGAATGAATATCTTCATGTTCGTCATGATGCGAACGTCGACT
GTGAACCTGACTCCCTGCGTGTTTGTTCTGCTCCTCGTCATTTCTGCTTATTTCATGGAATTGTTTATTATGCATACATTTATAGAC
TTAATGAATAGATATTTTGTGTAGTTGTGTGGTTGTGAGTGATTTAATATTATGTGATTGTCCATTAAAGGGATAAAGTGATGATG
GAGATGTATTACACGTGTTGTGATGTTGGCTTGAAAGCTTTATACATGGGTTTGTACCGTATATACATGCATATACGGATAGAAA

ACGGATAAGAAAAGGAAAAACGGAACTGAAGAGGAAAGACAAAGAAAAGAATCAAGGATATATTATTTATGAAAGAAATGG
GAGCGCAGCGAATCGAAACAGGAAAACCCAAGGAAAGAGAAAAAATAAAAAGTAAAGAGAAAAAAAaTAtAAATTCGAAA
ACGTCATCGTTTGAGAGTAATAAAAAAAAGAAAAAACAAAAACATATCCGAAAACGTTGTCGTTTGAAGGTGACTGTGTGGT
TTTACCATTTACTGTGTGGTAAATGGTAAGTGATTGAAGAATAGTTAAAAAATGGAGACACCGATAGGTAAATTGTCTCCAATA
TATTGGAGAACAATTGGGGACTCATAAATTGCCTTTCCTAAAATACCCCCGCTTTTGTGTCTAAGAGGCGCGTCGGAGTGCGC
CAATAAAGTTAACATTCTCTCTCCTCTTTTGATCTCCAATACAATTTCCCGGTGATCGGAGTCGAAttttccgacacgcgcggcggtgtgt
acccctgggagggtaggtaccactacgctacgctaGGCTTAGCTACGCCGAGCTTAGCTCGCCCACGTTCTAATATT

Appendix III

Sequence of DA Isolate:

GGTTTCTAACAATGTTGTAAATGAGCCCACTCGGGGGATTGTGTAAGAAGTTTCTTCGTGATAGGGATCAAGTTATGCGCAAA
TGGAATACAACTGTGACAGACGGTCTGTATGCATCGAAGGAGCAGGCTCTCGTGAAGAAGTTTATTAGGGTAATAAATATGTT
GTGTACAATCAGCAAGAGGCTGGCAAGTATGAGAATCATACTGAGAATGCATTGATGTTGTATATGGCGTGTACCCACGCCTCT
AACCCCGTGTATGCTACATTGAAGATACGGATCTACTTCTATGATTCAGTATCGAATTAATAAATATTAAATTTTATTGAATATGAT
TGGTTTACATATACAACATGGTGTAATACATTCCATAATACATGATCAACTGCCCTAATTACATTGTTAATACTGATAACTCCTAAC
ATATTCAAATACTTAATCACTTGGGTCTTAAAGACCCTTAAGAAACGACCAGTCGGAGGCTGTGAAGTCATCCAAATTCGGAA
GGCTAGGAAACATTTGTGAATCCCCAACGCTTTCCTCAGGTTGTGGTTGAACTGTACTTGGACGGTTATGATGTCTTTGTTCAT
CAGGAATGGCCGGTTGTGGTGCTTTATGATCTTGAAATAGAGGGGATTTGGAACCTCCCAGGTATACACGCCATTCATCGCCT
GAGCTGCAGTGATGGGTTCCCCTGTGGGTGAATCCATGGTTGTGGCAATTGATGTGTACGTAGTATGAGCAGCCACAGTTTAG
ATCAACCCTCTTACGCCGGATGGCTCTACGTTTAGCAGATTTGTGTTGGACCTTGATTGGAACCTGAGTAGAGTGGGCCTTGG
AGGGAGATGAAGGTTGCATTTTTAAGTGCCCAGGGATTCAATGCGCTATTATTTTCATCGTCCAAAAACTCAGCAGAGTAGGA
ATTGGGGCCTGGCTAGCCGAGGAAAAAAAAGTGGGGACTGCCAGTCCCTTTGGGCCCCCATGAATTCTTTAAAGTGTAGTGG
GGAGCTAAGTCTCGTATGATGTTGTACCATGCCTCAGTATTCGTATCACCTTTGGACTAAGGTCTAGATGCCCACACAGATAATT
GTGTGGACCCAGTGACCTAGCCCACATCGTCTTGCCCGTTCCACTGTCCCCCTCTATGATTCTCGGAAGCCCACTCCTCAAGTT
CGTCCGGAACTTGATCAAAAGACAAGAAAAAAAAGGAGAAACATAAACCTCCTTTGGAGGAGTAAAACTCCTATCTAAATTG
GATTTTAAATTATGGTATTGAAAAATAAAATCTTTTGGGAGTTTTTCCCTAATTATTGCCAGAGCGGCTTCAGCTGAACCTGCAT
TTACGGCCTCTGCTGCAGCATCATTAGCTGTCTGTTGACCTCCTCGAGCAGATCGTCCATCGACCTGAAAATGACCCCACTCGA
TGTAATCACCGTCCTTCTCGATGTAGGACTTGACATCGGATGAGGACTTTGCTCCCTGGAAGTTTGGATGGAATTGGGTGGAT
GTATTAGGGTGAGTGACTCGTAATGTCTGGGGTTTCTGAACTGGGATTTACGCCTCGAATTGAATGAGGGCATGGACATGACT
AGACCCATCTTGGGTGCATTTCCTGTGAACCTCTGATGAATAATTTATCAGAAAGGCTAGATATGTTTTTTCGAGTTCGAGAAC
TTGCTCTTTGGGAATAGAGCATTTTGGATAAGTAAGGAAGACATTCTTGCGCATTAACTTGGGAATTGATGACGCCGAGGCAA
ATTAATTGGCTCCCTCTCCCTAAATGCTGAGAATTGGG

Appendix IV

Sequence of DAX Isolate:

ACAGTCAGCCAATGATGCTTACGCCCAAGCAATTAACACAGGCAGTAAGTCAGAGGCTCTTAGAGTCATTAGGGAACTAGCA
CCAAAGGATTATGTCTTACAATTTCATAATCTTAATGCTAATCTAGATAGGATTTTTGCACCTCCAATGGAGGTTTATGTTTCCCC
TTTTCTTTCTTCTTCCTTCGATCAAGTTCCAGAGGCCATAGAGGAATGGGCCTCTGATAATGTGATGGGTCCCGCTGCGCGGCC
ATTGAGACCTAAAAGTATCGTCATTGAGGGTGATAGTCGTACGGGGAAGACAATGTGGGCTAGGTCACTGGGTCCACATAATT
ACCTATGTGGCCATTTAGATCTGAGCCCTAAGATCTATTCAAATGATGCATGGTACAACGTCATTGATGACGTAGACCCCCACTA
CCTAAAGCACTTTAAAGAATTCATGGGGGCCCAGAGGGACTGGCAAAGTAACACCAAGTACGGGAAGCCAGTTCAAATTAA
AGGGGGCATTCCCACTATCTTCCTATGCAATCCTGGGCCCAATTCCAGCTATAAAGAGTTCCTCGATGAGGAAAAGAACTCAG
CACTGAAGTCTTGGGCTATACACAATGCAACATTCATCACCCTCACAGAGCCACTGTACTCAGGTGCCCATCAAGGTCCAACA
CAGGCTAGCCAAGCGTAGAGCCATCCGGCGTAAGAGGGTTGACCTCGAGTGTGGCTGCTCATACTACGTACACATCAACTGC
CACAACCATGGATTCACGCACAGGGGAATCCATCACTGCAGCTCAGGCAATGAATGGCGTGTATATCTGGGAAGTCCCAAATC
CCCTGTATTTCAAGATCCTCAACCACGACAAAAGGCCACTCAACATGAACCACGACATCATCCACATCAGGGTCCAGTTCAATC
ACAACCTCAGGAAGGAGCTGGGGATAATGAAGTGTTTCTTGGGCTTGAAGATCCACACTCGTTTACATCCTCAGACTGGGCA
TTTCTTAAGAGTCTTTAAGACTCAGTTTTAAGATATTTAGATAGTTTAGGCGTAATAGTATAAATAATTGGATTAGGGCATTTAAC
ATGTATTATATGATGTACTGGAGGGCCAATTGATGTAATGAGGAACATGAATAAAATTTAATTTTATTATTCGAAACTGAAACAA
AAAAAAAAATCTCTTTCTTTCCGGGGGGGCAACACCAGGGGTGAGAGGGGGGGGGAACACCCCAAAAAAAAATCAAGGAT
TTTTCCAAATAAATTTCAATTGCCCCCCTTTGGGGGGTGGAAAAAAAAAAAAATTTAAAACCCAAAAAAACCTTCCCCGAAAC

CCGCCCCTCTAGAGACAAAGGGCCCCGGTGGGTTTTTTCCTTTTTTAAAATAGAGAAAAAAAAAAAAATTTCTTCGCGGGGGGGGGGGTTTTAAAACTAA

Appendix V

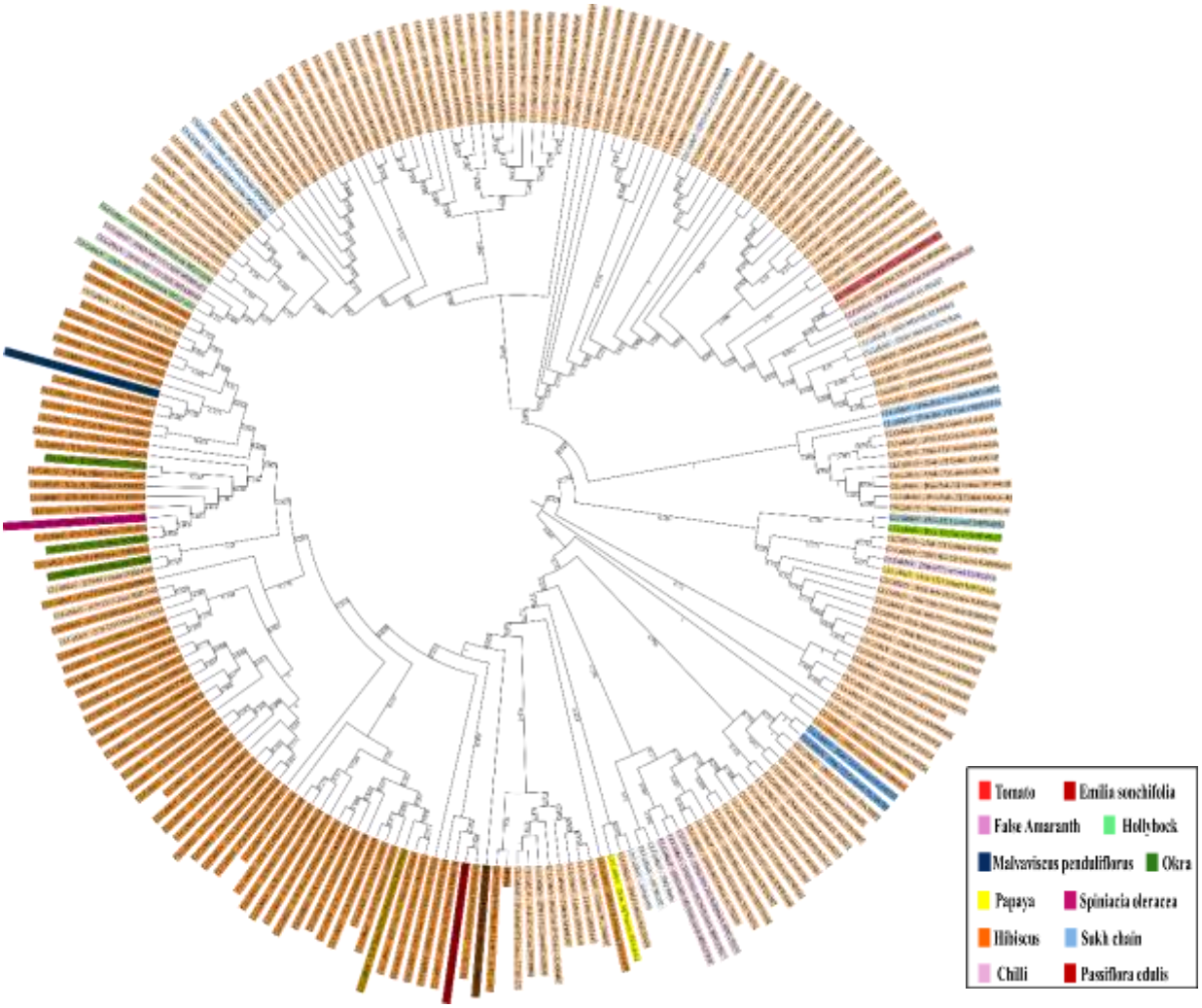

Figure: Cladogram showing the host range of CLCuMuV

turnitin

Prof. Dr. Muhammad Tahir
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Linta Jadoon
Assignment title:	Slot 01
Submission title:	MS-Thesis (Salliha Khalid)
File name:	For_Plagiarism_1.docx
File size:	181.54K
Page count:	42
Word count:	12,359
Character count:	69,047
Submission date:	05-Aug-2023 04:23AM (UTC-0400)
Submission ID:	2138493904

Detection, Amplification and Sequence Analysis of Begomovirus(es) from different Plant Species

By:

Salliha Khalid

Registration # 00000364461

A thesis submitted in partial fulfilment of the requirement for the degree of Master of Science

in

Plant Biotechnology

Under the Supervision of
Prof. Dr. Muhammad Tahir
Professor, Department of Plant Biotechnology

Atta-Ur-Rahman School of Applied Biosciences
National University of Sciences & Technology
Islamabad, Pakistan
2023

MS-Thesis (Salliha)

Prof. Dr. Muhammad Tahir
Deptt of Plant Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

ORIGINALITY REPORT

14%	10%	11%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	2%
2	prr.hec.gov.pk Internet Source	1%
3	Submitted to Higher Education Commission Pakistan Student Paper	1%
4	dokumen.pub Internet Source	1%
5	link.springer.com Internet Source	1%
6	Plant Virus and Viroid Diseases in the Tropics, 2013. Publication	<1%
7	The Whitefly Bemisia tabaci (Homoptera Aleyrodidae) Interaction with Geminivirus-Infected Host Plants, 2011. Publication	<1%
8	Kiran Vilas Dhobale, Bharatheeswaran Murugan, Rishav Deb, Sanjeev Kumar,	<1%