GENOME ORGANIZATION AND INFECTIOUS CLONE CONSTRUCTION OF A GEMINIVIRUS CAUSING YELLOW VEIN MOSAIC IN OKRA

THESIS

By

MORE PRASHANT DIGAMBAR
(A-2008-30-03)

Submitted to

CHAUDHARY SARWAN KUMAR
HIMACHAL PRADESH KRISHI VISHVAVIDYALAYA
PALAMPUR – 176 062 (H.P.) INDIA

in

partial fulfilment of the requirements for the degree

of

MASTER OF SCIENCE IN AGRICULTURE
(DEPARTMENT OF AGRICULTURAL BIOTECHNOLOGY)
(AGRICULTURAL BIOTECHNOLOGY)

2011

UNIVERSITY LIBRARY
C.S.K.H.P. KRISHI VISHVAVIDYALAYA
PALAMPUR

Acc. No. 105188
Class No. 635.6488 Book No. D44G
Auther Digambar, more Prashant
Title Genome organization and infectious clone vein mosaic okra.

Borrower's No.	Due Date	Borrower's No	Due Date

CSK HP KRISHI VISHVAVIDYALAYA
LIBRARY, PALAMPUR (H.P.)
Accession No... 105188......
Date 15.7.11.. Price..........
Source The Dean, PGS, CSKHPKV, Palampur
Bill No .... ——— Date ——.......
Accessioned by Checked by

635.6488
D44G

23/8/11

Is there anything I can say,
anything I can give
or do for you…..
Because all that I'm
all that I have
I owe to you…..

Affectionately Dedicated
to my
Revered Parents

Who have always sacrificed
their present
to make my future better

Dr Vipin Hallan
Scientist

Plant Virology Lab, Division of Floriculture,
Council of Scientific and Industrial Research
(CSIR), Institute of Himalayan Bioresource and
Technology (IHBT), Palampur.

CERTIFICATE – I

This is to certify that the thesis entitled **"Genome organization and infectious clone construction of a geminivirus causing yellow vein mosaic in okra"** submitted in partial fulfilment of the requirements for the award of the degree of **Master of Science (Agriculture)** in the discipline of **Agricultural Biotechnology** of CSK Himachal Pradesh Krishi Vishvavidyalaya, Palampur is a bonafide research work carried out by **More Prashant Digambar (A-2008-30-03)** son of Shri More Digambar Tukaram under my supervision and that no part of this thesis has been submitted for any other degree or diploma.

The assistance and help received during the course of this investigation have been duly acknowledged.

Vipin Hallan

(Dr Vipin Hallan)
Major Advisor

Place: Palampur
Dated: May 3, 2011

CERTIFICATE- II

This is to certify that the thesis entitled **"Genome organization and infectious clone construction of a geminivirus causing yellow vein mosaic in okra"** submitted by **More Prashant Digambar (A-2008-30-03)** son of Shri More Digambar Tukaram to the CSK Himachal Pradesh Krishi Vishvavidyalaya, Palampur in partial fulfilment of the requirements for the degree of **Master of Science (Agriculture)** in the discipline of **Agricultural Biotechnology** has been approved by the Advisory Committee after an oral examination of the student in collaboration with an External Examiner.

13-6-2011
(Dr. T.R. Sharma)
Chairperson
Advisory Committee

13/6/2011
( )
External Examiner

(Dr A A Zaidi) 13.6.11
Member

13/6/2011
(Dr Nageswer Singh)
Member

13/6/11
(Dr L S Sudhakar)
Member-*cum*-Dean's nominee

16/6/2011
Head of the Department

16/6
Dean, Postgraduate Studies

ACKNOWLEDGEMENTS

In this highly complex society, no work can be accomplished by a single individual but it needs inspiration and sincere gratitude of intellectuals as well as the grace of that Almighty. With limitless humility, I would like to praise and thank 'GOD', the merciful, the compassionate, who bestowed me with health, tenacity and enough courage to go through this crucial juncture.

I am grateful to "God", for bestowing me with affectionate parents, whose love, dedication & inspiration encouraged me to undergo higher studies. Their sacrifice, heartiest blessings & firm faith have made this documentation a puny remuneration to translate their and my dreams into reality.

With the overwhelming sense of legitimate pride and genuine obligation which gives me exuberant pleasure and privilege to express my eternal gratitude to my learned and revered advisor Dr Vipin Hallan (Scientist EII), Plant Virology Lab, Division of Floriculture, Institute of Himalayan Bioresource and Technology (IHBT), CSIR, Palampur and Chairman of my Advisory Committee for his excellent and praiseworthy guidance, keen interest, adroit admonition, juvenile encouragement and parental affection during whole course of my study and particularly at times of research and preparation of this manuscript. With a conscientious supervision, he has saved me from the taste of several errors by his frank and unsparing criticism. I shall always remain indebted to him. It is my proud privilege to express my gratitude and obligation to my esteemed co-supervisor Dr A. A. Zaidi (Scientist and Head, Floriculture Division, IHBT (CSIR) a virologist of repute. It is more than mere thanks that I acknowledge his guidance. It is his constructive criticism, fruitful discussion and critical analysis that have helped me to get this thesis in proper shape.

It is my sole prerogative to place on record my indebtedness and everlasting gratitude to intelligent and professional dexterity of the members of my Advisory Committee, Dr Nageswer Singh (Assistant Scientist) and Dr L. S. Sudhakar (Professor) for their prompt anticipation, wise counsel and ever helping attitude steered the completion of this work.

It is my sole prerogative to place on record my indebtedness and everlasting gratitude to Mr. Yogesh, who introduced me to the basics of Molecular Biology techniques and I enjoyed working under his guidance.

I owe my special thanks to Dr S. R. Thakur, Head of Department of Agricultural Biotechnology and Dr P.S. Ahuja, Director of IHBt (CSIR) for extending all necessary facilities as and when required. I am equally grateful to the esteemed faculty members of the Department of Agricultural Biotechnology, Late Dr P. Plaha, Dr R.K. Kapila, and Dr K.D. Sharma for their cordial co-operation extended towards the course of this study.

I highly acknowledge Dr. Raja Ram, for his suggestions and kind cooperation during the period of investigation and my seniors Tanuja, Salik, Sunny, Rahul, Surinder, Puja, Yashika, Anuradha, Renu, Vijay Laxmi, Smita, Prachi and Yashika for their immense help during the course of my M.Sc. work in field evaluation and day-to-day lab work.

To me research is not one man's endeavor. At each stage of work, my seniors especially Dr Rohit, Dr Vishal, Dr Naresh, Dr Prabhakar, Anil, Rajesh, Jintu, Vivek, Munish, Lokender, Murali, Arjun, Divakar and my friends Shrvan, Ashwani, Vishal, Sunil, Manoj, Varun, Gaurav and Ajay who boosted me up in periods of mental stress & strain, & made this place a home away from home. It would be inappropriate if I commit a mistake not to mention name of my juniors Subhash, Prakash, Suresh, Mamta, Harish, Sankhyan, Sumit, Ramesh, Pushpender, Saurav, Kuldeep, Praveen, Tushar, Omkar, Ajit and Rohit who kept me in exalted state during moments of despondency.

Words can't express my profound thanks to my parents Sh. Digambar More and Smt. Asha More, my sisters Pallavi, Tejaswi and all my family members, who has shared the joys and sorrows of my life and who has been continuous source of inspiration even in moment of utter melancholy.

In particular I am thankful to the Department of Biotechnology (DBT), Govt. of India, for providing the financial assistance in the shape of fellowship during the course of study.

Moreover, the services rendered by Mr. Jagadeep and Mr. Ranjeet are deeply acknowledged who provided me the best support and service during stress hours.

Mr. Ajay Walia deserves special thanks for his untiring efforts in bringing the manuscript to this form. Acknowledgements are inherently endless & incomplete, and I request indulgence from many friendly & helpful people whom I could not name here, due to paucity of space.

Lastly, it is worth mentioning the beautiful town Palampur and Shivalik (PG) Hostel for some unforgettable experiences and memories.

Needless to say, all omissions and errors are mine.

(More Prashant Digambar)

Place : Palampur
Dated : May 3, 2011

TABLE OF CONTENTS

LIST OF ABBREVIATIONS USED

Abbreviation	Meaning
µg	microgram
µl	microlitre
bp	base pair
CTAB	hexadecyl-trimethyl-ammonium bromide
DNA	deoxyribonucleic acid
DMSO	dimethyl sulfoxide
EDTA	ethylenediamine tetra acetic acid
ELISA	enzyme linked immunosorbent assay
g	gram
g	gravity
h	hour
IPTG	isopropyl-ß-D-thiogalactopyranoside
ISEM	immunosorbent electron microscopy
kb	kilobase
L	litre
LB	Luria-Bertani broth
LA	Luria-Bertani agar
M	molar
MAb	monoclonal antibody
mg	milligram
min	minute(s)
mL	milliliter
m*M*	millimolar
NaCl	sodium chloride
NCBI	National centre for biotechnology information
ng	nanogram
nt	nucleotide
°C	degree Celsius
PAb	polyclonal antibody
PCR	polymerase chain reaction
pH	puissance de hydrogen (ion conc.)
ppm	parts per million
PVP	polyvinypyrrolidone
RNA	ribonucleic acid
RFLP	Restriction fragment length polymorphism
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
sec	Seconds
SSC	saline sodium citrate
TAE	tris-acetate-EDTA
Taq	*Thermus aquaticus*
TBE	tris/borate/EDTA buffer
TE	tris EDTA buffer
Tris	tris (hydroxy methyl) amino methane
UV	ultraviolet
V	volts
X-Gal	5-bromo-4-chloro-3-indolyl-ß-D-galactoside

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

Department of Agricultural Biotechnology, COA
CSK Himachal Pradesh Krishi Vishvavidyalaya
Palampur – 176 062 (HP)

Title of thesis : Genome organization and infectious clone construction of a geminivirus causing yellow vein mosaic in okra
Name of the student : More Prashant Digambar
Admission number : A-2008-30-03
Major discipline : Agricultural Biotechnology
Minor discipline : Biochemistry
Date of thesis submission : May 3 , 2011
Total pages of the thesis : 131
Major Advisor : Dr Vipin Hallan

ABSTRACT

Okra is one of the important vegetables in India and the yellow vein mosaic disease is a serious problem in okra production. This disease is known to be caused by a geminivirus. Geminiviruses are most damaging viruses in tropical and subtropical regions but now these viruses are spreading in temperate region also. A begomovirus was detected causing disease in okra from sub-temperate region of Himachal Pradesh for very first time. The complete genome of viral DNA was amplified by using rolling circle amplification (RCA) technique. The virus was found to be monopartite but the satellite molecule was also found to be associated with the virus called betasatellite. The viral genomic components were successfully cloned and sequenced. The complete nucleotide sequence of the virus is determined to be 2739 nucleotide long and of betasatellite was determined to be 1332 nucleotide long. The genome organization was shown as that of Old World begomoviruses. The genome sequence showed 96% similarity with *Okra yellow vein mosaic virus* (OYVMV-[PK:Fai201:95]) and found to be a strain of this earlier characterized virus. Betasatellite was found to be highly variable showing less than 78% identity with all other reported betasatellites, which is the species demarcation threshold for a distinct betasatellite, suggesting it to be a distinct begomoviral betasatellite species. The present betasatellite also showed some similarity with DNA-A molecule which might be due to recombination event between DNA-A and betasatellite molecule. The DNA-A was also found to be recombinant molecule generated from OYVMV-[PK:Fai201:95] and CLCuMV[PK:62:05]. Similarly the Betasatellite was found to be recombinant generated from OLCuIB-[IN:Mun:EL41:06] and BYVB-[IN:Abd:OY164:06]. Agroinoculation of infectious clone of DNA-A shows that this DNA is not responsible for typical symptom induction but causes yellowing of leaves with leaf distortion and it is replicating and moving in plant tissue, confirming monopartite nature of the virus.

Vipin Hallan

(More Prashant Digambar)
Student
Date: 03.05.2011

(Dr Vipin Hallan)
Major Advisor
Date: 03.05.2011

Head of the Department

Introduction

1. INTRODUCTION

Okra (*Abelmoschus esculentus* L.) belongs to family Malvaceae and is grown for its tender fruits which are widely used as vegetable. Okra is one of the important vegetables in India and is cultivated in all parts of India. Okra is rich in vitamins, calcium, potassium, and other minerals. India ranks at first position in okra with 3.5 million tonnes (70% of the total world production) of okra produced from over 0.35 million ha land (Anonymous 2008). The area under okra cultivation in Himachal Pradesh and India during the last five years is found to be increasing from year to year. (Table: 1.1, Anonymous 2010).

Table 1.1: Area and production of okra in Himachal Pradesh and India during last five years (A: Area in 000'HA; P: Production in 000'MT)

	2005-06		2006-07		2007-08		2008-09		2009-10	
	A	P	A	P	A	P	A	P	A	P
Himachal Pradesh	1.7	19.7	1.8	21.7	1.8	21.7	2.2	26.2	2.2	26.2
India	390.7	3951.4	396.4	4070.3	406.9	4179.1	431.6	4527.9	435.6	4541.2

Viruses have emerged as major plant pathogens as 47% of all emerging infectious diseases of plants have been reported to be caused by viruses (Anderson *et al.* 2004). Okra is prone to damage by various insects, fungi, nematodes and viruses. Some of the important insects are fruit and shoot borer, aphids, whiteflies, ants, etc. The okra is also subjected to the attack of different viral diseases affecting leaves, flowers and fruits. Yellow vein mosaic, enation and leaf curl are common viral diseases in okra. Yellow vein mosaic disease is one of the major limitations in the production of important vegetable crop okra. This disease is mainly caused by OYVMV (Zhou *et al.* 1998) or

Bhendi yellow vein mosaic virus (BYVMV) (Jose and Usha 2003). The disease is characterized by a homogenous interwoven network of yellow veins enclosing islands of green tissue within. In extreme cases, the infected leaves become totally yellowish or cream colored. The plants infected in early stages remain stunted and bear very few deformed, pale yellow colored and small fruits with very low market value (Ali *et al.* 2005). The disease causes heavy loss in yield if the plants get infected within 20 days after germination.

OYVMV belongs to genus *Begomovirus* in family *Geminiviridae*. Plant viruses are divided into more than 15 families, of which *Geminiviridae* is the second largest family (Yadava *et al.* 2010). Geminiviruses are characterized by twinned icosahedra shaped particles approximately 18 nm x 30 nm in size (Stanley *et al.* 2005), which encapsidate single-stranded circular DNA (sscDNA) genome (Lazarowitz 1992). Geminiviruses constitute one of the economically most important groups of plant viruses and are transmitted mainly by whitefly (*Bemisia tabaci*) and leaf hoppers in persistent manner (Capoor and Varma 1950) but are not sap or seed transmitted (Bock 1982).

The genus *Begomovirus* is the largest and consists of more than 180 species and several unassigned isolates (Fauquet *et al.* 2008). Begomoviruses may have monopartite or bipartite genome. So far, three types of begomoviruses have been identified that are distinguishable by their genome composition. One group consists of bipartite begomoviruses that contain DNA-A and DNA-B components; a second group is consists of monopartite begomoviruses, which lack a component equivalent to DNA-B; and the third group is represented by monopartite begomoviruses that harbor a satellite molecule called betasatellite (Jose and Usha 2003).

Most geminiviruses were mainly reported in tropical and sub-tropical regions of India. Increase in international commodity trade, intercontinental transportation networks and a changing global climate have contributed to the spread of this virus and its whitefly vector (Moffat 1999). The complete genome sequence of the virus and its molecular analysis provided a clearer picture about its diversity and geographical distribution and such a data may further be utilized for developing resistance strategies against that virus. Genome sequence analysis also gives the idea about the intraspecies as well as interspecies recombination events that might have happened during viral evolution. Similarly, infectious clones of begomoviruses are an advanced molecular tool to deduce host range of the virus, as these viruses are usually not transferred mechanically.

The geminiviruses have now been reported in sub-temperate regions of Himachal Pradesh (Kumar *et al.* 2008). Begomovirus symptoms were not prevalent earlier, and the symptoms on various crops were first observed in 2004-2005. Whiteflies are also occurring in large populations in the greenhouses. The yellow vein mosaic disease was also observed. Seedlings are often transported from tropical and sub-tropical regions, which could account for the emergence of the disease in this region and hence its molecular characterization is necessary to control its spread and to develop a clearer picture of the geographical distribution of the virus.

Keeping in view the economic importance of okra and the emergence of the yellow vein mosaic virus in this area, following objectives were defined in the present study:

i. Complete genome organization of the virus causing yellow vein mosaic in okra
ii. Construction of infectious clone of the virus
iii. Elucidation of genetic nature of the virus by analyzing genome sequence by various bio-informatics tools

Review of Literature

2. REVIEW OF LITERATURE

Okra yellow vein mosaic virus (OYVMV) is most serious virus infecting okra, as up to 95.7% yield losses were reported in India under severe infection (Pun and Doraiswamy 1999). The total estimated loss of vegetable by OYVMV menace was around 20-30% but if it is uncontrolled by any means then up to 90% yield loss was reported (Thompson and Kelly 1957). Infection of OYVMV disease drastically reduces yield, quality of the produce and the extent of damage declines with delay in infection (Fajinmi and Fajinmi 2010). The Yield loss due to the virus ranged from 50 to 90 percent depending on the stage of the crop growth at which infection occurs, if the plants infected 50 and 65 days after germination then they suffer a loss of 84 and 49% respectively (Sastry and Singh 1974). The disease is more during summer months when the crop is sown during January to May (Singh 1980). OYVMV in okra was first reported from Mumbai in India (Kulkarni 1924), which implies that it might have originated in India.

Whitefly-transmitted viruses (prior to the discovery and classification of the genus *Begomovirus*) were referred to as “rugaceous” viruses (Bird and Maramorosch 1978). Their novel “geminate” particle morphology was revealed (Goodman 1981) in the 1970s when the first electron micrographs were produced. Begomoviruses contain a single stranded circular genome and were also discovered in late 1970s (Goodman 1977). Presently, only three whitefly species: *Bemisia tabaci, Trialeurodes vaporariorum*, and *T. abutilonia*, are known as vectors of plant viruses. Of the three virus transmitting whiteflies, *B. tabaci* is the most commonly demonstrated vector of over 100 different viral diseases in the tropics and subtropics, belonging mostly to the genus *Begomovirus* (Jones 2003).

The literature pertaining to different aspects of the present investigation has been reviewed under the following headings:

2.1 Geminiviruses
2.2 Genome organization of begomovirus
2.3 Replication and Recombination
2.4 Detection and amplification of viral genome
2.5 Infectious Clone Construction

2.1 GEMINIVIRUSES

Geminiviridae is one of the largest families of plant viruses, containing total 672 isolates (Fauquet *et al.* 2008). The diseases caused by the viruses that constitute the family *Geminiviridae* were recognized in the early 1960's, by Costa (1965) in Brazil. The name "geminivirus", proposed by Harrison *et al.* (1977) and Matthews (1979), is derived from the genimante structure of the viral capsid (Fig. 2.1). Members of this family have circular, single-stranded DNA genomes that are approximately 2.5-3.0 kb in size. Geminiviruses can be either monopartite, if their genome contains only one DNA molecule, or bipartite if it consists of two molecules (Stanley *et al.* 2005). Many economically important virus diseases of crops are caused by geminiviruses (Moffat 1999).

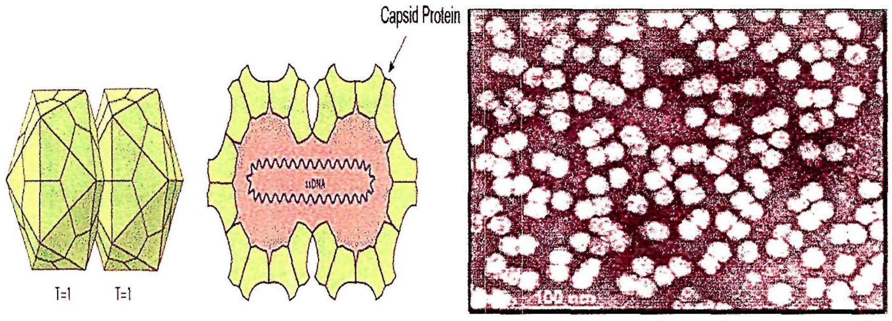
2.1.1 Taxonomy

Based on the insect vector, genome organization, and host range (Padidam *et al.* 1995a), geminiviruses are divided into four genera: *Begomovirus, Mastrevirus, Curtovirus* and *Topocuvirus* (Table 2.1).

Table 2.1. The features of different genera of geminiviruses

Virus genus	Type species	Genome	Vector	Host Plant
Mastreviruses	*Maize streak virus* (MSV)	Monopartite	Leafhopper (Mostly Of genus *Cicadulina*)	Monocotyledonous plants
Curtoviruses	*Beet curly top virus* (BCTV)	Monopartite	Leafhoppers (*Circulifer tenellus*)	Dicotyledonous plants
Begomoviruses	*Bean golden yellow mosaic virus* (BGYMV)	Monopartite or Bipartite	Whiteflies (*Bemisia abaci*)	Dicotyledonous plants
Topocuviruses	*Tomato pseudo-curly top virus* (TPCTV)	Monopartite	Treehopper (*Micrutalis malleifera*)	Dicotyledonous plants

Of these four genera, the genus *Begomovirus* is increasingly important with 188 species (181- assigned, 7-unassigned) (Fauquet *et al.* 2008).



A) B)

Figure 2.1. Morphology of geminiviruses.

A) Cartoon showing the geminate particle morphology of geminiviruses (adapted from http://expasy.org/viralzone/all_by_species/111.html) [10th April 2011]

B) Electron micrograph of *Geminiviridae* (adapted from http://www.ncbi.nlm.nih.gov/Images/em_gemin.jpg) [10th April 2011]

2.1.2 Strain demarcation and nomenclature of geminiviruses

In recent years the number of geminivirus genomic sequences deposited in GenBank have increased significantly which lead to the complexity in geminivirus taxonomy and nomenclature. Taxonomic and nomenclatural updates are published at regular intervals (Fauquet *et al.* 2000; 2003; 2008). In 2003, due to difficulties inherent in species identification, the ICTV *Geminiviridae* Study Group proposed new species demarcation criteria. They proposed 89% nucleotide identity threshold between full-length DNA-A component nucleotide sequences for begomovirus species. In same year, a new revised criterion for demarcation of geminiviruses was published (Fauquet *et al.* 2003).

In 2008, a new update was published (Fauquet *et al.* 2008), proposing demarcation criteria and guidelines to classify and name geminiviruses below the species level. They proposed following guidelines for the classification of geminivirus isolates in variants, strains and species:

i. After pairwise sequence comparison of a new begomovirus isolate with all known sequences representative of species, using ClustalV algorithm;
    - if its nucleotide identity is <88%, it belongs to a new species.
    - if its nucleotide identity =88-89%, it belongs tentatively to the closest species.
    - if its nucleotide identity >89%, it belongs definitively to that species.

ii. After pairwise sequence comparison of a new begomovirus isolate with all known sequences representative of strains and variants in the identified species, using ClustalV algorithm;
    - if its nucleotide identity <93% to all known members, it is a member of a new strain in that species,
    - if its nucleotide identity >94% to an existing isolate, it is a variant of that strain in that species.

2.1.2 Classification and Nomenclature of betasatellites

Since betasatellites were first identified in 2000, over 260 full-length sequences (~1,360 nucleotides) have been deposited in databases, and this number increases daily. This has highlighted the need for a standardized, concise and unambiguous nomenclature for these components, as well as a meaningful and robust classification system. So the recommendations were proposed (Briddon *et al.* 2008) for classification and nomenclature of betasatellites of begomoviruses. According to this criterion, species names take the form "host-disease symptom-[origin]-beta", for example, the betasatellite component associated with *Cotton leaf curl Multan virus* would become *Cotton leaf curl Multan beta* (CLCuMB). For species demarcation sequence identity of

78% after pair-wise comparisons of all available full-length Betasatellite sequences using ClustalV algorithm was proposed as the species demarcation threshold for a distinct Betasatellite.

2.2 GENOME ORGANIZATION OF BEGOMOVIRUSES

Based on phylogenetic studies and genome arrangement, begomoviruses have been divided broadly into two groups: the Old World (OW) viruses (eastern hemisphere, Europe, Africa, Asia) and the New World (NW) viruses (western hemisphere, the Americas) (Padidam *et al.* 1999; Paximadis *et al.* 1999; Rybicki 1994). Begomovirus genomes have a number of characteristics that distinguish Old World and New World viruses. All indigenous New World begomoviruses are bipartite, whereas both bipartite and monopartite begomoviruses are present in the Old World. In addition, DNA-A of bipartite begomoviruses from the New World lacks an AV2 ORF (Rybicki 1994; Stanley *et al.* 2005).The genomes of the bipartite begomoviruses consist of two circular single stranded DNA molecules referred to as DNA-A and DNA-B of nearly 2.6–2.8 kb size (Stanley *et al.* 2005). Haber *et al.* (1981) postulated for the first time that two nucleic acid components were to be carried in each paired particle of geminivirus. The DNA A and DNA B components share little sequence similarity, except for ~170 nucleotide (nt) of sequence in the intergenic region (IR), termed the common region (CR) (Hanley-Bowdoin *et al.* 1999). The CR has an origin of replication (Ori) including the stem-loop structure containing the conserved nonanucleotide TAATATTAC sequence. The Ori posses a tandemly repeated motif (iteron) located at variable distances from the conserved hairpin, which is bound specifically by its cognate Rep protein and functions as a major recognition element of the replication origin in begomoviruses and curtoviruses (Behjatnia *et al.* 1998; Fontes *et al.* 1994a; Fontes *et al.* 1994b). DNA-A typically harbors six open reading frames (ORF): AV1 (known as AR1; coat protein, CP) and AV2 (known as AR2; AV2 protein or movement protein, MP) on the virion-sense strand; AC1 (known as AL1; replication protein, Rep), AC2 (known as AL2; transcriptional activator, TrAP), AC3 (known as AL3; replication enhancer, REn) and AC4 (known as AL4; AC4 protein) on the complementary-sense strand. DNA-B contains two ORFs encoding proteins involved in movement: BV1 (known as BR1; nuclear shuttle protein, NSP) on the virion-sense strand and BC1 (known as BL1; movement protein, MPB) on the complementary-sense strand (Seal *et al.* 2006; Stanley *et al.* 2005). The genome organization of begomoviruses is as shown in Figure 2.2.

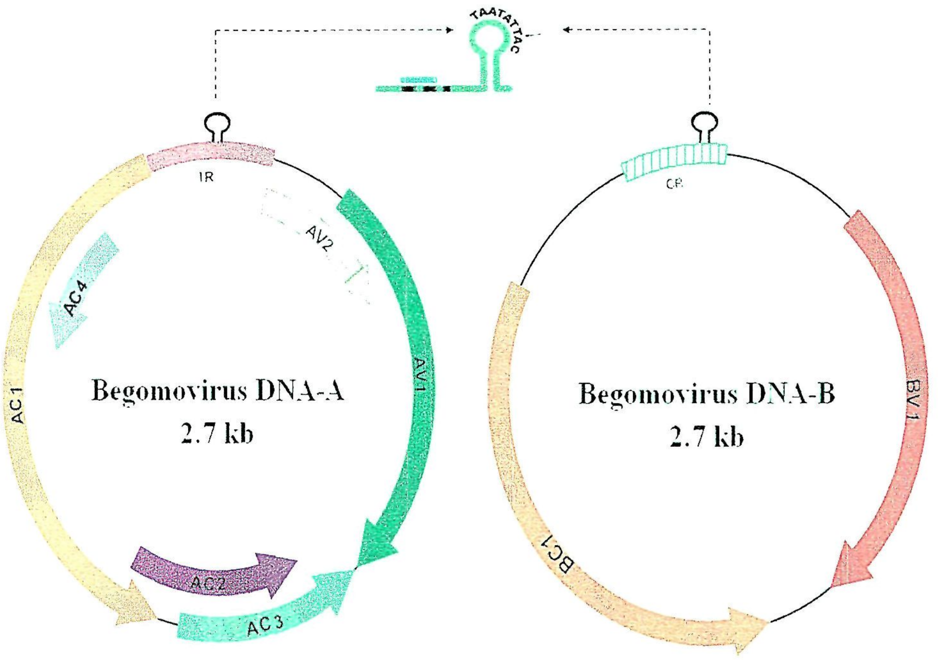


Figure 2.2. Genome organization of *Begomovirus*. The ORFs are denoted according to their orientation as V (virion-sense) or C (complementary-sense). The ORFs of the monopartite begomoviruses should not have the "prefix" A. The diagonal lining of the V2 ORF of DNA-A of the genus *Begomovirus* indicates that this ORF is absent in members from the New World. The common region (CR) of the two components of the genus *Begomovirus* is illustrated in more detail showing stem-loop structure with a solid arrow indicating the nicking position.

2.2.1 Circular ssDNA molecules associated with geminiviruses

The first circular DNA molecule associated with a geminivirus was *Tomato leaf curl virus* (ToLCV) satellite DNA isolated from ToLCV-infected tomato in Australia (Dry *et al.* 1997). This satellite DNA comprised 682 nts and contained two stem-loop structures (I and II). Stem-loop I had a nonanucleotide sequence TAATATTAC, identical to that of other geminiviruses while stem-loop II contained a Rep-binding motif (iteron; within the loop) GGTGTCT, identical to that of

ToLCV. Another iteron (AGACACC) was found upstream to the stem-loop II but in a reverse complement orientation. This satellite did not contain any significant ORF, shared no sequence similarity with the genome of its cognate virus (ToLCV), was completely depended on the cognate virus for replication, systemic movement and encapsidation, and was not essential for ToLCV replication (Dry *et al.* 1997).

2.2.1.1 Nanovirus-like DNA-1

The first nanovirus-like DNA molecule associated with geminiviruses was isolated from cotton infected with *Cotton leaf curl Multan virus* (CLCuMV) in Pakistan (Mansoor *et al.* 1999). Subsequently, similar molecules were found in many other plants infected with monopartite begomoviruses from the Old World (Briddon *et al.* 2004). These DNA molecules, named nanovirus-like DNA-1 (Fig. 2.3), comprised 1375 nt and had a common genome organization including (1) a predicted stem-loop structure containing a conserved TAGTAATAT nonanucleotide sequence within the loop, typical to that of nanoviruses, (2) a single large ORF in the positive sense encoding a homologue of the nanovirus replication-associated protein (Rep) (typically 315 amino acids) and (3) an adenine rich (A-rich) region immediately downstream of the coding region (typically 100-200 nts), the only feature different from nanovirus Rep components (Briddon *et al.* 2004). The Rep sequences of nanovirus-like DNA-1 are highly conserved (greater than 86 % amino acid sequence similarity; Briddon *et al.* 2004). Nanovirus-like DNA-1 molecules can replicate autonomously, but similar to ToLCV-sat, they depend on helper viruses for systemic movement, encapsidation and play no role in symptom induction (Briddon *et al.* 2004; Mansoor *et al.* 2003).

It has been suggested that the nanovirus like-DNA molecules were possibly captured by geminiviruses during mixed infection by trans-encapsidation. This allowed them to be transmitted by geminivirus vectors and therefore increased their host range (Mansoor *et al.* 1999; Saunders *et al.* 2002).

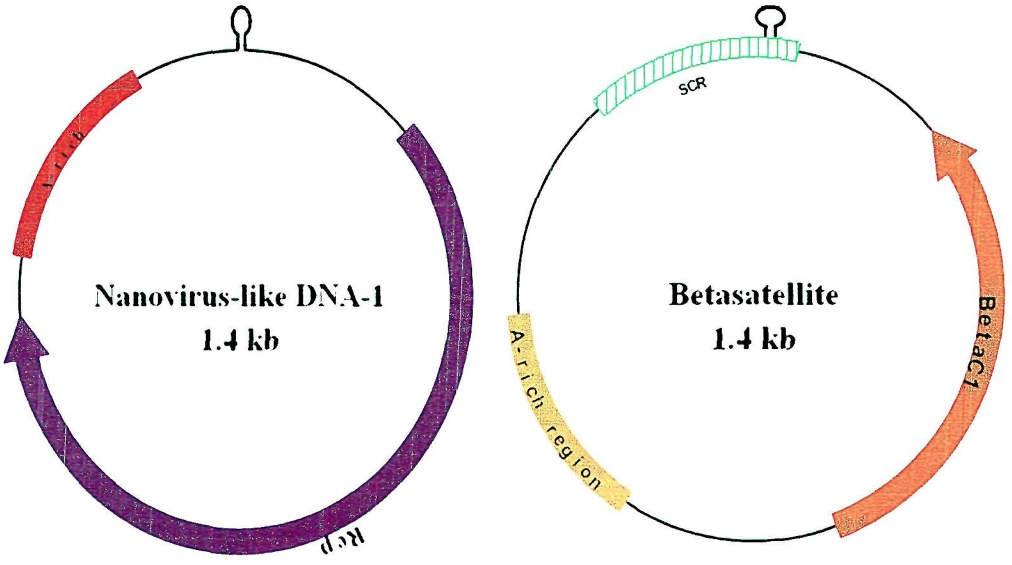
2.2.1.2 Betasatellites

Recently, another group of novel circular ssDNA molecules, named betasatellite (Fig. 2.3), have been found associated with many monopartite begomoviruses infecting a diverse range of plants (Briddon *et al.* 2003; Zhou *et al.* 2003). The first symptom modulating satellite molecules for DNA viruses were shown to be associated with the monopartite begomoviruses *Ageratum yellow vein virus* (AYVV) and CLCuMV (Saunders *et al.* 2000; Briddon *et al.* 2001). These two full-length Betasatellite molecules are just below half the length of their helper begomoviruses (49.2% for the AYVD-associated molecule originating from Singapore and 49.5% for the CLCuD-associated molecules originating from Pakistan). These Betasatellite molecules have a genome of approximately 1350 nucleotides for the full-length forms or approximately 700 nucleotides for the deleted forms, and contain three characteristic regions (Briddon *et al.* 2003).

1) Satellite conserved region (SCR). The satellite conserved region, a region of 200 nts, contains a putative stem-loop structure containing a nonanucleotide TAG/ATATTAC sequence typical of the nanoviruses and a very highly conserved region of over 100 nts located on the 5' side of the stem-loop (Briddon *et al.* 2003). This conserved region has a very high GC content (~ 70 %) (Zhou *et al.* 2003).

***2)* Adenine rich region (A rich region).** The betasatellite molecules contain an A-rich region (typically 160-180 nts; Briddon *et al.* 2003) located between nucleotide 700 and 1000 (Zhou *et al.* 2003). It was suggested that this region may be present to increase the size of these molecules to become a fraction (either half or quarter) of the typical genome size of geminiviruses (Mansoor *et al.* 2003). In doing so, the molecules could be tolerated during systemic movement which operates through a stringent size-selection mechanism (Etessami *et al.* 1989; Rojas *et al.* 1998)

3) Potential coding region. The Betasatellite molecules contain an ORF (BetaC1) on the complementary strand on 3' side of the stem-loop. This ORF encodes a protein of approximately 118 amino acids. Through mutation analysis, Zhou *et al.* (2003) demonstrated that the BetaC1 gene product is associated with symptom induction.



a) b)

Figure 2.3. Genome organization of the begomovirus satellites. SCR; satellite conserved region; A-rich: Adenine rich region; BetaC1: replication associated protein)

Mansoor *et al.* (2001) proposed that there is an association of a disease complex involving a begomovirus DNA A and a distinct Betasatellite with leaf curl disease of okra in Pakistan and according to their knowledge this was the first report of the association of a whitefly-transmitted begomovirus/DNA A-Betasatellite complex with okra leaf curl disease.

Zhou *et al.* (2003) demonstrated that Betasatellite is widely associated with begomoviruses from tobacco, tomato, *Malvastrum coromandelianum* and *Siegesbeckia orientalis* in China. The complete nucleotide sequences of 18 Betasatellite molecules were determined. Comparison showed that three types of betasatellite molecules were present among the isolates and that they are associated with the three begomovirus species *Tomato yellow leaf curl China virus* (TYLCCNV), *Tobacco curly shoot virus* (TbCSV) and *Malvastrum yellow vein virus* (MYVV). In the same year, Jose and Usha (2003) cloned and sequenced the 2.7 kb genome of BYVMV DNA-A, but the DNA-B component was not found, suggesting that BYVMV is a monopartite begomovirus. However, they successfully isolated a Betasatellite component from BYVMV-infected okra tissue and then it was amplified, cloned and sequenced. They also found that this satellite DNA is necessary for producing the typical symptoms in okra.

2.2.2 Functions of begomoviral genes

2.2.2.1 Replication-associated protein (Rep, AC1 and C1)

The Rep protein of geminiviruses is a multifunctional protein with a number of important functions including:

i) DNA binding: For initiation of replication, Rep is required to bind to the dsDNA template. Rep recognizes its cognate DNA (*ori*) in a sequence and site-specific manner, and this process involves iteron sequences upstream of the stem-loop structure (Fontes *et al.* 1994b). Although the natural substrate for Rep binding *in vivo* is dsDNA, it can bind to ssDNA *in vitro* (Fontes *et al.* 1994a). Jupin *et al.* (1995) demonstrated that the 116 N-terminal amino acids of TYLCV Rep are responsible for binding. Rep initiates the virion-strand replication by introducing a nick between nucleotides 7th and 8th of the nonanucleotide sequence (TAATATTAC) in the stem-loop (Laufs *et al.* 1995b). The Rep domain responsible for cleavage activity was mapped to the first 211 amino acids of TYLCV Rep (Heyraud-Nitschke *et al.* 1995) and the first 120 amino acids in *Tomato golden mosaic virus* (TGMV; Orozco *et al.* 1997). This N-terminal domain contains three motifs conserved among the Reps of all geminiviruses (Laufs *et al.* 1995a). Motif I (FLTY) has an unknown function, motif II (HLH) is a putative metal ion binding site, and motif III contains a highly conserved Y residue that is essential for both cleavage and joining activities (Laufs *et al.* 1995a; Orozco *et al.* 1997). The linear ssDNAs generated from RCR replication are recircularized into circular ssDNAs by the joining activity of Rep by transferring the 5' terminal phosphate of the linear ssDNA to the 3'OH end (Laufs *et al.* 1995b). These ssDNAs can either be encapsidated or go back into the replication cycle.

ii) Oligomerization: Formation of protein complexes is an essential property for origin recognition and replication in many organisms that replicate by RCR. Settlage *et al.* (1996) showed that the Rep of TGMV and BGMV formed oligomers. The authors demonstrated that this oligomerization occurred in a virus non-specific manner as Reps of the two viruses formed a complex with each other and the addition of heterologous Rep had no effect on the efficiency of replication. The Rep domain responsible for the oligomerization was mapped between amino acids 120 and 181 in TGMV. This domain contained two characteristic α-helices that were essential for the oligomerization (Orozco *et al.* 1997, 2000).

iii) Interaction with host factors associated with replication machinery: Replication of geminiviruses can occur in plant tissue that is not actively dividing (Horns and Jeske, 1991). In addition, geminiviruses do not encode a nucleic acid polymerase. Therefore, after establishing an infection, geminiviruses need to induce the replication machinery of the host cells. There is some evidence demonstrating the interaction between viral Rep and replication factors of the host. Ach *et al.* (1997) found that Rep of TGMV can bind to the RRB1, a maize retinoblastoma-related protein that is a negative regulator responsible for the G1 to S phase transition of the cell cycle. Recently, Rep of *Tomato yellow leaf curl Sardinia virus* (TYLCSV) was found to interact with PCNA (proliferating cell nuclear antigen) in *Arabidopsis*. PCNA is a ring-like protein that functions as a mobile platform or "sliding clamp" for docking of enzymes necessary for the replication of DNA (Castillo *et al.* 2003). The interaction of RepA with RRB is performed through an LXCXE motif that is located close to the splicing site (Boulton 2002).

iv) ATPase activity: The Rep C-terminal region contains a conserved motif similar to the P-loop motif of other NTP hydrolyzing proteins. Desbiez *et al.* (1995) showed that Rep of TYLCV exhibited an ATPase activity *in vitro* and demonstrated that the ability of Rep to bind and hydrolyse ATP was essential for replication. However, the nature of this activity in the replication process remains unclear because it was shown that cleavage and ligation activities do not require the participation of ATPase activity (Heyraud-Nitschke *et al.* 1995).

2.2.2.2 Coat protein (CP, AV1 and V1)

The CP of geminiviruses is a multifunctional protein required for a range of functions associated with encapsidation, accumulation of viral ssDNA, insect transmission and both intra- and inter-cellular movement (Boulton 2002). However, these functions vary from genus to genus.

i) Encapsidation: The most important function of CP is to form the shell in which genomic ssDNA is encapsidated. Initial studies on geminivirus capsid structure were done on *Chloris striate mosaic virus* (Hatta and Francki 1979). A study based on *Maize streak virus* (MSV), using cryo-electron microscopy and three dimensional image reconstruction (Zhang *et al.* 2001), revealed that geminate particles are assembled from 110 protein subunits, organized as 22 pentameric capsomers forming 2 abutting incomplete T=1 icosahedra joined together (Fig. 2.1-B). Assembly and stability of the geminivirus particles relies on interactions between CP molecules. It was suggested that the N- terminal region of one CP binds to the C-terminal amino acids of another (Hallan and Gafni 2001).

ii) Transmission by vectors: The CP plays a key role in vector transmission and in determination of vector specificity. One important experiment to prove this role was conducted with two members of different genera, in which the CP gene of *African cassava mosaic virus* (ACMV), a begomovirus, was replaced with that from BCTV, a curtovirus (Briddon *et al.* 1990). This chimeric genome produced symptoms typical of ACMV infection. The CP gene of BCTV was also expressed in plants and was shown to encapsidate the hybrid ACMV genomic ssDNA. Interestingly, *Circulifera tenellus* (the vector of BCTV) transmitted hybrid ACMV to *N. benthamiana* seedlings, but not the original ACMV. This indicated that the CP influences the vector specificity (Briddon *et al.* 1990). Similarly, a whitefly non-transmissible strain of *Abutilon mosaic virus* (AbMV), with the CP replaced with that from *Sida golden mosaic virus* (SiGMV), was acquired and transmitted by whitefly to various host plants, indicating a crucial role of CP in the transmission process (Hofer *et al.* 1997). The region associated with vector transmission was identified within positions 124-174. The mutation in this region altered virus transmission by the vector by either preventing particle assembly, or inhibiting passage of the virus from gut to haemocoel or from the haemocoel to the salivary gland of vectors (Harrison *et al.* 2002).

iii) Intra-cellular targeting: During infection, many virus-associated products (genomic DNA, replication intermediates, and proteins) need to move to particular sites in the cells. It has been suggested that this transport is conducted with the participation of viral proteins, host cytoskeletal elements and possibly host nuclear shuttle proteins (Gafni and Epel, 2002). Because geminiviruses replicate in the nucleus of infected host cells, following their inoculation into the cytoplasm by vectors, the virus needs to be transported into the nucleus for replication. Although it is still not clear if geminiviruses enter the nucleus in the form of intact virions or as nucleoprotein complexes, the presence of only the viral CP in the nucleus following initial cellular entry suggests it may be involved in the nuclear import of viral DNA. The trafficking of the viral DNA-protein complex between the nucleoplasm and protoplasm occurs through a complex structure called the nuclear pore complex (NPC) and is mediated by host transport receptors known as karyopherins that link to virus-associated proteins and then become associated with the NPC. To be recognized by host receptors, these virus-associated proteins must contain nuclear localizing signals (NLS; Gafni and Epel, 2002). Such signals have been determined for both monopartite and bipartite geminiviruses and are mainly located in the N-terminal region of the CP; 63 amino acids for TYLCV (Kunik *et al.* 1998), 5-22 amino acids for MSV (Liu *et al.* 1999) and 54 amino acids for ACMV (Unseld *et al.* 2001). For ACMV, two other domains containing NLS signals, which are located in the central (100-127 amino acids) and C-terminal (201- 258 amino acids) regions, were also determined (Unseld *et al.* 2001). The CP of geminiviruses also participates in exporting the viral genome from the nucleus to the cytoplasm. In this case, a nuclear export signal (NES) is required for recognition by a host receptor. A NES signal located in the C-terminal half of the TYLCV CP has been identified (Rhee *et al.* 2000). For bipartite begomoviruses, although nuclear export is the responsibility of the BV1 gene product, one NES was identified in the central region of the ACMV CP (Unseld *et al.* 2001). In brief, CP, in terms of intra-cellular targeting function, serves as a nuclear shuttle protein for monopartite geminiviruses and as a nuclear import protein for bipartite begomoviruses.

2.2.2.3 Pre-coat protein (PCP, MP, AV2 and V2)

This protein is only observed in Old World begomoviruses and is absent in New World begomoviruses. As this protein is located before Coat protein in its genomic position, it is referred as Pre-coat protein (PCP) and this protein is mainly involved in movement of virus hence it is also called as movement protein (Padidam *et al.* 1996). By doing mutational analysis, Wartig *et al.* (1997) showed that TYLCV V2 protein is required for a successful infection process in tomato and V2 gene products appeared to play a role in the accumulation of viral single-stranded DNA. Its function for bipartite begomoviruses is not fully understood.

2.2.2.4 AC4 protein (C4)

i) AC4 protein is involved in movement of monopartite begomoviruses: Through mutation analysis, Jupin *et al.* (1994) found that the protein encoded by the TYLCV C4 ORF was necessary for viral systemic movement. Using microinjection and transient expression assays, Rojas *et al.* (2001) suggested that the TYLCV C4 protein that contains an N-terminal putative myristoylation domain could deliver viral DNA to the plasmodesmata and mediate cell-to-cell transport.

ii) AC4 protein is involved in symptom expression of monopartite begomoviruses: Rigden *et al.* (1994) showed that plants agro-inoculated with constructs containing ToLCV C4 ORF initiation codon mutants showed significantly less symptoms than controls. For bipartite begomoviruses, the ACMV AC4 protein has been shown to bind miRNA (Chellappan *et al.* 2005) and suppress post transcriptional gene silencing (PTGS) (Vanitharani *et al.* 2004).

2.2.2.5 Replication enhancer protein (REn, AC3 and C3)

i) REn enhances replication: Sunter *et al.* (1990) observed that mutation of the AL3 ORF of TGMV created a large reduction in the levels of ss- and dsDNA. They proposed that the association between the AL3 ORF and replication depends on the interaction between Rep and the AL3 protein. Such an interaction has been found in TGMV and BGMV (Settlage *et al.* 1996).

ii) REn interacts with cell cycle regulator proteins: The interaction of the AL3 protein of TGMV (a bipartite begomovirus) with a maize retinoblastoma homolog (pRBR1) was demonstrated by Settlage *et al.* (2001). Using a yeast two-hybrid system, Castillo *et al.* (2003) found that REn of TYLCSV, a monopartite begomovirus, also interacted with PCNA of *Arabidopsis thaliana* and tomato.

2.2.2.6 Transcriptional activator protein (TrAP, AC2 and C2):

i) TrAP is a transcriptional activator protein: The AC2 protein is required for efficient transcription of virion sense viral genes such as CP and the BR1 protein (Sunter and Bisaro, 1992). Using transgenes consisting of complete and truncated versions of the CP promoter of TGMV fused to the GUS reporter gene, Sunter and Bisaro (1997) found that TrAP activated the CP promoter in mesophyll cells but repressed it in phloem tissue. The biochemical properties of TrAP were also elucidated showing that it (1) had ability to bind to ssDNA in a sequence non-specific manner and to zinc ions, (2) was phosphorylated and (3) contained a minimal transcriptional activation domain comprising 15 C-terminal amino acids (Hartitz *et al.* 1999).

ii) TrAP is a potential silencing suppressor: The ability of TrAP to act as a suppressor of post-transcriptional gene silencing was first shown with ACMV (Voinnet *et al.* 1999). Sunter *et al.* (2001) demonstrated that transgenic tobacco plants expressing the AL2 ORF of TGMV showed enhanced susceptibility to infection of TGMV, BCTV and *Tobacco mosaic virus* (TMV), an unrelated RNA virus. This function seemed to be independent of the transcriptional activity because the activation domain located in the C-terminal region was truncated in the AL2 transgene.

2.2.2.7 Genes on DNA-B of bipartite begomoviruses.

DNA-B of bipartite begomoviruses encodes two proteins, BV1 (NSP) and BC1 (MP), both involved in viral movement.

2.2.2.7.1 Nuclear shuttle protein (NSP, BV1)

i) BV1 is a nuclear shuttle protein. BV1 functions as a nuclear shuttle protein that is responsible for transporting viral ssDNA into and out of the nucleus. However, it is not involved in the nuclear import of viral ssDNA during initial infection, which is facilitated by the CP (Gafni and Epel 2002). The NLS of *Squash leaf curl virus* (SqLCV) BV1 was mapped to the N-terminal 113 residues (Pascal *et al.* 1994) and contained a sequence of 22 amino acids containing the motif SLEKDLLIDLH, resembling the NES of other nuclear shuttling proteins (Ward and Lazarowitz 1999).

ii) BV1 interacts with BC1 for cell-to-cell movement. BV1 enters the nucleus to form a complex with viral ssDNA that moves to the cytoplasm and is trapped by BC1. The complex BV1:BC1:ssDNA then moves to the plasmodesmata and is transferred to the adjacent cell (Gafni and Epel 2002). The C-terminal region of the SqLCV BV1 was shown to be essential for interaction with BC1 (Sanderfoot *et al.* 1996).

2.2.2.7.2 Movement Protein (MP, BC1)

i) BC1 is a movement protein. The function of BC1 as a MP was demonstrated in two cases: (1) *Bean dwarf mosaic virus* (BDMV) BC1 increased the size exclusion limit (SEL) of plasmodesmata (Noueiry *et al.* 1994; Rojas *et al.* 1998), and (2) SqLCV BC1 induced formation of a tubular structure derived from the endoplasmic reticulum that facilitated viral translocation between cells (Ward *et al.* 1997). As mentioned above, it has been proposed that BC1 interacts with BV1 through the BV1: BC1: ssDNA complex for cell-to-cell movement (Gafni and Epel 2002).

ii) BC1 is involved in pathogenicity. The association of BC1 with pathogenicity has been proven in transgenic experiments. Tobacco and tomato plants transformed with the BC1 gene of *Tomato mottle virus* (ToMoV) and BDMV respectively, expressed characteristic visible symptoms of viral infection. The BC1 genomic region responsible for induction of pathogenicity was mapped to the C-terminus since transgenic lines containing a deletion of this region (e.g. 119 amino acids for BC1 of TMoV), were all symptomless (Duan *et al.* 1997; Hou *et al.* 2000).

2.3 REPLICATION AND RECOMBINATION

2.3.1 Replication

Geminiviruses have a small genome and encode only a few proteins. Therefore, their DNA replication cycle relies largely on the use of cellular DNA replication proteins. Geminiviruses replicate via a rolling-circle mechanism (Stenger *et al.* 1991) with the virion strand replication origin (V-*ori*) located in a conserved nonanucleotide motif (TAATATT↓AC), and conserved iterated sequences (iterons) required for specific recognition and binding by Rep during replication (Arguello-Astorga *et al.* 1994).

The geminivirus replication cycle can be subdivided in several functionally distinct stages characterized by specific events. Early during the infection process, viral particles are injected by the insect vector, presumably uncoated, and the viral genome is transported into the host cell nucleus by some unknown mechanism (Lazarowitz *et al.* 1992; Palmer and Rybicki 1998). Inside the nucleus, amplification of the viral genome, which involves an efficient DNA replication process, occurs in three distinct stages (Fig. 2.4; Stanley and Townsend 1985; Stenger *et al.* 1991). First stage (A) involves the conversion of the genomic circular ssDNA [(c)ssDNA] into supercoiled covalently closed circular dsDNA intermediates [(ccc)dsDNA] or replicative form I (RFI). Second stage (B) is the amplification of the dsDNA intermediates by a rolling-circle mechanism. Third stage (C) is the production and encapsidation of mature genomic circular ssDNA into viral particles. This strategy is remarkably similar to that of many prokaryotic replicons (Baas and Jansz 1988; Novick 1998). Besides RCR, recombination-driven replication (RDR) mechanism also contributes to the viral DNA synthesis to some extent (Werner *et al.* 2003).

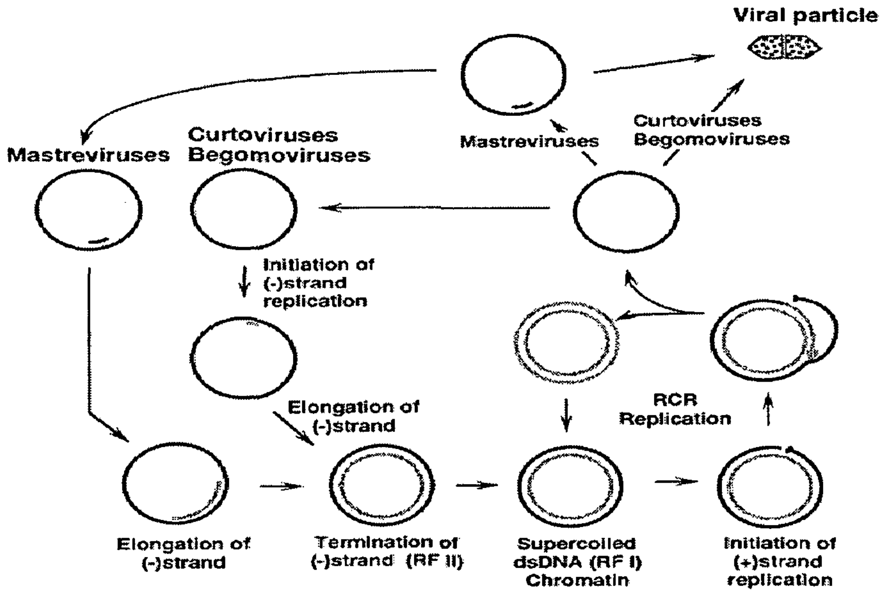


Figure 2.4. Summary of the geminivirus replicative cycle. Proposed stages of the geminivirus DNA replication cycle (adapted from Gutierrez 1999)

2.3.2 Recombination

Genetic recombination is a ubiquitous biological process that is both central to DNA repair pathways (Cromie *et al.* 2001; Michel 2001) and an important evolutionary mechanism. The rapid replication of viruses and the lack of proofreading or repair mechanisms give the potential for much mutagenic variation in both RNA and DNA viruses. Three factors contribute to the success of recombination, giving rise to new viruses: mixed infections, high levels of viral replication, and increased host range of the vector (Hull 2009). By generating novel combinations of preexisting nucleotide polymorphisms, recombination can potentially accelerate evolution by increasing the population wide genetic diversity upon which adaptive selection relies. The viability of recombinants is apparently largely dependent on how severely recombination disrupts the coevolved intragenome interaction networks (Escriu *et al.* 2007; Jain *et al.* 1999; Martin *et al.* 2005). These networks include interacting nucleotide sequences that form secondary structures, sequence-specific protein-DNA interactions, inter-protein interactions, and amino acid-amino acid interactions within protein three-dimensional folds. One virus family where such interaction networks appear to have a large impact on patterns of natural interspecies recombination is the single-stranded DNA (ssDNA) geminiviruses. Genome organization and rolling circle replication (RCR), the mechanism by which geminiviruses and many other ssDNA viruses replicate, seem to have a large influence on basal recombination rates in different parts of geminivirus genomes (Garcia-Andres *et al.* 2007; Jeske *et al.* 2001; Owor *et al.* 2007). One of the earliest pieces of evidence for recombination amongst geminiviruses was obtained from studies of a severe mosaic disease of cassava in Uganda (Zhou *et al.* 1997). Sequence analysis revealed that a virus responsible for the disease, *East African cassava mosaic virus*–Uganda (EACMV-UG) had probably arisen by interspecific recombination between EACMV and ACMV. Using a program to detect gene conversion (GENECONV), Padidam *et al.* (1999) searched for recombination events among geminiviruses from sequences representing 64 distinct species. In total, the analysis identified 420 statistically significant recombinant fragments distributed across the viral genomes. The fragments (391) detected between viruses from different continents and between begomoviruses and curtoviruses were located in the N-terminal region of Rep, suggesting that these are old events that presumably occurred before the geographical isolation. This important analysis suggested that interspecific recombination has resulted in remarkable diversity among geminiviruses and could be a major cause of the emergence of new geminivirus diseases. At present, the number of new geminiviruses arising as a consequence of recombination is increasing (Fauquet *et al.* 2005; Garcia-Andres *et al.* 2006; Girish and Usha 2005; Idris and Brown 2005; Kon *et al.* 2006; Rojas *et al.* 2005; Rothenstein *et al.* 2006 ; Were *et al.* 2005). In some cases, the recombinants exhibited a new pathogenic phenotype which is often more virulent than the parents. For example, a natural recombinant between TYLCSV and TYLCV has been detected which has a wider host range than for the individual viruses and which is becoming prevalent in geminivirus populations infecting tomato in Spain (Monci *et al.* 2002).

2.4 EVOLUTION OF GEMINIVIRUSES

The evolution of geminiviruses is subject to point mutations and recombination. The mutation rate of geminiviruses is dependent on the type of virus, host plant, age of the inoculated plant and inoculum homogeneity (Duffy and Holmes 2009; Ge *et al.* 2007 and Isnard *et al.* 1998). More importantly, it is believed that recombination among different DNA-A components is the main source of molecular variation among geminiviruses (Padidam *et al.* 1999) and can result in gain of virulence for the helpers and acquisition of new satellite molecules.

Begomoviruses acquired betasatellite components at some point during their evolution, although the origin of the satellite DNA remains unknown. However, the presence of an A-rich region (Fig. 2.3) could indicate that betasatellite adapted from a preexisting component. This might have occurred either by association with a monopartite begomovirus (Fig. 2.5) or by displacement of DNA B from an Old World bipartite begomovirus as demonstrated experimentally for SLCMV (Saunders *et al.* 2002). Bipartite begomoviruses might have evolved from an ancestral monopartite virus by component duplication and the acquisition of novel genetic material (Fig. 2.5). Begomoviruses must have acquired nanovirus components during mixed infections (Mansoor *et al.* 1999; Saunders and Stanley 1999). Mansoor *et al.* (2003) suggested that nanovirus-like components were captured after Betasatellite because they have not been found in its absence.

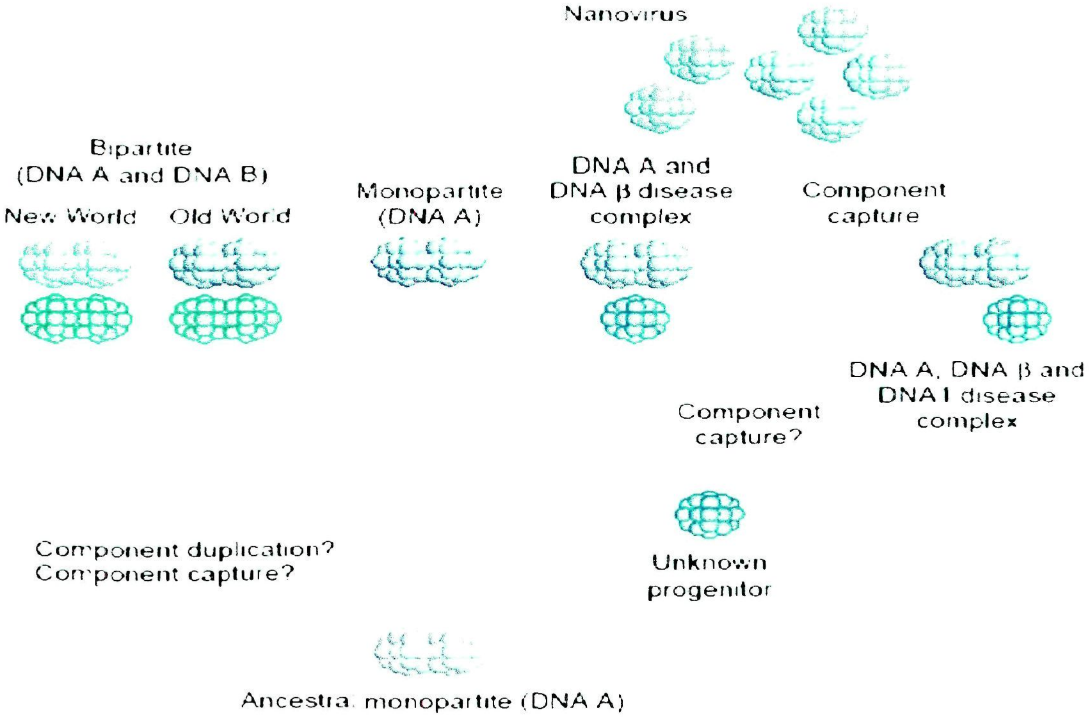


Figure 2.5. Possible evolutionary relationships between begomoviruses, DNA β and nanovirus-like components (adapted from Mansoor *et al.* 2003)

2.5 DETECTION AND AMPLIFICATION OF VIRAL GENOME

Several methods, particularly those based on protein or nucleic acid detection, have been developed to identify geminiviruses.

2.5.1 Serological detection

Serological detection techniques, such as ELISA or its variants, are based on the antigenic properties of the viral coat protein. Traditionally, the techniques have been a primary means of virus detection and diagnosis. For mastreviruses, polyclonal and monoclonal antibodies has been used to detect and differentiate the isolates of SMV (Bosque-Perez 2000; Peterschmitt *et al.* 1991; Pinner and Markham 1990), three distinct viruses, namely *Chloris striate mosaic virus* (CSMV), *Paspalum striate mosaic virus* (PaSMV) and *Digitaria striate mosaic virus* (DDSMV) infecting graminaceous plants from Australia (Pinner *et al.* 1992) and a Syrian chickpea isolate of *Chickpea chlorotic dwarf virus* (CpCDV), a dicot mastrevirus (Kumari *et al.* 2006). For begomoviruses, however, serological-based diagnostics have met with limited success because the particles are only moderately immunogenic, are purified with difficulty from plant materials, and occur in only low to moderate concentration in plants tissue. Consequently, serological techniques using polyclonal antibodies (pAbs) lack both the specificity and sensitivity required for accurate diagnosis (Harrison and Robinson 1999; Harrison *et al.* 2002 and Pico *et al.* 1996). These problems have been resolved, to a certain degree, by several methods including techniques such as ISEM and ELISA using a range of polyclonal and monoclonal antibodies (mAb) (Harrison *et al.* 2002; Pico *et al.* 1999). Finally, recombinant pAbs against BGMV, *Cabbage leaf curl virus* (CaLCuV), ToMoV and TYLCV have been generated using the CP expressed in *Escherichia coli* as immunogenic sources. These antibodies are inexpensive and have high sensitivity for detection of begomoviruses (Abouzid *et al.* 2002). Harrison *et al.* (2002) identified more than 50 distinct begomoviruses originating from over 30 countries on six continents using selected mAbs.

2.5.2 Genomic DNA based detection

2.5.2.1 DNA Hybridization:

Hybridization techniques have been widely used in the diagnosis of plant viruses. These techniques are based on the base pairing between viral nucleic acid sequences (target) and labeled probes whose sequence is complementary to that of the targets. Hull (1993) summarized in detail the factors affecting hybridization including temperature, nucleic acid composition, sequence length and base mismatch, salt concentration, pH and organic solvents. There are three major formats for DNA hybridization techniques; the dot blot technique, the tissue print (or squash plot) technique and the Southern blot technique. In the dot blot technique, the DNA extracts are dotted onto a nylon membrane for hybridization (Gilbertson *et al.* 1991; Harper and Creamer 1995; Polston *et al.* 1989; Polston *et al.* 1999; Potter *et al.* 2003; Stonor *et al.* 2003). The dot blot can also be used to estimate relative differences in viral nucleic acid titres in infected tissues (Gilbertson *et al.* 1991). In the tissue print (or squash plot) technique, the infected tissue is squashed directly onto a nylon membrane for hybridization. This method provides a specific, rapid, and simple means to detect virus without DNA extractions. The technique has been use to detect geminiviruses in field samples (Czosnek and Laterrot 1997; Gilbertson *et al.* 1991; Pico *et al.* 1996) and in assessing virus resistance (Maruthi *et al.* 2003; Rubio *et al.* 2003). In the Southern blot technique, the DNA extracts are electrophoresed through an agarose gel followed by transfer onto nylon membrane for hybridization. This technique enables the detection of the characteristic replicative forms of viral DNA present in plants, including open circular dsDNA, supercoiled dsDNA and circular ssDNA. The technique has been widely used to characterize new viruses (Bigarre *et al.* 2001; Lotrakul *et al.* 1998), investigate the presence of viruses in whitefly vectors (Ghanim *et al.* 1998), study gene functions (Briddon *et al.* 1990; Noris *et al.* 1998; Orozco and Hanley-Bowdoin 1996; Padidam *et al.* 1996; Petty *et al.* 2000; Wartig *et al.* 1997), study the interaction between host factors and virus (Pooma *et al.* 1996), to discover the replication intermediates of viruses (Jeske *et al.* 2001) and to confirm the presence and role of satellite molecules in disease induction (Briddon *et al.* 2004; Briddon *et al.* 2001; Bull *et al.* 2004; Mansoor *et al.* 1999).

2.5.2.2 Polymerase chain reaction (PCR)

The most widely used amplification method is the polymerase chain reaction (PCR), of which many specific protocols exist for various applications. PCR has been widely used in detection and diagnosis of plant viruses because of its rapidity, sensitivity, specificity and reliability (Henson and French 1993; Martin *et al.* 2000). PCR, using degenerate (or universal) primers designed from highly conserved regions of virus genomes, has become a rapid and reliable way to screen mixed infections or to detect new geminiviruses in plants or vectors (Guo and Zhou 2005; Harrison *et al.* 1997; Lyttle and Guy 2004; Rampersad and Umaharan 2003; Rojas *et al.* 1993; Wyatt and Brown 1996) and their satellites (Briddon *et al.* 2002; Bull *et al.* 2003; Zhou *et al.* 2003).

Usually, however, PCR produces relatively short amplicons and is dependent on specific primer sequences. As the quality of genome analysis increases with the length of the generated sequences, techniques enabling the amplification of whole genomes are most advantageous.

2.5.2.3 Rolling circle amplification (RCA)

Among the new available techniques for whole genome amplification, rolling-circle amplification (RCA) using the DNA polymerase of bacteriophage phi29 is showing great value for the analysis of viral DNA genomes. This polymerase possesses several features, such as strand displacement activity, proofreading activity and generation of very long synthesis products (Fig. 2.6), which make it most suitable for the efficient amplification of circular DNA molecules from complex biological samples (Johne *et al.* 2009)

In 2004, RCA using Φ29 DNA polymerase was first applied to viral genomes. Rector *et al.* (2004) showed that the papillomavirus genome, which is composed of circular double-stranded DNA, could be efficiently amplified from tissue samples using this technique. At the same time, Inoue-Nagata *et al.* (2004) amplified and cloned the complete circular DNA genome of a begomovirus, based on the characteristic that the bacteriophage ΦDNA polymerase amplifies circular DNA in a rolling circle amplification mechanism. Total DNA extracted from infected tissue was used as the template of an amplification reaction using the commercial kit TempliPhi (Amersham Biosciences). The amplified DNA could be used for direct sequencing and was cloned after digestion with a single cutting restriction endonuclease. The use of this enzyme simplified the cloning steps and increased the cloning efficiency of the complete genome of a circular plant DNA virus.

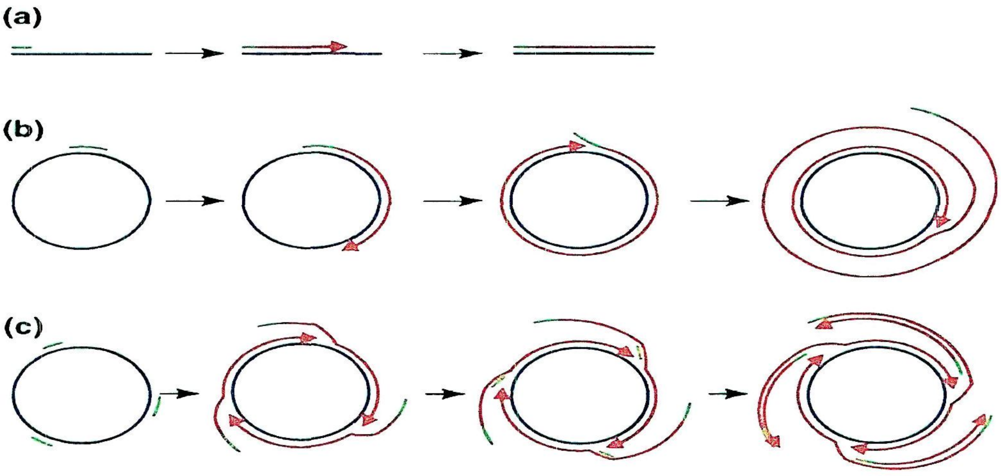


Figure 2.6. Principle of rolling-circle amplification. Blue lines denote target DNA sequences, green lines represent oligonucleotide primers and red lines represent new DNA synthesized by the polymerase. Arrow heads indicate 3'-ends of growing DNA strands.

(a) Linear template and single primer. After primer binding, the polymerase synthesizes one complementary strand.

(b) Circular template and single primer. The polymerase synthesizes a complementary strand beginning at the bound primer. After one round, the primer and the synthesized strand are displaced and DNA synthesis continues for additional rounds, producing a long concatemeric single-stranded DNA

(c) Circular template and multiple random primers. The synthesis is initiated at multiple primers bound to the template. DNA synthesis using strand displacement is carried out as in (b). However, primers still present in the reaction mixture bind to the displaced strand and are used as additional initiation points for DNA synthesis. The multiple products are long concatemeric molecules of double-stranded DNA (adapted from Johne *et al.* 2009)

RCA has been applied to detect several different known and unknown geminiviruses, including monopartite and bipartite species. An investigation of plant tissues revealed four novel geminivirus species from two different genera using direct sequencing of RCA products (Schubert *et al.* 2007; Knierim and Maiss 2007). A diagnostic system using RCA combined with restriction enzyme analysis has been established by Haible *et al.* (2006) for the detection and differentiation of known geminivirus species. They had shown that RCA in combination with RFLP is a highly reproducible tool for geminivirus diagnosis, largely independent of viral genome organization, source plant type and origin, and sample preparation. The combination of RCA/RFLP is vastly superior to PCR-based procedures. Neither specific primers are needed to start the reaction, nor expensive equipment like a thermocycler.

Shepherd *et al.* (2008) showed that this method could also be successfully applied to dried plant tissue. They cloned 77 genomes of geminiviruses using RCA products from 102 dried plant samples, some of which had been stored for 10 years. In same year, Kumar *et al.* (2008) characterized a distinct bipartite begomovirus species *Tomato leaf curl Palampur virus* (ToLCPMV) infecting tomato in sub-temperate region in India. The complete DNA-A and DNA-B components were amplified through rolling circle amplification (RCA) using Φ-29 DNA polymerase and then sequenced the whole genome of the virus.

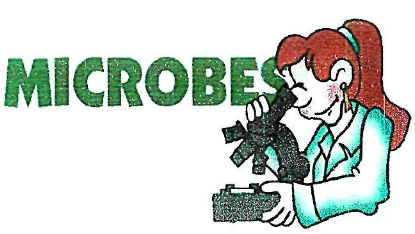
Advantages of rolling circle amplification:

No prior sequence information of the target is required, which increases the possibility of discovering novel viruses. As RCA takes place at isothermal conditions, costly thermocyclers are not required for the amplification. Minimal reagents are required in RCA and it avoids the generation of false-positive results (Rector *et al.* 2004; Wang *et al.* 2005). The only natural source of false positive in plants is small mitochondrial plasmids (Homs *et al.* 2008). RCA also avoids artificial recombination of DNAs, which may arise during PCR. Proof-reading activity of Φ-29 DNA polymerase makes RCA a very reliable method. Direct sequencing of the RCA products can be performed with high precision, which helps in quick detection of the virus. All infecting circular DNA components can simultaneously be amplified in a single RCA reaction, including satellites and defective interfering DNAs as well as other circular DNA viruses present in mixed infections. This feature of RCA makes it much more informative than PCR (Haible *et al.* 2006). Another advantage of RCA is that it can work dried leaf samples and mixed-infected plants (Schubert *et al.* 2007). Last but not the least; RCA can be used to construct infectious clones of geminiviruses in a simple one-step cloning approach (Wu *et al.* 2008; Ferreira *et al.* 2008).

2.6 INFECTIOUS CLONE CONSTRUCTION

Routine infectious clone construction relies upon cloning of the whole genome, which is then sub-cloned as a tandem one-and-a-bit-mers or two genome (containing two replication origins) cassette into a vector prior to agroinoculation. Mandal *et al.* (1997) constructed the infectious clone of a bipartite geminivirus *Mungbean yellow mosaic virus* (MYMV) and showed that the tandem dimeric constructs of DNA-A and DNA-B were infectious when inoculated together and induced systemic yellow mosaic symptoms. In case of monopartite begomoviruses, Jose and Usha (2003) cloned and constructed an infectious clone of a *Bhendi yellow vein mosaic virus* (BYVMV) using the partial tandem repeat constructs of both DNA-A and Betasatellite. The BYVMV DNA-A clones were unable to induce symptoms typical of yellow vein mosaic disease in okra. But together with Betasatellite causes typical yellow vein disease symptoms in okra, showing that yellow vein mosaic disease in India is caused by association of a Betasatellite satellite with a Begomovirus. In another approach, Chakraborty *et al.* (2003) constructed an infectious clone of the virus by partial tandem dimeric constructs of DNA-A and DNA-B and were co-inoculated by particle bombardment. They found that DNA-A alone also is infectious, but symptoms were milder and took longer to develop. Ueda *et al.* (2004) developed the infectious clone of a *Tomato yellow leaf curl* Yaizu isolate (TYLCV-SzY). They amplified the full-length 2.8-kbp PCR products using two sets of abutting primers. Monomer PCR products obtained using abutting primers then treated with restriction enzymes and then cloned in binary vectors for infectious clone construction.

The construction of cassettes containing two replication origins is, however, a time consuming process. Ferreira *et al.* (2008) and Wu *et al.* (2008) independently developed an improved method for rapid construction of agro-infectious begomovirus clones, which employs limited restriction digestion of multimeric viral genomes produced by RCA, followed by direct cloning of DNA fragments containing genome dimers into appropriate vectors. All agroinoculated plants became infected, confirming the infectivity of the clones. This approach was proven to be extremely fast and useful for the production of infectious clones.



Materials and Methods

3. MATERIALS AND METHODS

The present study was carried out in the Plant Virology Lab, Division of Floriculture, Institute of Himalayan Bioresource and Technology (IHBT), Council of Scientific and Industrial Research (CSIR), Palampur, Himachal Pradesh. The material used and the methodology adopted to achieve the objectives of the investigation are presented in this chapter.

3.1 PLANT MATERIAL

In the present study, okra leaves were collected from different locations in Kangra and Bilaspur districts of Himachal Pradesh. Collections were made during the months of August-September. The collected samples were stored at -80°C until use.

3.2 EXTRACTION OF PLANT GENOMIC DNA

Okra contains the high amounts of polysaccharides and phenolic compounds. When cells are disrupted, these cytoplasmic compounds can come into contact with nuclei and other organelles (Loomis 1974). In their oxidized forms, polyphenols covalently bind to DNA giving it a brown color and making it useless for most research applications (Katterman and Shattuck 1983; Guillemaut and Maréchal-Drouard 1992; our observations). So the modified CTAB DNA extraction protocol was developed to get the good quality DNA from okra. A modification of the Doyle and Doyle (1990) CTAB extraction procedure was adopted, which uses a higher CTAB and NaCl concentration to remove polysaccharides (Lodhi *et al.* 1994), thereby preventing their interaction with DNA. CTAB is a cationic detergent that binds to nucleic acids, denatures proteins and separates nucleic acids from proteins. Rapid freezing of the material using liquid nitrogen (N_2) stops oxidation and makes the tissue brittle. Incubation of disrupted samples at 65°C inactivates most cellular enzymes immediately. Tris prevents precipitation by acids and EDTA binds Mg^{2+} and other cations required for nuclease activity. NaCl prevents the precipitation of CTAB nucleic acid complexes. PVP and sodium sulphite were also added to prevent oxidation of phenolic compounds. Extractions with chloroform remove proteins and lipids, and isoamyl alcohol is added to improve phase separation. Nucleic acids remain in the aqueous phase which can be precipitated by isopropanol. Washing with 70 % alcohol reduces the salt content in the pellet. The standardized protocol is as follows:

(i) The plant sample (50-100 mg) was ground in liquid N_2 in a pestle and mortar and the powder was taken in a 1.5 ml sterile tube

(ii) 500µl of CTAB buffer was added, mixed by vortexing

(iii) 200µl of 5% sarcosyl buffer was added and mixed again by vortexing

(iv) The mixture was shaken for 1h at 65°C in water bath

(v) The mixture was centrifuged at maximum speed (14000 rpm) for 5 min and the supernatant was collected in new tube

(vi) Equal volume of CI solution was added to the supernatant and mixed by inverting the tube

(vii) The mixture was spun down at maximum speed (14000 rpm) for 5min

(viii) The aqueous phase was collected and transferred to a new tube

(ix) Equal volume of Isopropanol was added

(x) The reaction tube was inverted three-four times for precipitation

(xi) The pellet was spun down at maximum speed for 10min at 4°C

(xii) The pellet was washed with 70% ethanol (EtOH)

(xiii) The mixture was spun down at maximum speed for 5min at room temperature

(xiv) Ethanol was evaporated for 10 min at room temperature and the pellet was dissolved in 50 µl of 18.2 Ω sterile H_2O (ddH_2O) (stored at -20°C)

Note: Interphase material should not come out with the aqueous phase.

CTAB buffer: 200mM Tris-HCl, pH 8.0, 50mM EDTA, 2.2 M NaCl, 2.5% CTAB, 4% mercaptoethanol, 0.6% sodium sulphite, 2% PVP

CI solution: Chloroform: Isoamyl alcohol (24:1)

RNase A treatment of total nucleic acids

The total nucleic acids were incubated with DNase and protease-free RNase A (Fermentas, Maryland, USA) at 37°C for one hour as described below:

Total nucleic acids	50 µl (35-40 µg)
RNase A (10 mg/ml stock)	0.5 µl

After RNase A treatment the nucleic acids were electrophoresed in 0.7 % agarose gel after mixing with loading dye for analyzing their integrity and quality. Quantification of nucleic acids was performed by using NanoDrop® ND-1000 UV–Vis Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

3.3 DETECTION OF BEGOMOVIRUSES

3.3.1 PCR based detection

For detection and characterization of begomovirus DNA components from the samples, PCR was carried out in a GeneAmp® PCR system 9700 (Applied Biosystems, California, USA) using degenerate primers for DNA-A (Table 3.1).

Table 3.1 Primers used for detection of different genomic components of begomoviruses

	Primer Name	Primer Sequence	Product size	Reference
DNA-A	AV494	5' GCCYATRTAYAGRAAGCCMAG 3'	~570 bp	Wyatt and Brown (1996)
	AC1048	5' GGRTTDGARGCATGHGTACATG 3'		

The pattern of thermal cycling for each primer pair is also summarized (Table 3.2). A reaction mixture of 50 µl consisted of:

Total DNA	5 µl (200-500 ng)
10x Taq buffer A	5 µl
10mM dNTP mix	2 µl
Forward primer	1 µl (200 ng)
Reverse primer	1 µl (200 ng)
Taq DNA polymerase	0.5 µl (3 units/µl)
ddH_2O to	50 µl

10x Taq buffer A: 100 mM Tris (pH9.0), 500 mM KCl, 15 mM $MgCl_2$, 0.1% Gelatin

Nucleotides at degenerate positions are represented by a single letter of the IUPAC ambiguity code: D = A,G,T; H = A,C,T; K = G,T; M = A,C; N = A,C,G,T; R = A,G; W = A,T; Y = C,T

Table 3.2 Thermal cycling pattern for different primer pairs

Primer Pair	Thermal cycling pattern
AV494 AC1048	Initial denaturation at 94°C for 3 min followed by 35 cycles consisting of 94°C for 30 sec, 60°C for 30 sec, 72°C for 45 sec and final extension at 72°C for 10 min

3.3.2 Agarose gel electrophoresis

Gel electrophoresis was carried out using a submarine horizontal agarose slab gel as described by Sambrook *et al.* (1989). Appropriate amount of agarose (0.8-2.0 %, as per requirement) was completely dissolved in 1x TAE or 0.5x TBE buffer by boiling. The melted agarose was cooled to 50°C and poured into an appropriate gel casting tray. The slot-forming comb was inserted, making sure that no bubbles were trapped underneath the combs and all bubbles on the surface of agarose were removed before the gel sets. After solidification of the agarose, the comb was removed and the gel was placed in the gel tank for electrophoresis. PCR product (20 µl) were properly mixed with 4 µl of gel loading dye and loaded onto the wells of the gel. DNA ladder was also loaded to compare the size of the PCR product. Electrophoresis was carried out at a constant voltage of 5 V/cm. After electrophoresis, the gel was stained with ethidium bromide (EtBr; 1.0 µg/ml) and visualized using a UV trans-illuminator and photographed using gel documentation system (Cell Biosciences, Alpha Innotech Corp., CA, USA).

Sample loading dye (6X): 15% Ficoll 400; 0.25% bromophenol blue and xylene cyanol and stored at room temperature.

TBE (10X per liter): 108 g Tris base; 55 g boric acid and 9.3 g EDTA.

TAE (50X per liter): 242 g Tris base; 57.1 ml glacial acetic acid and 100 ml EDTA (0.5 M, pH 8.0).

3.4 RCA BASED DETECTION

RCA is widely used for detection of circular genomic components from complex biological samples. The RCA was carried out using a TempliPhi^{TM} Amplification Kit (GE Healthcare, UK). . All the steps are followed as given in kit manual. The RCA includes mainly three thermocycling steps as in PCR but it doesn't need any thermocycler machine. First step is denaturation to destroy residual enzymes (nucleases, proteases etc.) and to denature dsDNA. After denaturation the secondary structures of cccDNA are disrupted to a certain extent, which provides more templates for RCA. In second step, annealing of random hexamer primers (supplied with the kit buffer) to every ssDNA was carried out by cooling to room temperature and in last step; the subsequent polymerization was carried out at 30°C for 18-20 hrs. In contrast to PCR; during RCA, both the annealing and polymerization steps were carried out simultaneously at 30°C. The standardized RCA protocol is described below:

(i) 5 µl of TempliPhi Sample Buffer was transferred to an Eppendorf tube.

(ii) Sample DNA (isolated as described before) was transferred to the dispensed TempliPhi Sample Buffer. Amounts of sample DNA varied (20-200 ng) depending on the source of the starting material. However, total volume of the DNA never exceeded 1.5 µl.

(iii) The sample was denatured by heating at 95°C for 3min and cooled down to room temperature or 4°C immediately.

(iv) In a separate tube, 5 µl of TempliPhi Reaction Buffer was combined with 0.2 µl of TempliPhi Enzyme Mix.

(v) 5 µl of this mixture was transferred to the cooled, denatured sample (step 3).

(vi) The reaction mixture was incubated at 30°C for 4-18 h in a water bath.

(vii) The enzyme (Φ-29 DNA polymerase) was inactivated by heating the reaction mixture at 65°C for 10 min. The mixture was cooled down to 4°C.

The RCA product (~150ng) was then loaded in agarose gel for electrophoresis to detect the presence of high molecular weight amplified viral DNA.

3.4.1 RFLP of RCA amplified viral DNA components

Restriction fragment length polymorphisms are caused by variations in the DNA sequence. Its analysis involves fragmentation of genomic DNA by a restriction enzyme and gel electrophoresis for separation of the cleaved DNAs according to their length. Class II restriction enzymes which recognize four or six nucleotide-sequences were used for this purpose. Because a four base recognizing enzyme has higher probability of restriction sites in a begomovirus component, *Hpa*II (recognizes C↓CGG) was used to generate distinct restriction patterns for different viruses as described by Schubert *et al.*, (2007). This method was quite useful in differentiating distinct viral species in different samples based on their polymorphic restriction patterns in a 2 % agarose gel. The method also provided ideas about the genomic nature (monopartite/bipartite) of the virus under study. Restriction reactions were set up as follows:

RCA product	1.5 µl
10x Reaction Buffer	2 µl
Restriction enzyme	0.5 µl (5 Units)
ddH_2O to	20 µl

The reaction mixtures were incubated at 37°C for 3-4 h and the enzyme was inactivated by incubation at 65°C for 10 min before loading directly to the gel.

3.5.3 Cloning RCA amplified complete viral DNA components

For cloning a complete viral DNA component into suitable vector, the primary requirement is identification of a restriction enzyme that has a single cut site in the entire component.

3.5.3.1 Preparation of the viral DNA for cloning

Aliquots (1 µl each) of the RCA product were digested with several restriction enzymes (as described above) in separate reactions and electrophoresed in 1.5 % agarose gel. The restriction enzymes, having single restriction site in multiple cloning site (MCS) of pBluescriptII KS+, were selected for digestion. The enzyme that produced either ~2.7 kb (indicating complete DNA-A or DNA-B) or ~1.4 kb band (indicating alpha or betasatellite) was selected for cloning. More amount of the RCA product (5 µl) was digested with that enzyme and enzyme inactivation was performed as per manufacturer's instructions. The digested product was then purified by PCI (Phenol: Chloroform: Isoamyl alcohol) purification method. This method enhanced cloning efficiency.

3.5.3.2 PCI purification

To the digested product, an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added. The mixture was the mixed by inverting the tube several times. Then it was centrifuged at 13000 rpm for 5 min at 4^0C. After centrifugation, the upper aqueous phase was carefully transferred to new eppendorf tube. Two volumes of absolute ethanol was added to this aqueous phase and mixed by inverting the tube several times. This mixture was then stored at -80^0C for 30 min. then it was centrifuged at 13000rpm for 20 min. at 4^0C. Supernatant was discarded and the DNA pallet was washed with 70% ethanol. Then the pallet was dried at room temperature to remove any traces of ethanol. The pallet was then dissolved in sterile water.

3.5.3.3 Preparation of the vector DNA for cloning

The pBluescriptII KS+ vector DNA (1-2 µg) was digested with the restriction enzyme selected for cloning followed by enzyme inactivation as per manufacturer's instructions. The reaction mixture was purified using PCI purification method as mentioned above. To remove the 5' phosphate residues from the digested vector fragments, the purified DNA was treated with shrimp alkaline phosphatase (SAP) (USB, Ohio, USA) in a following reaction:

Digested vector DNA	1 to 1.5 µg
10x SAP reaction buffer	5 µl
Shrimp alkaline phosphatase	1 µl (1 unit)
ddH$_2$O to	50 µl

The reaction mixture was incubated at 37°C for 30 min followed by incubation at 65°C for 15 min to inactivate SAP. Removal of the 5' phosphate residues prevents the digested vector to self-ligate and circularize thereby increasing the cloning efficiency many folds. The SAP treated linearized vector DNA was purified using PCI purification (as described above). Only 50-100 ng of the prepared vector was used in a ligation reaction and the remaining amount was stored at -20°C for future use.

3.5.3.4 Ligation reaction

Ligation reaction of viral and vector DNA was set up in an eppendorf tube as follows:

Vector DNA	20-50 ng
Insert DNA	50-150 ng
10x T4 DNA Ligase Buffer (Fermentas)	2 µl
T4 DNA Ligase (Fermentas)	1 µl (Weiss units)
Water to	20 µl

The reaction mixture was incubated at 16°C overnight in a water bath followed by storage at -20°C until transformation.

3.6 PREPARATION OF COMPETENT CELLS AND THEIR TRANSFORMATION

Since DNA is a very hydrophilic molecule, it won't normally pass through a bacterial cell's membrane. In order to make bacteria take in the plasmid, they must first be made "competent" to take up DNA. This is done by creating small holes in the bacterial cells by suspending them in a solution with a high concentration of calcium. For the transformation, *E. coli* DH5α strain was used.

This procedure must be performed under sterile conditions. During the preparation of competent cells, all the solutions were chilled, pipette tips cooled and cells always kept on ice.

Preparation of competent cells which can be stored at -70°C

- E. coli cells (DH5α) were Streaked on an LB plate and allowed cells to grow at 37°C overnight
- One colony was placed in 5 ml LB media and grown overnight at 37°C
- 250 ml LB was inoculated with 1 ml of overnight culture of *E.coli* and grown at 37°C on a shaker till the $O.D._{600}$ was around 0.5.

- Culture was cooled on ice immediately and cells were harvested by centrifugation at 6000 rpm for 5 min at 4°C.
- Supernatant was removed carefully and traces of it were removed by inverting the centrifuge tube on paper towels
- However, the tube was never removed from ice for longer time. The bacterial pellet was resuspended in 50-70 ml of ice-cold 0.1M $CaCl_2$ (sterilized by filtration or autoclaving) and incubated on ice for 30 min.
- Cells were recovered by centrifugation as above, resuspended in 50-70 ml of 0.1M $MgCl_2$ and incubated on ice for 30 min.
- The cells were finally recovered and suspended in 10 ml of 0.1 $CaCl_2$ containing 10% glycerol.
- Aliquots of 200 µl each were prepared and immediately stored at -70°C

3.7 TRANSFORMATION

The process by which cells take up exogenous DNA from the outside is called transformation. The purpose of this technique is to introduce a foreign plasmid into bacteria and to use those bacteria to amplify the plasmid in order to make large quantities of it. Apparently, the treatment induces a transient state of "competence" in the recipient bacteria, during which they are able to take up DNAs derived from variety of sources. Many variation of this basic technique have since been described, all directed toward optimizing the efficiency of transformation of different bacteria strains by plasmid.

3.7.1 Transformation of DH5α strains of E. coli

- Frozen competent cells were thawed on ice for 15-20 minutes.
- To these competent cells, ligated product was added and mixed gently and the mixture was incubated on ice for 30 minutes.
- The cells were given heat shock at 42 °C for exactly 90 sec and immediately transferred to ice for at least 5 min.

- 800 µl of Luria broth was added and tube was incubated at 37^0C for 1 hour in a rotary shaker set at 200rpm.
- The cells were pelleted down and pelleted cells were resuspended in 100-150ul of LB.
- Transformed cells were pipetted onto LB agar plate containing appropriate antibiotics for selection of positive clones containing plasmids and spreaded uniformly using sterilized glass beads or bent glass rod spreader. For selection of the recombinant plasmids by blue-white selection, IPTG and X-gal were used in LB agar plate in appropriate amounts.
- Plates were incubated upside down at 37^0C for 12-16 h till colonies are big enough for transfer.

X-gal/IPTG: For each plate containing 20-25 ml media, 40µl of X-gal (20 mg/ml stock in dimethyl formamide) and 10µl of IPTG (20 mg/ml stock in ddH_2O) were plated before cell plating.

Concentration of antibiotics used: Ampiccilin-100 µg/ml in water; Kanamycin- 50 µg/ml in water

3.7.2 Selection of recombinant clones

Recombinant clones were selected by blue/white screening. Colonies that turned blue were left out while the colonies that remained white (recombinant clones) were transferred on to a fresh plate containing appropriate antibiotics. Plasmid DNA was then isolated from the white colonies as given below and confirmation of the recombinant clones was done by either/both of the two strategies. One strategy involves digestion with those restriction enzymes that have their sites at the flanking ends of the insert (in order to release the insert from the vector) and another strategy deals with digestion using *Pvu*II restriction enzyme. In pBluescriptII KS+ vector, *Pvu*II has a site at ~200 bp upstream/downstream of each end of the insert. Therefore, a self-ligated vector produces a band at ~400 bp after digestion with *Pvu*II unlike the recombinant clones (in most cases). This strategy was followed because viral genomic components and the vectors being used have approximately similar sizes and are difficult to separate on agarose gels.

3.8 PLASMID ISOLATION FROM TRANSFORMED CELLS

Transformed colonies were picked and inoculated on LA-ampicillin plates with the help of sterile toothpick and same was used for the inoculation in LB-Amp media. These plates and tubes were incubated overnight at 37^0C.

Principle

Bacterial plasmids DNA are widely used as cloning vehicles in recombinant research. Method denatures high molecular weight chromosomal DNA while covalently closed circular DNA (cccDNA) remains double stranded.

3.8.1 Boiling miniprep method (Holmes and Quingley 1981)

- 1.5 ml of the overnight grown culture was pelleted down in a microfuge by centrifugation at 14,000rpm for 30 sec.
- Supernatant was discarded and the pellet was resuspended in 110µl of STET buffer.
- Then freshly prepared 10.0µl of Lysozyme (10mg/ml stock in 10mM Tris pH 8.0) was added to the resuspended cells and thoroughly mixed and incubated at room temperature for 5 minutes.
- Heat shock was given for 42 seconds in boiling water followed by centrifugation at full speed for 20-min at room temperature.
- Pellet containing the cell debris was removed with the help of a sterile toothpick.
- DNA in the supernatant was precipitated by adding one volume of isopropanol (i.e. 120 µl).
- Precipitated DNA was collected immediately by centrifuging for 20 minutes at full speed (14,000 rpm).

- The supernatant was completely removed and the pellet was air dried to remove excess of isopropanol.
- DNA was resuspended in 50µl nuclease free water.

Stet buffer: 8% sucrose; 0.5% triton X-100; 50mM Tris HCl (pH 8) and 50 mM EDTA (pH 8). Add triton x 100 after autoclaving rest of components in solution.

3.8.2 Digestion of the plasmid DNA for checking positive clones:

In order to find out positive clones among the transformed colonies plasmid were digested either with enzyme that was used for cloning into vector or *Pvu*II as follows:

Digestion mixture:

CONTENT	QUANTITY (µl)
Plasmid DNA (~2µg)	10.0
Buffer G(Roche)	5.0
PvuII(10U/µl)	0.5
Water	34.5
Total	50.0

Procedure:

- The digestion mixture containing plasmid DNA, buffer G, enzyme *Pvu*II and water was thoroughly mixed and centrifuged briefly.
- Mixture was incubated at 37°C for 3 hours.
- 2.5 volumes of cold ethanol was added in each mixture and incubated at -80°C for 30 min.
- Precipitated nucleic acid was centrifuged at maximum speed for 20 minutes at 4°C.

- The supernatant was removed, pellet was dried and resuspended in RNase free water and then checked on 1% agarose gel.

3.8.3 Purification of plasmid DNA for sequencing

Sequencing requires ultra clean DNA and therefore plasmid DNA was re-extracted using commercially available plasmid isolation kits.

Plasmid isolation using commercial kits

Plasmid was also isolated using GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) as per the manufacturer's instructions.

3.9 Sequencing of viral DNA components

Plasmid of recombinant clones was isolated using either GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) as per the manufacturer's instructions. End sequences of the cloned viral DNA components were obtained by sequencing the recombinant plasmids using vector specific M13 universal primers. Three clones from the same transformation experiment were sequenced for each component. As complete begomoviral components are ~2.8 kb in size, these could not be completely sequenced using vector specific primers only. Vector specific primers were used to sequence both the strands. Complete sequences were obtained by primer walking strategy using self-designed specific primers (described in results section) or from 1st BASE Laboratories (Malaysia).

The sequencing PCR was performed as follows:

CONTENT	**QUANTITY (µl)**
Plasmid (200-500 ng)	1.0
5X sequencing buffer	1.0
Reaction mixture	1.0
Primer	0.5
dH_2O	1.5
Total	5

Sequencing PCR was performed by using following cycling parameters: 94°C for 10 sec; 25 cycles of 54°C for 40 sec and 60°C for 4 min. The sample was precipitated with ethanol, washed with injection buffer for removing surplus dyes and taken for sequencing by ABI PRISM^{TM}, 310 Genetic analyzer.

3.10 SEQUENCE ANALYSIS USING BIOINFORMATICS TOOLS

Percentage homology searches for obtained sequences were performed *using Basic Local Alignment Search Tool (BLAST; Altschul et al. 1997) available* at NCBI website (http://www.ncbi.nlm.nih.gov/blast). The pairwise sequence comparison of a new begomovirus isolate with all known sequences representative of species was carried out with the software MegAlign of DNAStar (Lasergene) using Clustal V algorithm. The program BLASTP was used to search the amino acid sequence database. Phylogenetic analysis was performed by neighbor-joining method selecting 1000 bootstrap replicates, using ClustalW program (Larkin *et al.* 2007) version 2.1 either available online (http://www.ebi.ac.uk/tools/clustalw2) or MEGA 5.0 software or DNASTAR Megalign software and trees were viewed and edited using TreeView software version 1.6.6 (Page, 1996). For recombination analysis, RDP, GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN and 3SEQ methods were used in the recombination detection program 3 (RDP3; Martin *et al.,* 2010). Highly divergent sequences (i.e., those sharing <60% genome-wide sequence identity to any other sequences in a data set) were discarded. Default settings were used throughout, and only potential recombination events detected by three or more of the above methods coupled with phylogenetic evidence of recombination were considered significant. Various representative sequences of distinct begomoviral species and alpha/betasatellites were used for phylogenetic and recombination analysis (described in results section). Online restriction maps of the sequences were constructed using either NEBcutter version 2.0 (Vincze *et al.*, 2003) available at http://tools.neb.com/NEBcutter2/ or webcutter version 2.0 (http://users.unimi.it/~camelot/tools/cut2.html) software.

3.11 CONSTRUCTION OF AGRO-INFECTIOUS CLONES

For transmission of most of the begomoviruses, it is necessary to construct dimers or partial dimers (also called bitmers) of the viral DNA components that harbour at least two origins of replication. These constructs can directly (without restriction) be rubbed onto plant surfaces, as the unit genome component will be released by replication and/or less efficiently by recombination. These constructs can be transferred to agrobacterial plasmids and delivered to plants by agro-infiltration for increased infectivity (Stenger et a., 1991). Agro-infectious clones of these new species were prepared in binary vector pCAMBIA-1300 (Cambia Labs, Canberra, Australia) as described below.

The construction of infectious clones was performed in two steps. In first step, the RCA product was double digested with *Xba*l and *Sal*l (as per Fermentas Double Digest Tool). This digestion generates a fragment of 2.5kb of DNA-A. This partial viral genomic component is then ligated into binary vector pCAMBIA 1300, which was previously double digested with same enzymes *Xba*l and *Sal*l, to get recombinant vector pCAM-OYParXS. In second step, this recombinant pCAM-OYParXS is then used to ligate a full length DNA-A component generated by *Xba*l, by cutting the RCA product, to get a clone pCAM-OYVM-A (fig.3.1). The orientation of the inserts was then monitored with end sequencing and restriction digestion with different enzymes that are having single restriction site in DNA-A, so that all these enzymes should yield 2.7kb full length fragment. The positive clones were selected that contain tandemly repeated viral genomes having 2 *ori* sites, which is a prerequisite for infectious clone construction.

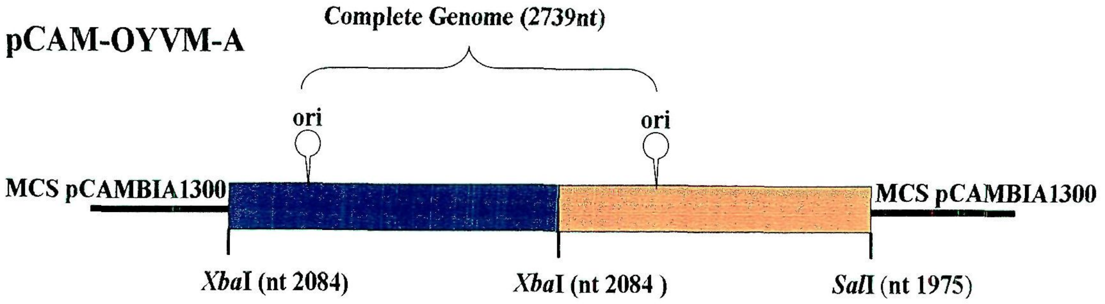


Figure 3.1. Schematic representation of infectious clones of okra yellow vein mosaic virus. 1.9 mer of viral genome cloned into binary vector pCAMBIA-1300 (ori: origin of replication; MCS: Multiple cloning sites)

3.11 *AGROBACTERIUM* TRANSFORMATION AND AGROINFECTION

Agrobacterium tumefaciens was transformed either by tri-parental mating or by freeze-thaw transformation of chemically competent cells.

3.11.1 Transformation of *Agrobacterium* competent cells by freeze-thaw method:

The competent cells were transformed with the recombinant pCAMBIA plasmids as follows:

(i) Competent cells were thawed on ice for 20 min.

(ii) 1 µg of the plasmid DNA was mixed with the competent cells.

(iii) Tubes were immediately put in liquid N_2 for 5 min for freezing.

(iv) After that the tubes were immediately put at 37°C for 25 min (heat shock).

(v) After the incubation, 1 ml of YEP medium (without any antibiotic) was added to the tubes.

(vi) Transformed cells were allowed to grow for 3 hrs at 28°C with continuous shaking (250 rpm) in a rotary shaker.

(vii) Cells were pelleted down by centrifugation at 12,000 rpm for 2 min.

(viii) Supernatant was discarded but approximately 150 µl supernatant was retained and cells were dispensed in it.

(ix) Cells were spread on YEPA plates containing streptomycin (100 µg/ml), rifampicin (25 µg/ml) and kanamycin (50 µg/ml).

(x) The plates were incubated at 28°C for 3-4 days.

3.11.2 Selection of transformed *Agrobacterium* by colony PCR

Colony PCR was performed to select the transformed *Agrobacterium* colonies using specific primers for different components. A single *Agrobacterium* colony was picked and placed it in 50 µl of sterile water. Boiled for 2 min and immediately placed on ice. PCR reaction was set up as follows:

105188

Boiled mixture	10 µl
10x Taq buffer A	2.5 µl
10mM dNTP mix	1 µl
Forward primer	0.5 µl (100 ng)
Reverse primer	0.5 µl (100 ng)
Taq DNA polymerase (Genei, Bangalore, India)	0.25 µl (3 units/µl)
ddH_2O to	25 µl

After completion, the reaction was directly electrophoresed in 1 % agarose gel to look for desired amplification.

3.11.3 Agro-infiltration of infectious clones

Agro-infiltration was performed by a method described by Zhou *et al.* (2003). Transformed *Agrobacterium* cultures were grown at 28°C for 48 hours (till OD_{600} =1) in 10 ml of YEP medium. The cultures (0.2-0.5 ml per plant) were injected into stems and petioles of the test plants at 4-6 leaf stage using 1 ml syringe and a fine needle. At least five test plants were inoculated with the inoculum. After agroinoculation, the plants were kept in an insect-free chamber at a constant temperature (25-28°C) under 14-16 h lighting. Plants were photographed 35 dpi (days after post inoculation).

3.11.4 Detection of viral DNA components in agro-inoculated plants

Viral components were detected from the systemically infected leaves of agro-inoculated plants by southern blotting and PCR detection.

3.11.4.1 Southern blotting and hybridization

Localization of particular sequences within genomic DNA is usually accomplished by the transfer techniques described by Southern (1975). The DNA-DNA hybridization forms its basis.

i. Blot preparation

The DNA was isolated from systemically infected leaf tissue of agro-inoculated plants by modified CTAB method and treated with RNase A (as described before) for detecting viral DNA components. The high concentration of DNA (~10 µg) was electrophoresed in 0.7 % agarose gel and transferred to the positively charged nylon membrane by capillary transfer method. After completion of electrophoresis, the DNA in the gel was denatured by placing the gel in a plastic tray containing the denaturation solution with slow shaking on a gel rocker for 30 min. After denaturation, the gel was soaked in neutralizing solution for 30 min with slow shaking. The two chambers of the electrophoresis tank were filled with transfer solution (10x SSC). A wick of Whatman® 3MM paper was soaked in 10x SSC and placed over the gel platform of the tank with its ends submerged in the solution. Three more sheets of Whatman® 3MM paper (slightly bigger than the gel) were soaked in 10x SSC and placed over the wick. The gel was carefully inverted and placed over the sheets avoiding any air bubbles between the gel and Whatman® papers. Nylon membrane was cut to the gel size, wetted in sterile water and placed over the gel avoiding any air bubbles. Three more sheets of Whatman® 3MM paper were cut to gel size, soaked in 10x SSC and placed over the membrane. A stack of gel-sized dry blotting papers (8-10 cm height) was placed over the 3MM papers and approximately 500 g weight was placed on top of the stack. The transfer was allowed to take place for 14-16 h.

After completion of the transfer, the blotting paper stack was removed, the membrane was marked so as to indicate the orientation and rinsed in 2x SSC briefly. The buffer was removed by placing the membrane between the folds of a dry 3MM sheet but the membrane was not completely dried. The membrane was wrapped in a cling film and placed under UV light in a UV cross-linker for 2 min for cross-linking the DNA to the membrane. The blot was then stored at 4°C until hybridization.

ii. Probe preparation

Partial CP of ToLCNDV cloned in pGEM®-T Easy vector was digested with *Eco*RI restriction enzyme and electrophoresed in 1% agarose gel. Viral DNA fragment was excised from the gel, purified and then incubated in boiling water bath for 10 min, for probe preparation. The following reaction was used:

Denatured DNA	200-500 ng
Random Primer	100 ng
10x Klenow buffer	3 µl
dNTP mix (-CTP)(3.3nM each)	4.5 µl
α-^{32}P dCTP (10µCi/µl,specific activity 3x10^3 Ci/mmole)	10 µCi
Klenow enzyme	5 units
Water to	30 µl

The reaction was incubated at 37°C for one hour. To this equal volume of Buffer-A was added. The mixture was then denatured by incubating the tube in boiling water bath for 7 min. The contents were immediately transferred to ice before adding to the hybridization bottle.

dNTP mix: (for α^{32}-P dCTP as the radioactive molecule, 100mM stock of dATP, dTTP and dGTP): 1+1+1+27µl water. 4.5µl of this mix was used for one reaction.

20x SSC buffer: 3M NaCl and 0.3 M Trisodium citrate

Buffer-A: 500mM Tris HCL (pH 7.5), 500mM NaCl, 5mM EDTA and 0.5% SDS.

TNE Buffer: Tris HCl- 20mM, NaCl- 10mM and EDTA- 10mM

iii. Pre-hybridization/Hybridization

Pre-hybridization was carried out for 1 hr at 42°C in the pre-hybridization buffer. Pre-hybridization buffer comprised of:

Formamide	50 % (v/v)
Na_2HPO_4 pH 7.2	120 mM
NaCl	250 mM
EDTA, pH 8.0	1 mM
SDS (w/v)	7 %

After pre-hybridization, probe was added to fresh pre-hybridization solution and was incubated at 42°C overnight (10-18 hrs).

iv. Washings of Blots

After hybridization was complete the blots were removed from the hybridization solution, rinsed briefly in 2x SSC and placed in the first washing buffer and was followed by 2^{nd} and third wash with progressive stringency as follows:

First Wash Buffer	2.0x SSC, 0.1% SDS (25°C) - two washings of 15 min each
Second wash buffer	0.5x SSC, 0.1% SDS (25°C) - two washings of 15 min each
Third Wash buffer	0.1x SSC 0.1% SDS (65°C) - two washings of 30 min each

After washing, the blots were placed on a Whatman filter paper to remove excess of liquid and wrapped in a saran wrap immediately, to prevent it from drying and for the purpose of autoradiography.

v. Autoradiography

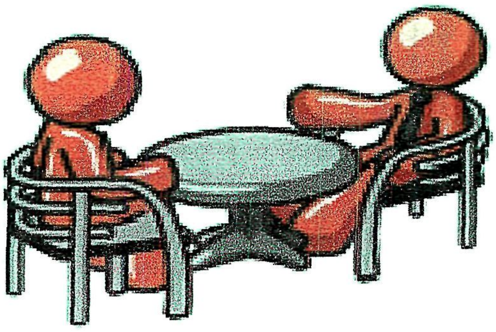
The blots were exposed to X-ray films (X-OMAT, XK-5, and Kodak, India) by putting them between two intensifying screens (Kiran Hi-speed, India) for the autoradiography in a cassette at -80°C. The films were developed after sufficient exposure, depending on the intensity of the signal prior to autoradiography as determined by a portable radioactivity monitor.

3.11.4.2 PCR detection

The DNA was isolated from leaves of agoinoculated plant using modified CTAB method and the PCR was carried out using the AV494 and AC 1048 degenerate primers specific to viral DNA, which was agroinoculated into plants. The pattern of thermal cycling for each primer pair is also summarized as described earlier (Table 3.1).

3.12 C, N, H, S ANALYSIS OF AGROINOCULATED AND NON-INOCULATED PLANTS

The percentage amount of C, N, H, and S in plant samples were carried out by using an instrument of CHNS analyzer (Elementar, *vario micro).* The fresh leaf and fruit samples (agroinoculated and healthy) were kept in an oven at 45°C until weight loss became constant. Dry plant tissues were crushed in a mortar and pestle to a very fine powder. Thereafter, about 10-15 mg of each plant samples were mixed with tungsten oxide (3-4 mg) as a catalyst and packed in a very thin tin board. Similar procedure was repeated for all the plant samples. The packed sample boats were kept in sample holder of instrument for analysis of C, N, H, and S. On placing the sample boats in holder, instrument was started and proper entry of parameters like weight of blank, standard and samples were done in a computer attached with the instrument. Sulfanilamide was used as a standard, which has a fixed percentage amount of C (41.8), N (16.25), hydrogen (4.65) and sulfur (18.62). After putting the weight of all the above said parameters, instrument was started and left for the analysis. A statistical factor was applied to deduct the blank from each plant sample.



Results and Discussion

4. RESULTS AND DISCUSSION

In Himachal Pradesh, okra is mainly grown in sub-temperate regions. Survey was carried out to collect samples from okra growing areas *viz.* Kangra and Bilaspur districts of Himachal Pradesh and also from Solapur (Maharashtra) and Jammu (Jammu & Kashmir) (Fig. 4.1). Samples were collected based on symptomatic observations. The collected okra samples had virus like symptoms showing vein clearing mosaic pattern (Plate 4.1), indicating the presence of viral infection on leaves. The samples were classified into 6 groups according to their collection location as follows:

1) Punder (Palampur); 2) Palampur; 3) Nurpur; 4) Bilaspur; 5) Jammu; 6) Solapur

These samples were then processed further to detect the presence of virus and to amplify its complete genome. The complete genome was sequenced and analyzed through various bioinformatics tools. Further, infectious clone of the virus was constructed to characterize infectivity of the virus. Results obtained on different aspects of the study have been presented under the following heads.

4.1 DNA ISOLATION FROM OKRA

Okra contains the high amounts of polysaccharides and phenolic compounds, a major problem in the purification of pure plant DNA. When cells are disrupted, these cytoplasmic compounds can come into contact with nuclei and other organelles (Loomis 1974). In their oxidized forms, polyphenols covalently bind to DNA giving it a brown color and making it useless for most research applications (Katterman and Shattuck 1983; Guillemaut and Maréchal-Drouard 1992; our observations). So the modified CTAB DNA extraction protocol was developed to get the good quality DNA from okra. The quality of the extracts was measured using a NanoDrop® ND-1000 UV–Vis Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and visualized by electrophoresis through 0.7% agarose gels (Plate 4.2). The average ratios of spectrophotometric absorbance A260/A280 were observed in range of 1.83-1.97

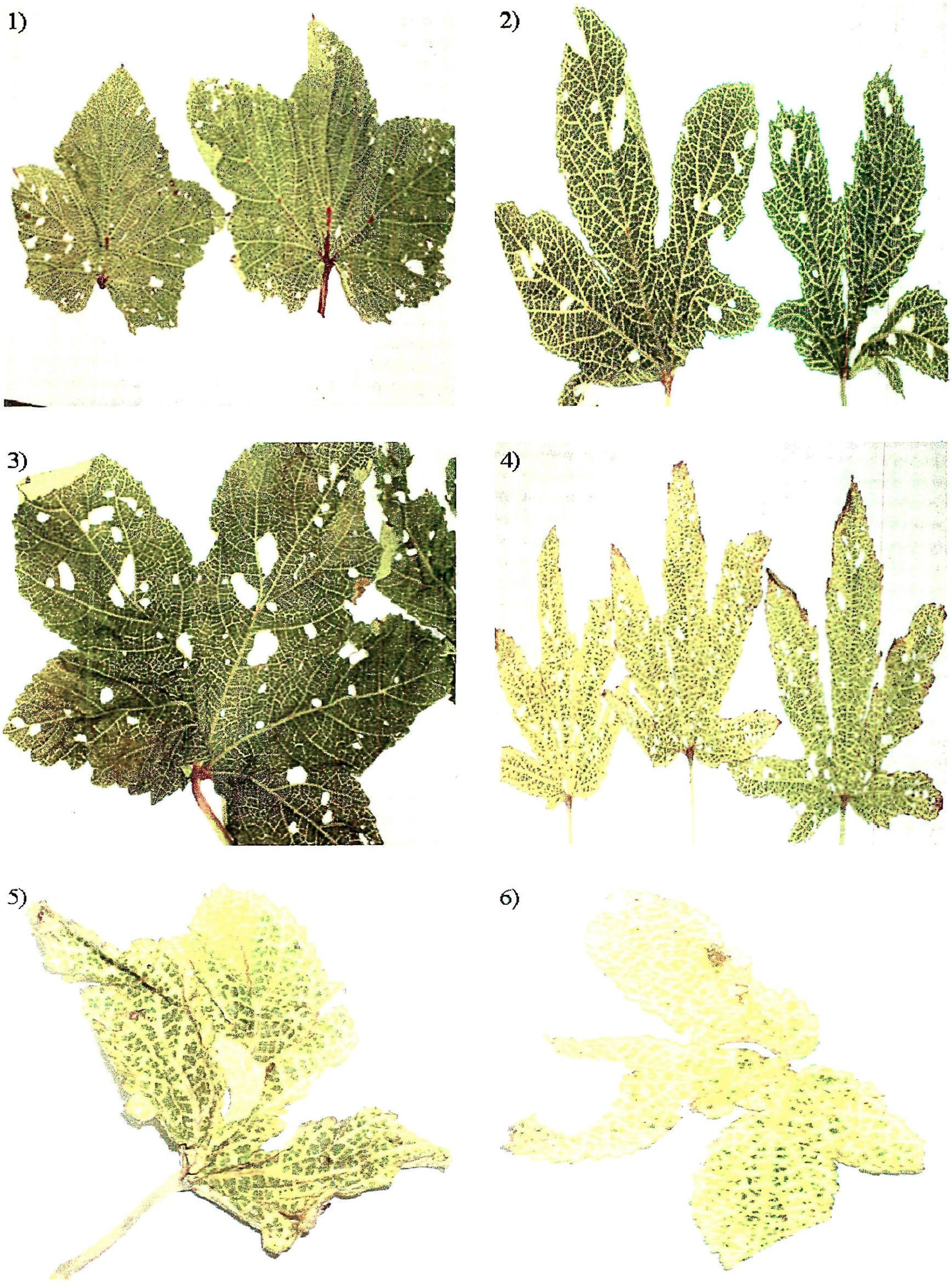


Plate 4.1. Symptoms of yellow vein mosaic disease. Samples were collected from Punder (1), Palampur (2), Nurpur (3), Bilaspur (4), Jammu (5) and Solapur (6)

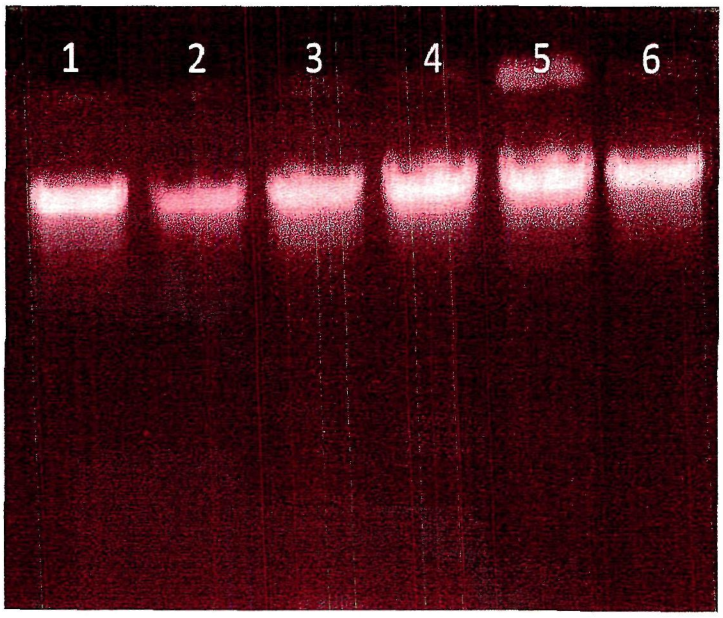


Plate 4.2. Quality of isolated DNA from okra samples: Lanes 1-6 contain samples from Punder, Palampur, Nurpur, Bilaspur, Jammu and Solapur

in all samples. The yield of DNA ranged from 20-40 µg per 100mg of leaf sample. The DNA isolated from this protocol was of good quality and the presence of virus was successfully detected by PCR and Rolling Circle Amplification (RCA) techniques.

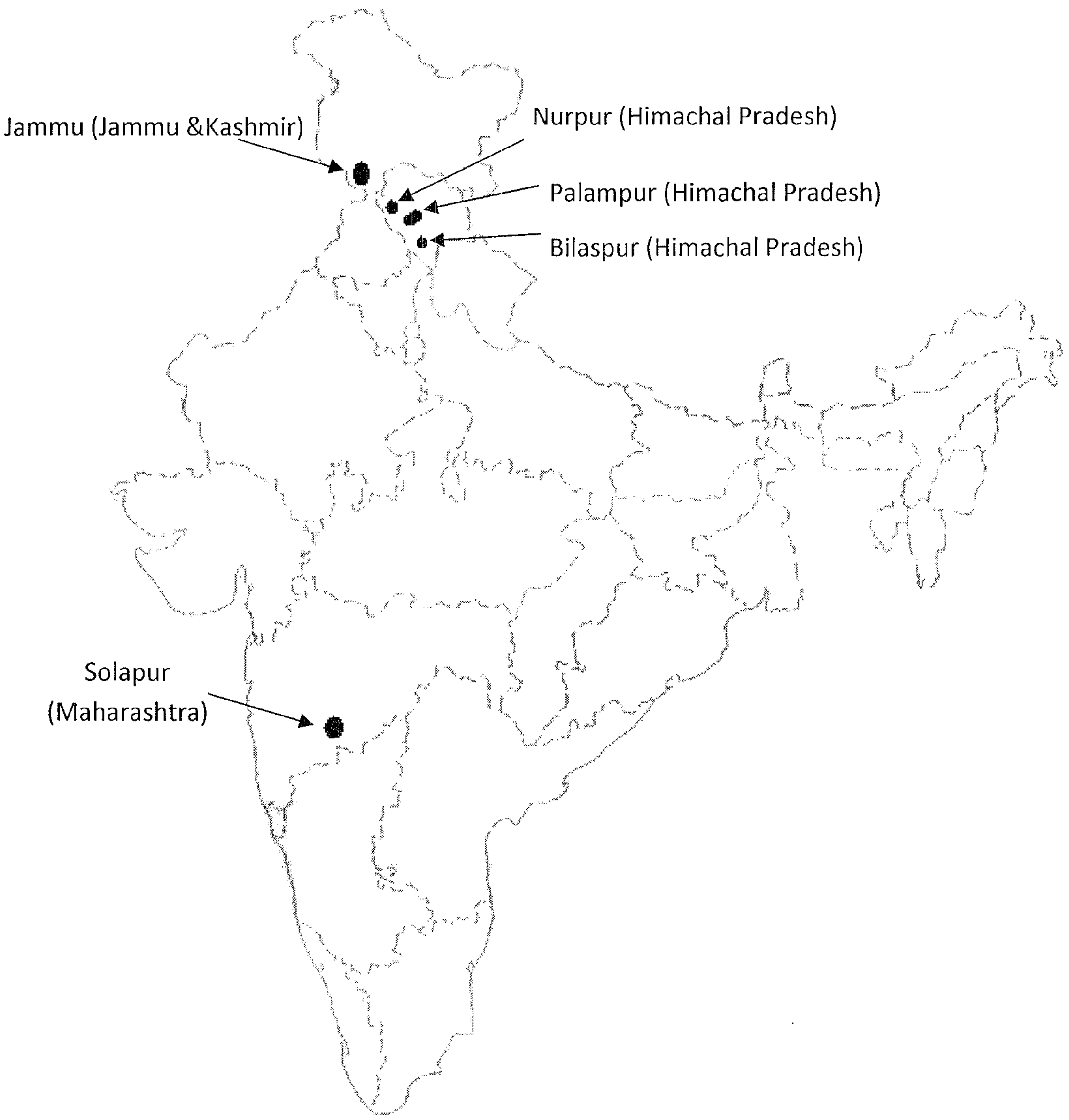


Figure 4.1. Map of India showing the different locations of sample collection

4.2 DETECTION OF VIRUS

Different molecular techniques like PCR and RCA were used for detection of viral DNA in samples. These techniques confirmed the presence of virus in collected leaf samples.

4.2.1 Polymerase chain reaction (PCR) based detection

Out of six different samples, 3 samples were found positive in primary detection PCR by using degenerate primers AV494 and AC1048 (Wyatt and Brown 1996) which amplifies ~570 bp DNA fragment of a geminivirus (Plate 4.3). The results suggested presence of geminivirus in three samples namely Nurpur, Palampur and Punder and all other three may be negative samples. So the further confirmation of positive virus carrying samples was carried out by using RCA.

4.2.2 RCA based detection

The positive samples detected by PCR were further confirmed by RCA. RCA also gave same result as that of PCR, confirming the presence of virus in 3 samples. The high molecular weight band of RCA amplified concatamers was detected in 3 samples (Plate 4.4).

4.2.3 RCA/Restriction fragment length polymorphism (RFLP)

RCA products were subjected to RFLP to check whether the 3 positive samples were having the same viral strains or not. Further, this will also help us identify if some of the samples contain new isolate of geminivirus infecting okra. The RCA products were digested by 4bp cutter (*Hpa*II) and then checked on 2% agarose gel. All samples showed different profile (Plate 4.5) suggesting that the samples were carrying different viral strains. Out of these, sample from Punder was then selected for further characterization.

These findings suggests that RCA/RFLP is most efficient method for detection of geminiviruses from plant samples as it is highly reproducible, independent of viral genome organization, source plant type and origin and sample preparation. Also neither specific primers are needed to start the

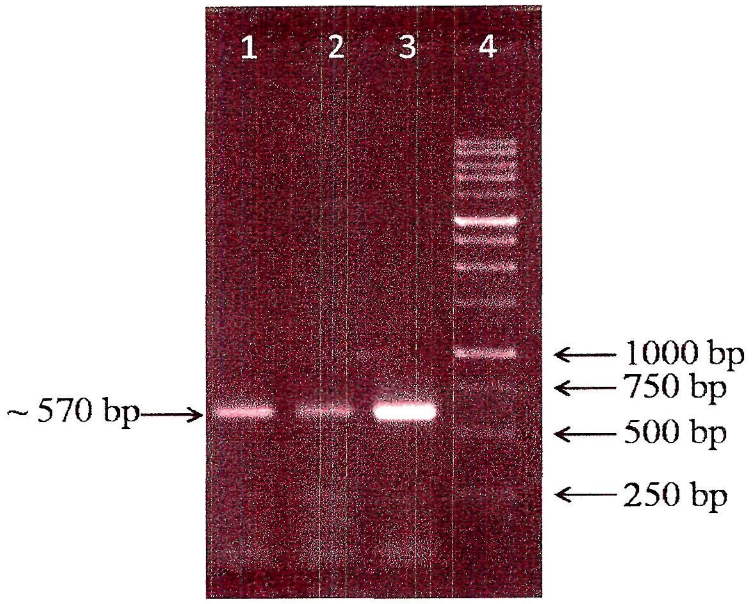


Plate 4.3. Detection of geminivirus by PCR using AV494 and AC1048 degenerate primers. Lanes 1-3 contain samples from Nurpur, Palampur and Punder; Lane 4: 1 kb marker

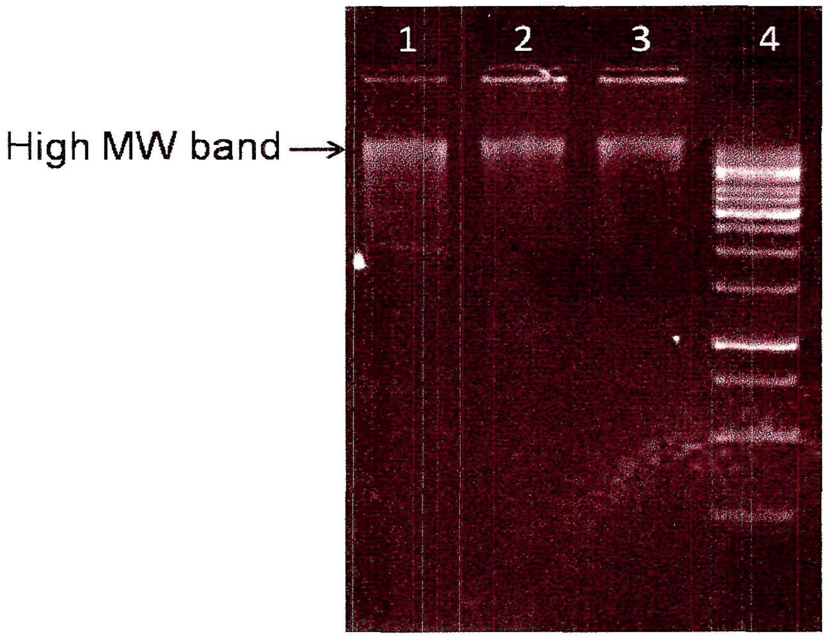


Plate 4.4. Detection of a geminivirus by RCA. Lanes 1-3 contains samples from Nurpur, Palampur and Punder; Lane 4: 1 kb marker

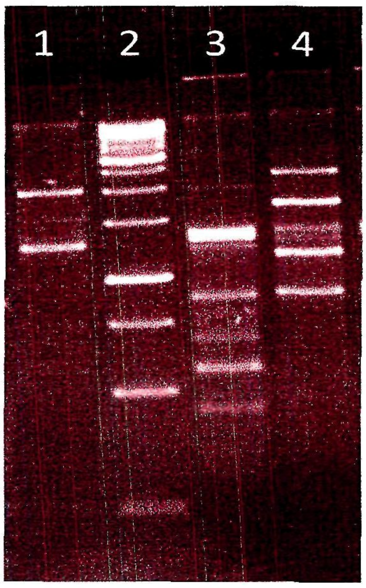


Plate 4.5. RCA/RFLP of positive samples. Lanes 1, 3-4 contains samples from Punder, Palampur, and Nurpur respectively; Lane 2: 1kb marker

reaction, nor expensive equipment like a thermocycler. This diagnostic system using RCA combined with restriction enzyme analysis has been previously established by Haible *et al.* (2006) for the detection and differentiation of known geminivirus species.

4.3 AMPLIFICATION AND CLONING OF COMPLETE VIRAL GENOME

The full genome of the virus from selected sample was amplified using RCA. RCA gave the high molecular weight products that were long concatameric genome molecules of double-stranded DNA. For cloning a complete viral DNA component, the RCA product was first digested with several enzymes to identify a unique restriction enzyme site in the entire component. *Hind*III, *Xba*I and *Sac*I were found to produce fragments of 2.7 kb while *Kpn*I was found to produce fragments of 1.3 kb (Plate 4.6). Hence out of these enzymes *Hind*III and *Kpn*I were selected for cloning of complete putative DNA-A and betasatellite fragment, respectively in pBluescript (KS$^+$) vector (Stratagene, La Jolla California, USA).

4.4 SEQUENCING OF COMPLETE VIRAL GENOMIC COMPONENTS

The *Hind*III and *Kpn*I cloned fragments (as mentioned above) were first end sequenced to confirm their identity. End sequences of both *Hind*III and *Kpn*I cloned fragments in BLAST search were found to be aligning with begomoviral DNA-A and betasatellite sequences, respectively. This confirmed their viral genetic nature and then the complete sequencing of both the fragments was carried out by primer walking. For complete sequencing of DNA-A component, two universal primers and two designed primers were used (Table 4.1). The universal primers were selected as per the vector, pBluescript KS+, and the two primers were designed from the end sequencing reaction obtained after sequencing by the universal primers so as to get complete sequence of DNA-A. Similarly for betasatellite also same universal primers were used and one designed primer to get the complete sequence of betasatellite (Table 4.1).

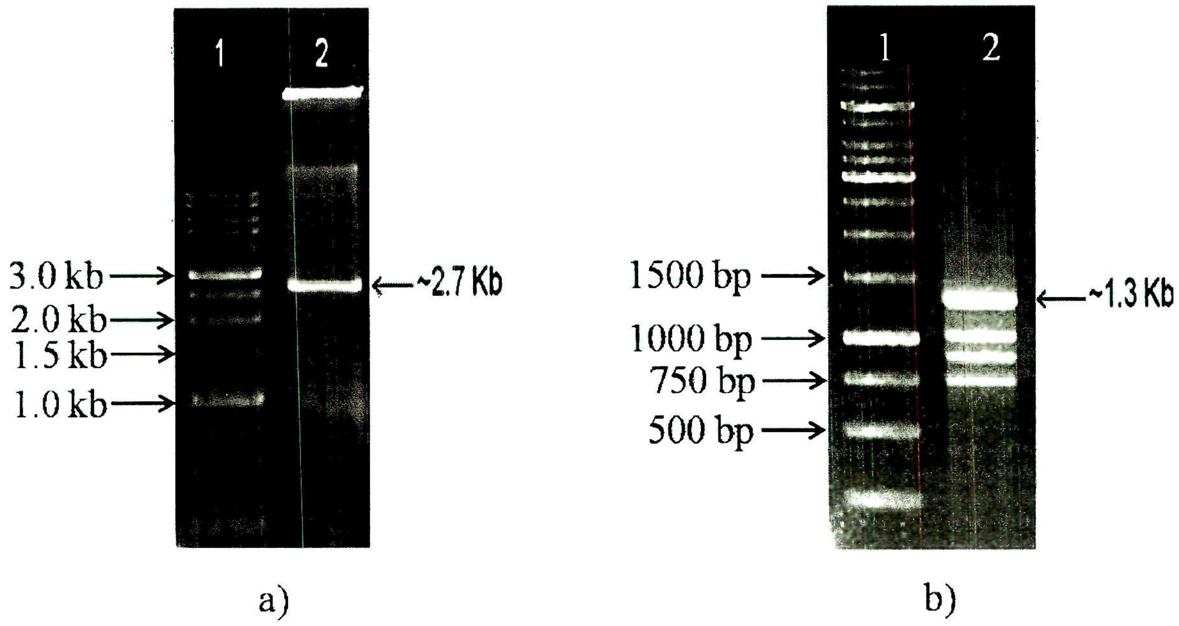


Plate 4.6. Restriction digestion of RCA product to get complete genome components. (a) Lane 1: 1kb marker, 2: *Hind*III digest; (b) Lane 1: 1kb marker, 2: *Kpn*I digest

Table 4.1 Primeres used for complete genome sequencing of present begomovirus

Sr. no.	Primer	DNA	Sequence
1	M13F (-20)	DNA-A and betasatellite	5' -GTAAAACGACGGCCAGT -3'
2	M13R-pUC (-26)	DNA-A and betasatellite	5' CAGGAAACAGCTATGAC 3'
3	BYVMVSeqF	DNA-A	5' GGT CAA GAA GTT TGT TAG GGT 3'
4	BYVMVSeqR	DNA-A	5' ATC TTC ACG TGC TCA TCC AG 3'
5	OYVMV-B For	Betasatellite	5' TTACTGGGGTTCAAAAGCAG 3'

The sequences obtained after sequencing were assembled into complete DNA-A and betasatellite sequences by using SeqEdit software of DNASTAR. The complete sequence of DNA-A was determined to be 2739 bases (Appendix I-1A) and that of betasatellite was determined to be 1332 bases in length (Appendix I-1B). The both DNA-A and betasatellite sequences were submitted to EMBL database and accession number FR694925 and FR823510 were obtained, respectively.

It is for the first time that a monopartite begomovirus species causing disease in okra was detected from sub-temperate region of Himachal Pradesh. The viral species is also found to be associated with a betasatellite and no DNA-B component was reported to be associated. These findings were similar to that reported earlier by Zhou *et al.* (1998) and Jose and Usha (2003). Zhou *et al.* (1998) also earlier determined the complete nucleotide sequence of OYVMV and reported it as a monopartite begomovirus but no betasatellite was reported to be associated with the virus. But later, Jose and Usha (2003) determined complete nucleotide sequence of BYVMV and reported that the betasatellite DNA molecule was associated with the virus.

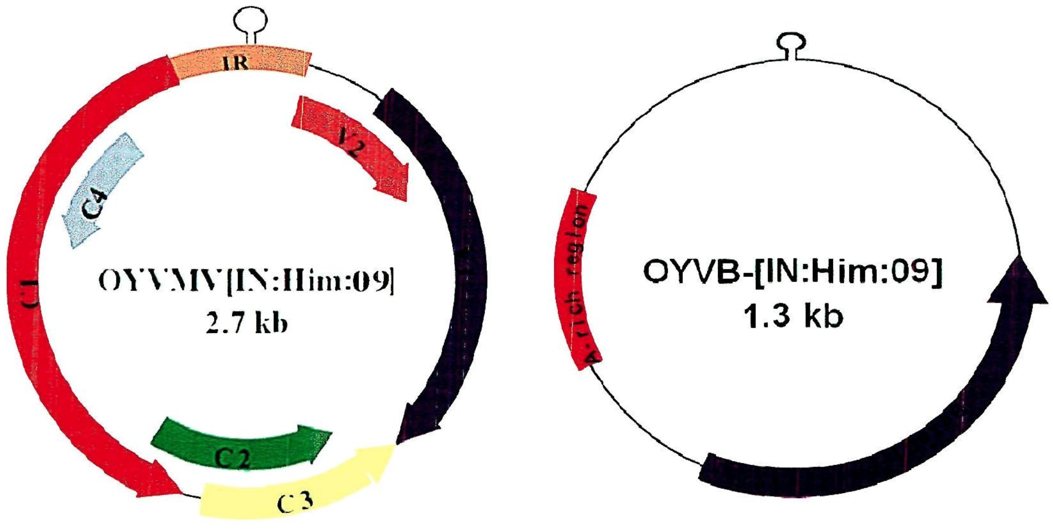
4.5 GENOME ORGANIZATION

4.5.1 Genome organization of DNA-A

The DNA-A of the virus was then searched for presence of different ORFs in the sequence using Expasy translation tool. Six different ORFs were found to be present on both virion and complementary strand. The position and orientation of all the six ORFs was then mapped on the begomovirus genome (Plate 4.7a). These ORFs were found to be separated by one intergenic region (IR). Out of six different ORFs, four ORFs were found in complementary sense (C1, C2, C3, and C4 ORF) and two were found in virion sense (V1 and V2 ORF). The C1 ORF was found to be of longest and C4 ORF was the shortest. The various features of the proteins were predicted by Protean analysis software of DNASTAR (Table 4.2). The sequences and the corresponding accession numbers of these ORFs is given in Appendix I-2.

Table 4.2. Features of OYVMV-[IN:Him:09] DNA-A ORFs

ORF	Position	Amino acid residues	Predicted structural class	Predicted Molecular weight (kDa)
V1	281-1051	256	Alpha	29.66
V2	121-486	121	Alpha	14.05
C1	1500-2591	363	Alpha	41.10
C2	1151-1603	150	Alpha	17.26
C3	1054-1458	134	Alpha	15.81
C4	2132-2434	100	Irregular (R)	10.92



a) b)

Plate 4.7. Genome map of (a) OYVMV-[IN:Him:09] and (b) OYVB-[IN:Him:09]. The ORFs are denoted according to their orientation as V (virion-sense) or C (complementary-sense). The direction of the arrow represents orientation of expression.

In addition to these six ORFs, one IR of 268 nucleotide (nt) was also found in the genome between the start codons of ORFs V2 and C1. The IR was found to have a conserved nonanucleotide sequence TAATATTAC that forms a stem-loop structure and contains a nick site *Ori* between the 7th and 8th nucleotide for initiation of replication of viral sense strand. The IR of present begomovirus has iterons or Rep binding sequence- CGGTACCC. These iterons are tandemly repeated and are present in two copies (Fig. 4.2). The first copy is present at 121nt upstream of *Ori* at the 5' end of IR and the second copy is located just immediately before TATA motif, which is located at 78 nt upstream of *Ori*.

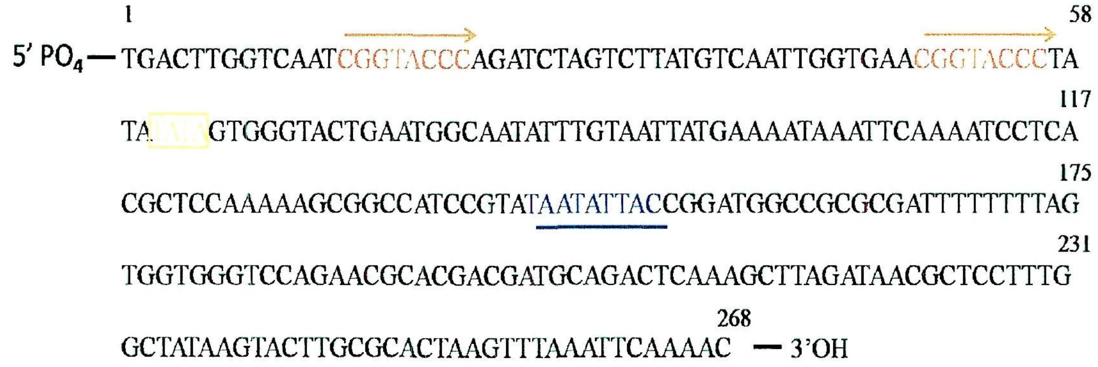


Figure 4.2. Arrangement of iterons in IR. The Ori of replication was shown by underline, the TATA motif was shown by box and iterons are shown by overhead arrows.

The genome organization was found as that of Old World begomoviruses (Padidam *et al.* 1999; Paximadis *et al.* 1999; Rybicki 1994). The AV2 ORF was also found to be present in DNA-A which was reported to be absent in New World begomoviruses (Rybicki 1994; Stanley *et al.* 2005). The length of genome was determined to be 2739 bases which is a typical length of all monopartite begomoviruses (Stanley *et al.* 2005). The IR contained a conserved nonanucleotide sequence TAATATTAC, which was typical to that of all other geminiviruses reported earlier. IR also found to have one iteron upstream of TATA motif. The iteron sequence (CGGTACCC) and its arrangement is similar to that of OYVMV-[PK:Fai201:95], as described earlier by Zhou *et al.* (1998).

4.5.2 Genome organization of betasatellite

The betasatellite of present isolate was searched for different ORFs and only one coding ORF (BetaC1) was found in complementary sense. The various features were predicted by Protean analysis software of DNASTAR (Table 4.3). The sequence and accession number of the proteins are given in Appendix I-2. The present betasatellite, OYVB-[IN:Him:09] also have an Adenine rich (A-rich) region of 181 bases in length between nucleotide position 902 to 1082 (Plate 4.7b) and has 62% A content while the overall A content of betasatellite molecules is 34%. This region has repeated blocks of up to 8 consecutive A residues. These findings were found similar to that reported earlier by Briddon *et al.* (2003). But the OYVB-[IN:Him:09] sequence had no satellite conserved region (SCR), which was previously reported to be present in betasatellites (Briddon *et al.* 2003, Zhou *et al.* 2003). A putative stem-loop structure containing a nonanucleotide TAATATTAC sequence typical of geminiviruses is present in OYVB-[IN:Him:09].

Table 4.3. Features of OYVB-[IN:Him:09] ORFs

ORF	Position	No. of amino acids	Predicted structural class	Predicted Molecular weight (kDa)
BetaC1	314-754	146	Alpha	13.7

4.6 RESTRICTION MAPPING OF GENOME

4.6.1 Restriction mapping of the viral genome

Sequence analysis showed that the viral genome had 51 unique restriction enzyme sites (Plate 4.8). However, the C3 ORF was found to have only four restriction sites, while ORF C1 was found to have maximum number (25) of restriction sites.

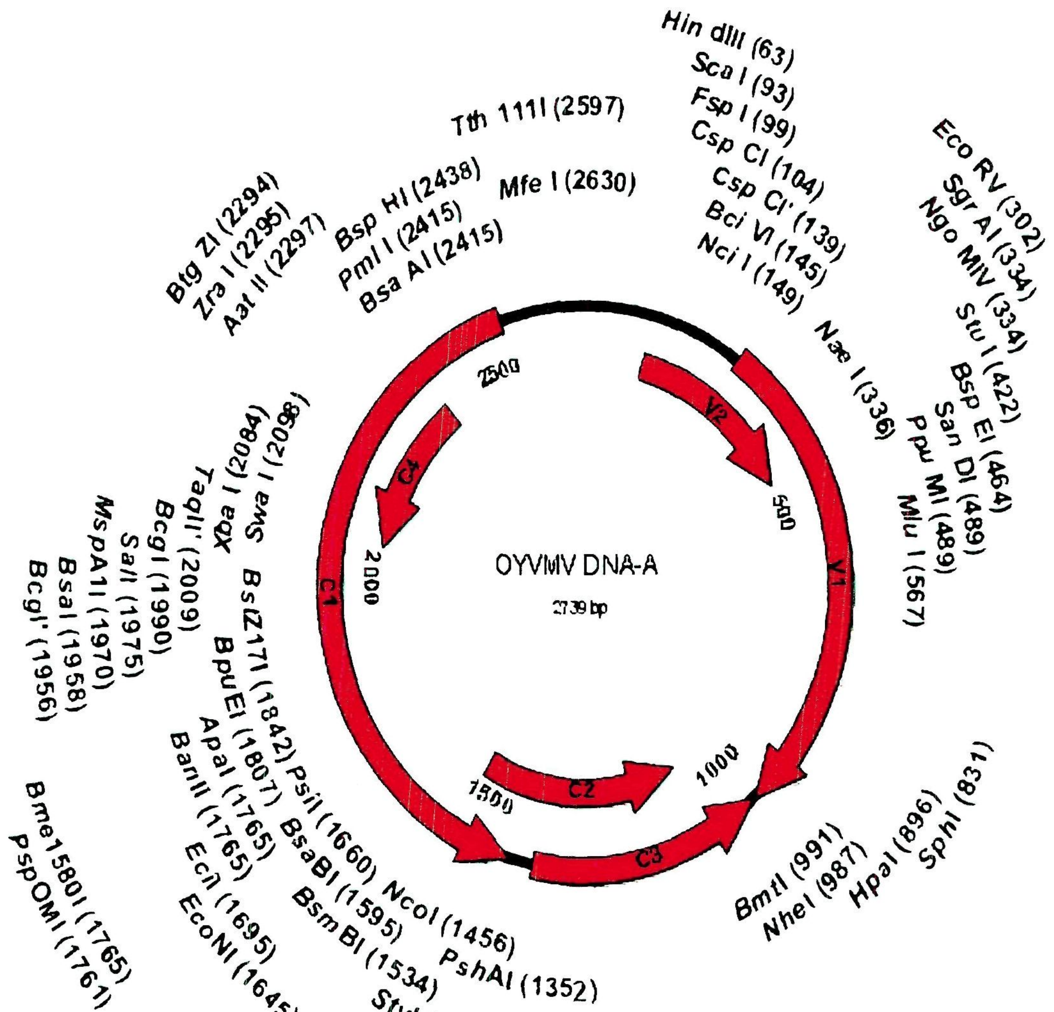


Plate 4.8. Restriction map of OYVMV DNA-A. The restriction sites of the corresponding enzyme were shown by a vertical line and the position of cutting site is shown in brackets.

4.6.2 Restriction mapping of betasatellite

The restriction map of betasatellite showed 37 unique restriction sites and out of that only 10 restriction sites are present in BetaC1 ORF. The region from Ori to start codon of BetaC1 has less frequency of restriction sites and only 9 restriction sites were detected. Similarly, the 200 nts downstream region (755-955 bases) of ORF BetaC1 was also found to have only 3 restriction sites (Plate 4.9).

4.7 SEQUENCE COMPARISON WITH OTHER BEGOMOVIRUSES

4.7.1 Sequence Comparison of DNA-A with corresponding begomoviral components

The nucleotide sequence of present begomovirus genome was compared to the complete DNA-A sequences of several other closely related begomoviruses available in GenBank database (Table 4.4). The pair-wise sequence comparison of the present begomovirus isolate with all other known sequences representative of species was carried out using ClustalV algorithm (DNASTAR) to know whether the present begomovirus belongs to a new species or is a strain or variant of a known species. The present begomovirus showed maximum nucleotide identity (96%) with that of OYVMV-[PK:Fai201:95], suggesting that it is a strain of OYVMV-[PK:Fai201:95] (Fauquet *et al.* 2008). Hence for present begomovirus OYVMV-[IN:Him:09] name was proposed as per the guidelines given by Fauquet *et al.* (2008). Among Indian isolates OYVMV-[IN:Him:09] showed maximum identity with BYVMV-[IN:OY177:09] (93%) followed by BYVMV[IN:Thd:05] (89%) isolates of okra. All other Indian begomovirus isolates showed less than 89% nucleotide identity (Table 4.5). The OYVMV-[IN:Him:09] also showed 90% maximum identity with begomovirus isolate of cotton, CLCuMV[IN:Lud:04].

The multiple alignment of OYVMV-[IN:Him:09] with several other begomovirus complete DNA-A sequences showed that the sequence was conserved in most of the coding regions but contained variability in IR but the region of *Ori* including stem-loop structure containing the nonanucleotide TAATATTAC sequence was found to be highly conserved in most of the begomoviruses (Appendix II).

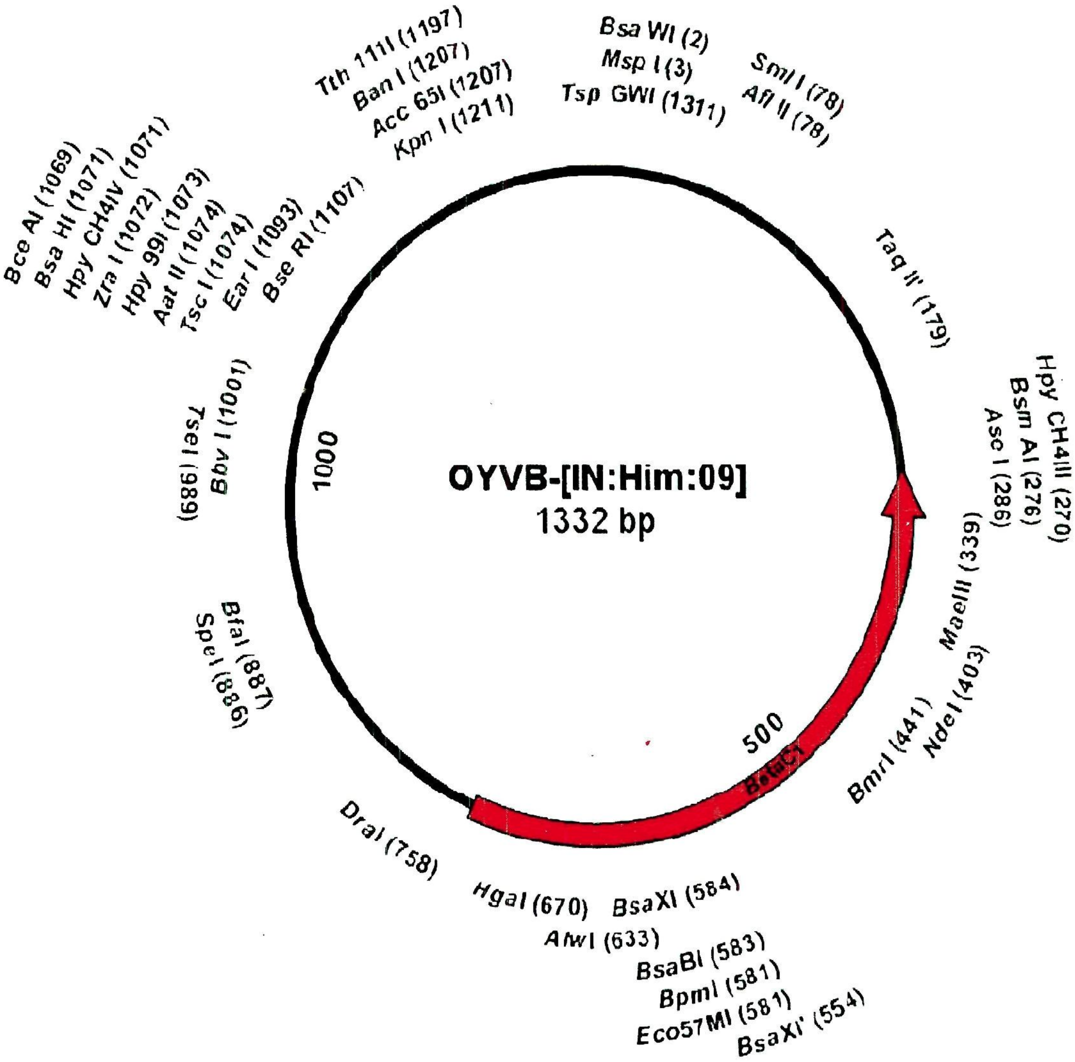


Plate 4.9. Restriction map of OYVB-[IN:Him:09]. The restriction sites of the corresponding enzyme were shown by a vertical line and the position of cutting site is shown in brackets.

4.7.2 Sequence comparison of betasatellite with other reported begomoviral betasatellites:

The nucleotide sequence of present betasatellite was compared using ClustalV algorithm to the complete betasatellite sequences of several other closely related betasatellite isolates available in GenBank database. The present betasatellite was found to be highly variable and showed only 65% maximum nucleotide identity with *Okra leaf curl betasatellite* (OLCuB-[PK:Goj:97]) which is below the threshold level of 89 % for species demarcation for a distinct betasatellite (Briddon *et al.* 2008). This data showed that the present betasatellite is a distinct satellite, which is associated with present OYVMV-[IN:Him:09] infecting okra from Himachal Pradesh (Northern India). Hence the name *Okra yellow vein betasatellite* (OYVB-[IN:Him:09]) was proposed as per the guidelines given by Briddon *et al.* (2008). Among Indian betasatellite isolates, BYVB-[IN:Gun:OY112:06] showed maximum nucleotide identity of 61%. All other isolates showed less than 60% identity (Table 4.5).

```
OYVMV-[IN_HIM_09]_FR694925      TGACTTGGTCAATCGGTACCC-AGATCTAGTCTTATGTCAATTGGTGAAC 49
OYVB-[IN_HIM_09]_FR823510       TGACTTGGTCAATCGGTACCCTATATATAGT---GGGT--ACTGAATGGC 45
                                ********************* * ** ****     **  * **     *

OYVMV-[IN_HIM_09]_FR694925      GGTACCCTATATATAGTGGGTACTGAATGGCAATATTTGTAATTATGAAA 99
OYVB-[IN_HIM_09]_FR823510       AATATTGTAATTATA-TAAATA--AAATCAAAATCCTCACGCTCCAAAAA 92
                                  **   **  **** *   **   ***   ***  *     *    ***

OYVMV-[IN_HIM_09]_FR694925      ATAAATTCAAAATCCTCACGCTCCAAAAAGCGGCCATCCGTATAATATTA 149
OYVB-[IN_HIM_09]_FR823510       GTAAAATCAAAATCCTCACGCTCCAAAAAGCGGCCATCCGTATAATATTA 142
                                 **** ********************************************

OYVMV-[IN_HIM_09]_FR694925      CCGGATGGCCGCGCGATTTTTTTTAGTGGTGGGTCCAGAACGCACGACGA 199
OYVB-[IN_HIM_09]_FR823510       CCGGATGGCCGCGCGATTTTTTTAAGTGGTTGGTCCAGATCTCGCG-CGA 191
                                *********************** ****** ******** * * ** ***

OYVMV-[IN_HIM_09]_FR694925      T 200
OYVB-[IN_HIM_09]_FR823510       T 192
                                *
```

Figure 4.3. The sequence alignment of part of IR of OYVMV-[IN:Him:09] with SCR replaced region of OYVB-[IN:Him:09]. The conserved sequence is shown by '*'.

Table 4.4. The different begomovirus sequences that were used for comparisons with the present sequence

Name of the Virus	Abbreviation	Accession
Okra yellow vein mosaic virus-[Pakistan:Faisalabad201:1995]	OYVMV-[PK:Fai201:95]	AJ002451
Bhendi yellow vein mosaic virus-[India:OY177:2007]	BYVMV-[IN:OY177:07]	FJ179371
Cotton leaf curl Multan viru -[India:Ludhiana:2004]	CLCuMV-[IN:Lud:04]	AY765257
Gossypium darwinii symptomless virus-[Pakistan:Dar1:2007]	GDSV-[PK-Dar1:07]	EU365613
Bhendi yellow vein mosaic virus-[India:Pandaralli:2005]	BYVMV-[IN:Pan:05]	FJ176236
Bhendi yellow vein mosaic virus-[India:Thadagan:2005]	BYVMV-[IN:Thd:05]	FJ176235
Okra leaf curl India virus -[India:New Delhi: 2008]	OkLCuIV-[IN:ND:08]	GQ245760
Cotton leaf curl Multan virus-*Hisar*[Pakistan:Faisalabad3:99]	CLCuMV-His[PK:Fai3:99]	AJ132430
Bhendi yellow vein mosaic virus-Pakistan[Pakistan:Multan301:1996]	BYVMVPK-[PK:M301:96]	AJ002453
Cotton leaf curl Multan virus-Hisar[Pakistan:Multan 311:Okra:1996]	CLCuMV-[PK:M311:Ok:96]	AJ002459
Bhendi yellow vein mosaic virus-*India*[India:Madurai:2000]	BYVMV-IN-[IN:Mad:00]	AF241479
Mesta yellow vein mosaic Bahraich virus-[India:Bhanga:2008]	MeYVMBV-[IN:Bhg:08]	FJ159268
Mesta yellow vein mosaic Bahraich virus -[India:Bahraich:2007]	MeYVMBV-[IN:Bah:07]	EU360303

Cotton leaf curl Multan virus-[Pakistan:Yazman 62:1995]	CLCuMV-[PK:Y62:95]	AJ002447
Cotton leaf curl Multan virus-[India:Bhatinda:2005]	CLCuMV-Bha[IN:Bha:05]	DQ191160
Cotton leaf curl Rajasthan virus-[Pakistan:Multan:2007]	CLCuRV-[PK:Mul:07]	EU384574
Okra yellow vein mosaic virus-[India:Aurangabad:2008]	OYVMV-[IN:Abd:08]	GU181356
Cotton leaf curl Alabad viru -[Pakistan:Kohiwala802a:1996]	CLCuAV-[PK:K802a:96]	AJ002455
Cotton leaf curl Multan virus-[India:New Delhi 2:2003]	CLCuMVRaj-[IN:ND2:03]	AY795605
Cotton leaf curl Bangalore virus-[India: Bangalore:2004]	CLCuBV-[IN:Ban:04]	AY705380
Gossypium punctatum mild leaf curl virus-[Pakistan:2007]	GPMLCV-[PK:07]	EU365617
Bhendi yellow vein Delhi virus-[India:OY131:2004]	BYVDV-[IN:OY131:04]	FJ515747
Malvastrum yellow vein virus-[China:Yunan340:2008]	MaYVBV-[CN:Y340:08]	FN806779
Cotton leaf curl Multan virus-[Pakistan:Sindh:LS6:2010]	CLCuSV-[PK:Sin:LS6:10]	FN552005
Kenaf leaf curl virus-[China:Yunan341:2008]	KeLCV-[CN:Y341:08]	FN806778
Malvastrum yellow vein virus-[China:Yunan206:Ageratum:2003]	MaYVV-[CN:Yn206:Age:03]	AJ744881
Bhendi yellow vein Bhubaneswar virus -[India:OYBHU:2003]	BYVBhV-[IN:OYBHU:03]	FJ589571

Table 4.5. Percentage nucleotide identities of OYVMV-[IN:Him:09] and amino acid identities of its protein with other sequences taken for the analysis

Abbreviation	DNA-A[a]	IR[a]	V1[b]	V2[b]	C1[b]	C2[b]	C3[b]	C4[b]
OYVMV-[PK:Fai201:95]	96	95	99	98	95	94	97	91
BYVMV-[IN:OY177:07]	93	92	98	95	92	92	94	85
CLCuMV-[IN:Lud:04]	90	78	98	97	89	84	86	77
GDSV-[PK-Dar1:07]	90	76	95	95	89	88	90	76
BYVMV-[IN:Pan:05]	89	72	98	95	84	92	94	68
BYVMV-[IN:Thd:05]	89	70	98	99	83	91	94	66
OkLCuIV-[IN:ND:08]	88	76	97	93	85	87	92	69
CLCuMV-His[PK:Fai3:99]	88	78	96	96	87	84	85	75
BYVMVPK-[PK:M301:96]	88	73	98	96	84	86	91	70
CLCuMV-[PK:M311:Ok:96]	88	78	97	96	90	85	88	80
BYVMV-IN-[IN:Mad:00]	87	70	98	98	81	92	96	71
MeYVMBV-[IN:Bhg:08]	87	67	98	94	85	84	90	77
MeYVMBV-[IN:Bah:07]	87	67	97	94	85	84	90	77
CLCuMV-[PK:Y62:95]	87	74	98	95	89	85	88	74
CLCuMV-Bha[IN:Bha:05]	86	61	97	95	90	85	88	78
CLCuRV-[PK:Mul:07]	86	61	92	96	91	82	81	75
OYVMV-[IN:Abd:08]	86	67	97	NA	83	82	89	44
CLCuAV-[PK:K802a:96]	86	78	99	98	74	88	90	45
CLCuMVRaj-[IN:ND2:03]	85	61	96	96	87	84	86	76
CLCuBV-[IN:Ban:04]	84	61	92	73	90	84	81	77
GPMLCV-[PK:07]	84	74	98	96	74	81	87	45
BYVDV-[IN:OY131:04]	84	50	83	66	88	88	89	70
MaYVBV-[CN:Y340:08]	83	78	89	74	87	86	76	75
CLCuSV-[PK:Sin:LS6:10]	82	57	93	72	90	84	84	77
KeLCV-[CN:Y341:08]	81	65	89	73	85	88	79	42
MaYVV-[CN:Yn206:Age:03]	81	77	89	74	84	87	77	45
BYVBhV-[IN:OYBHU:03]	81	55	82	67	88	68	73	80

[a] nucleotide identies [b] amino acid identies

Table 4.6. Percentage nucleotide identities of OYVB-[IN:Him:09] and amino acid identities of its protein with other sequences taken for the analysis

Betasatellite Name	Abbreviation	Acc. No.	DNA[a]	BetaC1[b]
Okra leaf curl betasatellite-[Pakistan:Gojra:1997]	OLCuB-[PK:Goj:97]	AJ316029	65	93
Bhendi yellow vein betasatellite-[India:Gun:OY112:2006]	BYVB-[IN:Gun:OY112:06]	GU111969	63	97
Okra yellow vein betasatellite-[India:Aura	OYVB-[IN:Abd:08]	GU233520	63	96
Bhendi yellow vein betasatellite-[India:Barrakpur:2006]	BYVB-[IN:Bar:06]	EF417919	63	98
Okra leaf curl India betasatellite-[India:New Delhi:2007]	OLCuIB-[IN:ND:07]	GQ245761	63	98
Bhendi yellow vein betasatellite-[India:Aurangabad:OY164:2006]	BYVB-[IN:Abd:OY164:06]	GU111988	63	97
Okra leaf curl betasatellite-[India:Chandigarh:OY177:2006]	OLCuB-[IN:Chn:OY177:06]	GU111981	62	93
Bhendi yellow vein betasatellite-[India:Muthuppatti:2000]	BYVB-[IN:Mut:00]	AJ308425	62	97
Bhendi yellow vein betasatellite-[India:Raipur:OY56B:2006]	BYVB-[IN:Rai:OY56B:05]	GU111968	62	96
Bhendi yellow vein betasatellite-[India:Bangalore:OYNun:2006]	BYVB-[IN:Ban:OYNun:06]	GU111991	62	92
Bhendi yellow vein betasatellite-[India:Tirupati:OY98:2006]	BYVB-[IN:Tir:OY98:06]	GU111978	62	94
Okra leaf curl India betasatellite-[India:Mun:EL41:2006]	OLCuIB-[IN:Mun:EL41:06]	GU111965	62	93
Okra enation leaf curl betasatellite-[India:Sonipat:EL10:2006]	OELCuB-[IN:Son:EL10:06]	GU111961	62	94
Bhendi yellow vein betasatellite-[India:Thadgaon:OY156:2006]	BYVB-[IN:Tha:OY156:06]	GU111987	62	84

Tomato leaf curl New Delhi betasatellite-[India:OY142:2007]	TLCNDB-[IN:OY142:07]	EU081883	62	94
Bhendi yellow vein betasatellite-[India:Jalagaon:OY142:2006]	BYVB-[IN:Jal:OY121:06]	GU111984	61	80
Bhendi yellow vein betasatellite-[India:Pan:OY168:2006]	BYVB-[IN:Pan:OY168:06]	GU111990	60	71
Bhendi yellow vein betasatellite-[India:Tri:OY115:2006]	BYVB-[IN:Tri:OY115:06]	GU111983	59	82
Bhendi yellow vein betasatellite-[India:Kolar:2006]	BYVB-[IN:Kol:06]	GU111973	57	84
Bhendi yellow vein betasatellite-[India:Karnataka:OY80A:2006]	BYVB-[IN:Kar:OY80A:06]	GU111979	55	91
Bhendi yellow vein betasatellite-[India:Kai:2006]	BYVB-[IN:Kai:06]	GU111976	55	67
Bhendi yellow vein India betasatellite-[India:Coimbatur:OY141:2006]	BYVIB-[IN:Coimb:OY141:06]	GU111986	54	74

[a] **nucleotide identies** [b] **amino acid identies**

The multiple sequence alignment of OYVB-[IN:Him:09] with several other betasatellites showed that the coding region (BetaC1) is conserved and the non-coding region is highly variable in most of the begomovirus betasatellites. Interestingly, the SCR was absent in OYVB-[IN:Him:09] and it was noticed that this region was not showing any kind of similarity with the other betasatellites. But it was found to show maximum identity with the IR of its helper virus i.e. OYVMV-[IN:Him:09] (Fig. 4.3). Previously it was reported (Mansoor *et al.,* 2003) that the betasatellites do not show any kind sequence homology with its helper virus except for the nonanucleotide TAATATTAC sequence.

Both OYVMV-[IN:Him:09] and OYVB-[IN:Him:09] showed highest identity with the isolates of yellow vein mosaic and leaf curl virus of Okra reported from Pakistan, respectively. This indicated that the present begomovirus might have acquired the betasatellite component at some point due to mixed infection in okra leading to a novel association. Keeping in view the geographical distribution, the present viral species could be introduced from Pakistan to Himachal Pradesh due to an increase in international commodity trade, intercontinental transportation networks and its adaptation due to changing global climate leading to spread of its vector as reported earlier by Moffat (1999).

4.8 PHYLOGENETIC ANALYSIS

4.8.1 Phylogenetic analysis of DNA-A

The phylogenetic analysis of present isolate with several other closely related begomovirus DNA-A isolates showed that these begomoviruses formed three main clusters that are host specific and the present isolate clustered with OYVMV-[PK:Fai201:95] and BYVMV-[IN:OY177:09]) (Fig. 4.4). The cluster-1 belongs mainly to begomoviruses of okra from India and Pakistan and cluster-2 belongs mainly to begomoviruses of cotton, while the cluster-3 is highly diverse and mainly contains begomoviruses from China except BYVBhV-[IN:OYBHU:03] and CLCuBV-[IN:Ban:04].

4.8.2 Phylogenetic analysis of IR

In the phylogenetic analysis of the OYVMV-[IN:Him:09] IR with other begomovirus isolates, IR was found to be different than the pattern of DNA-A tree as it did not show any host specific grouping. The OYVMV-[IN:Him:09] grouped with OYVMV-[PK:Fai201:95] and BYVMV-[IN:OY177:09]) isolates as that of complete DNA-A phylogenetic analysis, but here OYVMV-[IN:Him:09] is forming a separate branch than the other two isolates (Fig. 4.5).

4.8.3 Phylogenetic analysis of betasatellite

Phylogenetic analysis of betasatellite, OYVB-[IN:Him:09] with other several closely related complete betasatellite sequences showed that present betasatellite forms group with OLCuB isolates instead of BYVB isolates reported from Indian subcontinent. The clustering of OYVB-[IN:Him:09] with other OLCuB isolates is strongly supported by 100% bootstrap support but OYVB-[IN:Him:09] branched separately within this cluster, occupying a distinct place in a tree (Fig. 4.6). These differences clearly indicate that OYVB-[IN:Him:09] is a new species associated with OYVMV-[IN:Him:09].

Figure 4.4. Phylogenetic relationship between OYVMV-[IN:Him:09] and other begomovirus isolates in the complete DNA-A. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Evolutionary analyses were conducted in MEGA5

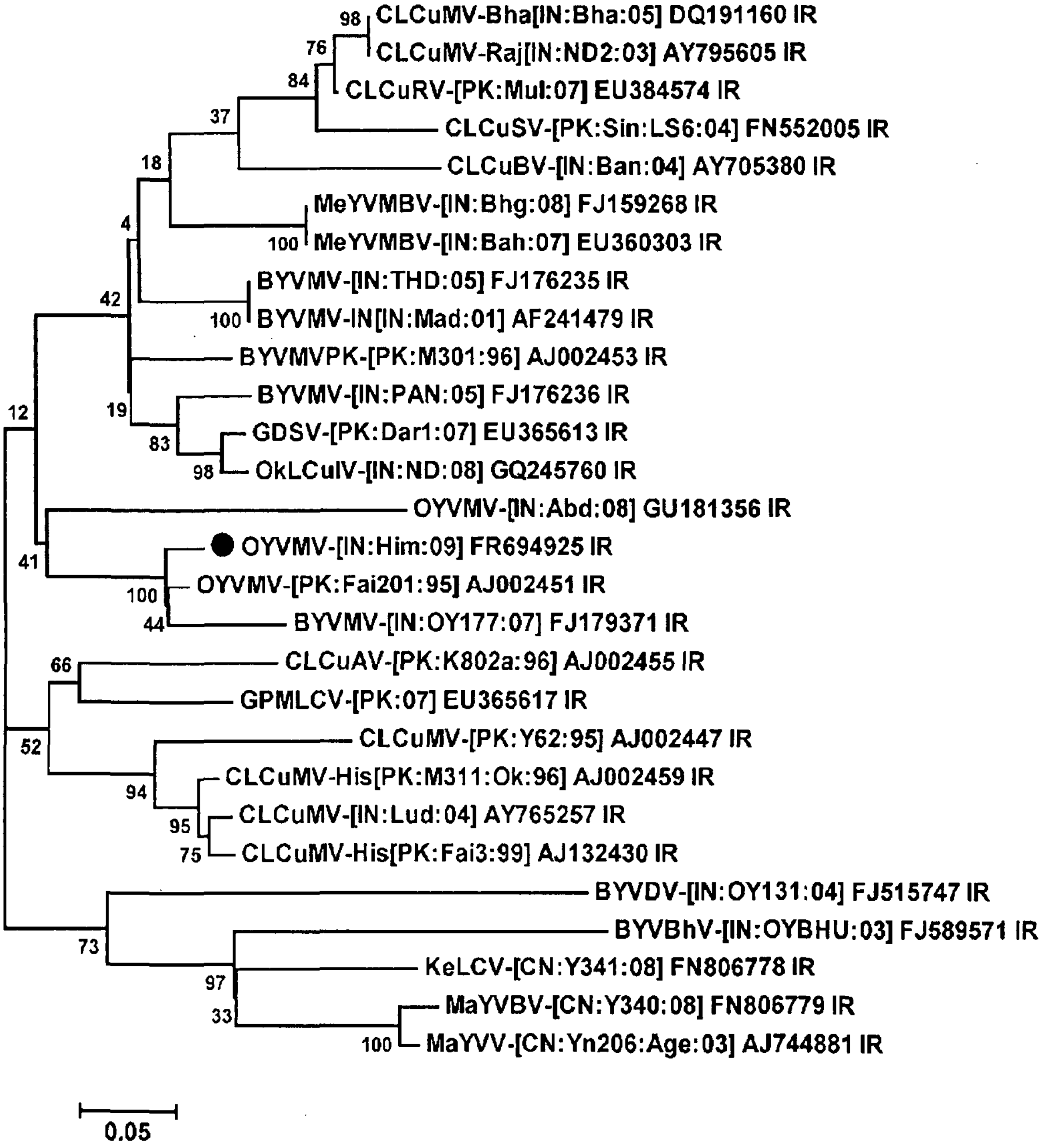


Figure 4.5. Phylogenetic relationship between OYVMV-[IN:Him:09] and other begomovirus isolates in the IR. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Evolutionary analyses were conducted in MEGA5

Interestingly BetaC1 sequence on phylogenetic analysis found to be grouping with BYVB isolates reported from India (Fig. 4.7), while the non coding region was grouping with OLCuB (Fig. 4.8). This result suggested that the non coding region may be came from OLCuB and coding region from BYVB by recombination to form present betasatellite, OYVB-[IN:Him:09] during the course of evolution.

In BLAST search it was found that the present betasatellite was showing some kind of similarity with DNA-A of some begomoviruses, hence the region (~1192-50 nt), which is a SCR of other reported betasatellites, had been taken along with part IRs of some closely related begomoviruses for phylogenetic analysis and interestingly it was found that this end region of the present betasatellite was clustered with begomovirus DNA-A isolates (Fig. 4.9). This suggests that there might be some recombination that had happened between DNA-A of OYVMV-[IN:Him:09] with some unknown betasatellite where the part of IR region of DNA-A was donated to the betasatellite replacing its SCR to form a present betasatellite i.e. OYVB-[IN:Him:09].

4.9 RECOMBINATION ANALYSIS

4.9.1 Recombination analysis of DNA-A

The recombination analysis of OYVMV-[IN:Him:09] with several closely related begomoviral sequences (Table 4.4) showed 6 best recombination events for present begomovirus with two perfect major recombination events (Fig. 4.10a). One of the event showed a recombination between OYVMV-[PK:Fai201:95] as a major parent and CLCuMV-[PK:62:05] as a minor parent to give rise to the daughter OYVMV-[IN:Him:09] (the present isolate; Fig. 4.10b). The recombination beginning breakpoint is at position 1474 (position 1524 in alignment) and the ending breakpoint is determined at position 1723 (position 1776 in alignment; fig 4.20). This recombination event was detected by 7 different methods (Table 4.7).

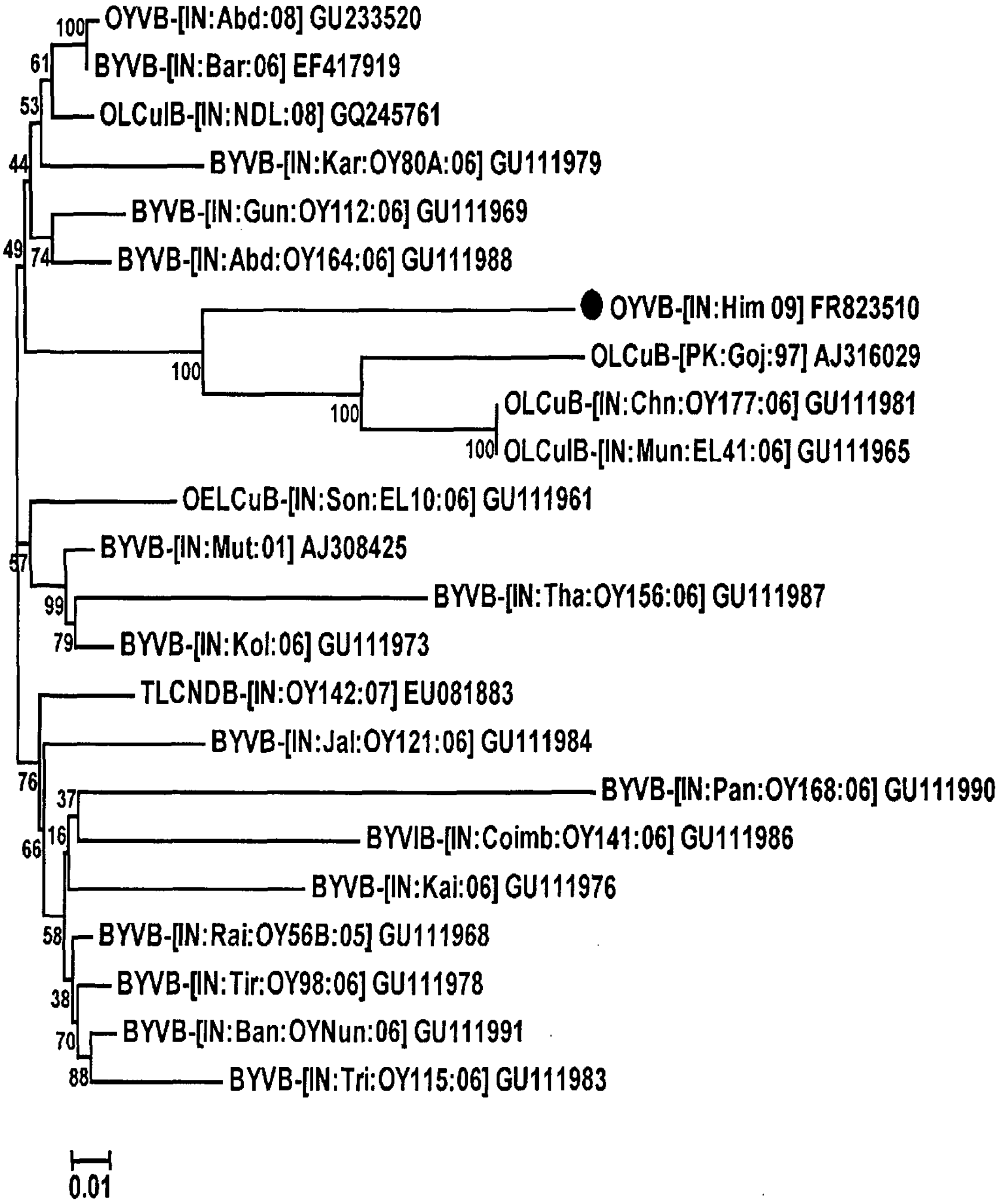


Figure 4.6. Phylogenetic relationship between OYVB-[IN:Him:09] and other begomovirus betasatellites. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Evolutionary analyses were conducted in MEGA5

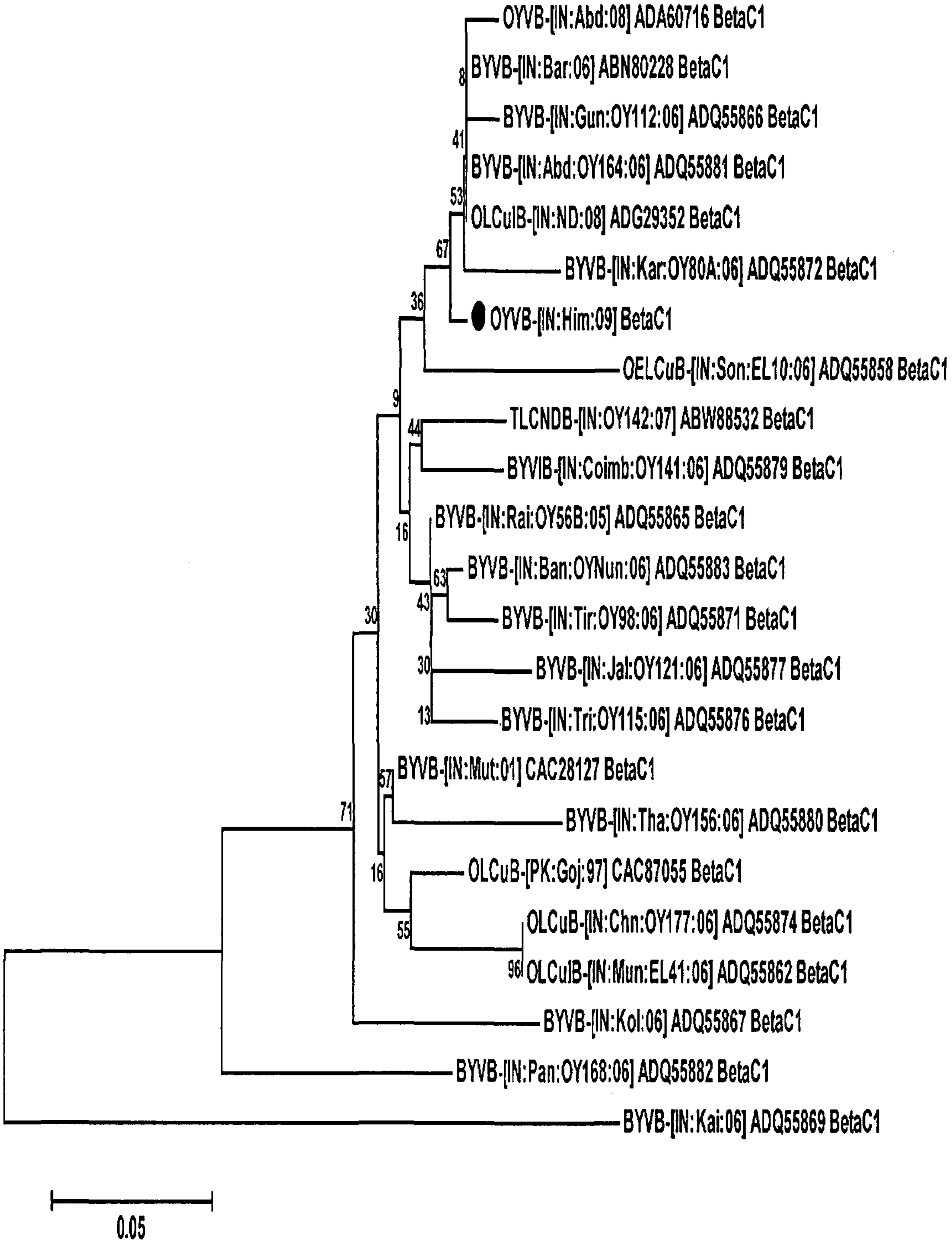


Figure 4.7. Phylogenetic relationship between OYVB-[IN:Him:09] BetaC1 ORF and other begomovirus betasatellites BetaC1 ORF. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Evolutionary analyses were conducted in MEGA5

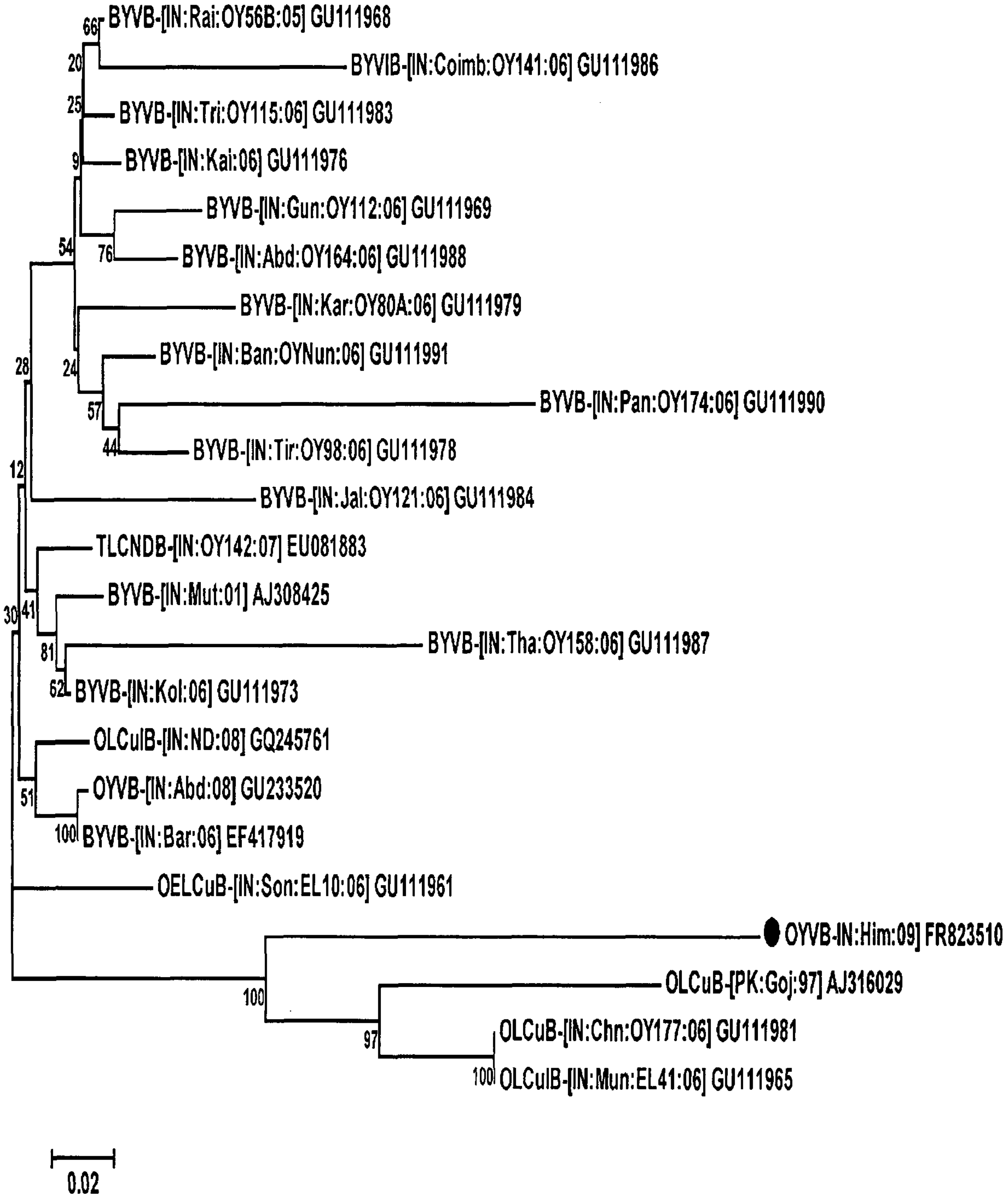


Figure 4.8. Phylogenetic relationship between OYVB-[IN:Him:09] non-coding region and other begomovirus betasatellites non-coding region. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Evolutionary analyses were conducted in MEGA5

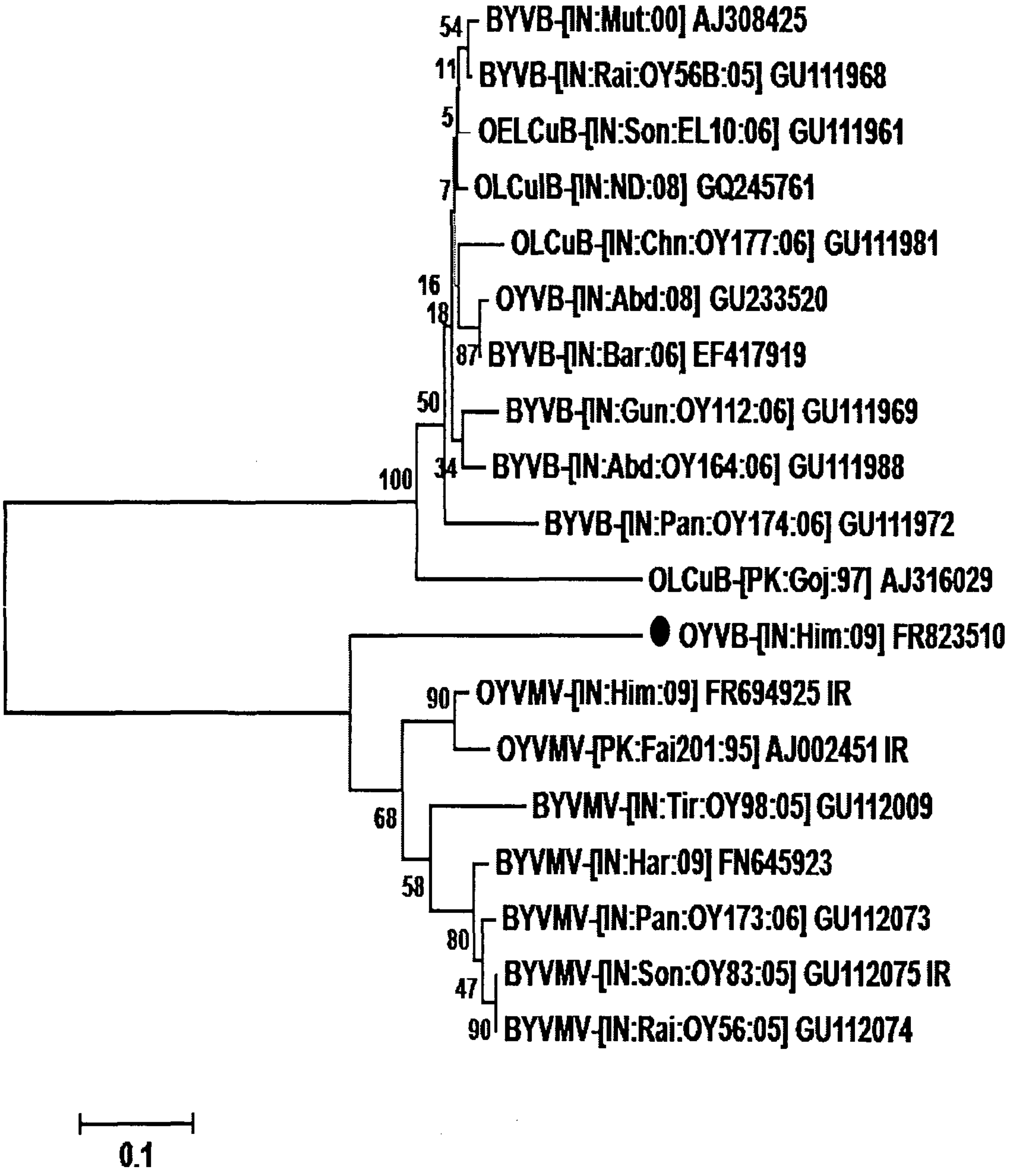


Figure 4.9. Phylogenetic relationship between OYVMV-[IN:Him:09] and other begomovirus isolates along with some IRs of begomoviruses. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Evolutionary analyses were conducted in MEGA5

The second recombination event showed a recombination between CLCuRV-[PK:Mul:07] as a major parent and OYVMV-[IN:Abd:08] as a minor parent to give rise to the daughter OYVMV-[IN:Him:09] (the present isolate; Fig. 4.10c). The recombination beginning breakpoint is at position 407 (position 448 in alignment) and the ending breakpoint is determined at position 1417 (position 1467 in alignment; Plate 4.11). This recombination event was detected by 7 different methods (Table 4.8).

These both recombination events showed that the OYVMV-[IN:Him:09] was evolved as a result of high rate of interspecies recombination between the geminiviruses as reported earlier by Zhou *et al.* (1997) and Padidam *et al.* (1999).

Table 4.7. Recombination confirmation table of OYVMV-[PK:Fai201:95] and CLCuMV-[PK:62:05] to give OYVMV-[IN:Him:09]. Analysis was conducted in RDP3 version

Methods	Events	Average P-value
RDP	1	1.359×10^{-15}
GENECONV	1	1.372×10^{-10}
BootScan	1	4.415×10^{-14}
MaxChi	1	7.083×10^{-07}
Chimaera	1	1.472×10^{-07}
SiScan	1	1.086×10^{-05}
3Seq	1	1.272×10^{-11}

Table 4.8. **Recombination confirmation table of CLCuRV-[PK:Mul:07] and OYVMV-[IN:Abd:08] to give rise to the OYVMV-[IN:Him:09]. Analysis was conducted in RDP3**

Methods	Events	Average P-value
RDP	2	$9.844 \text{X} 10^{-21}$
GENECONV	2	$1.847 \text{ X } 10^{-18}$
BootScan	2	$1.857 \text{ X } 10^{-20}$
MaxChi	2	$5.093 \text{ X } 10^{-10}$
Chimaera	2	$7.383 \text{ X } 10^{-11}$
SiScan	2	$1.453 \text{ X } 10^{-18}$
3Seq	2	$6.399 \text{ X } 10^{-19}$

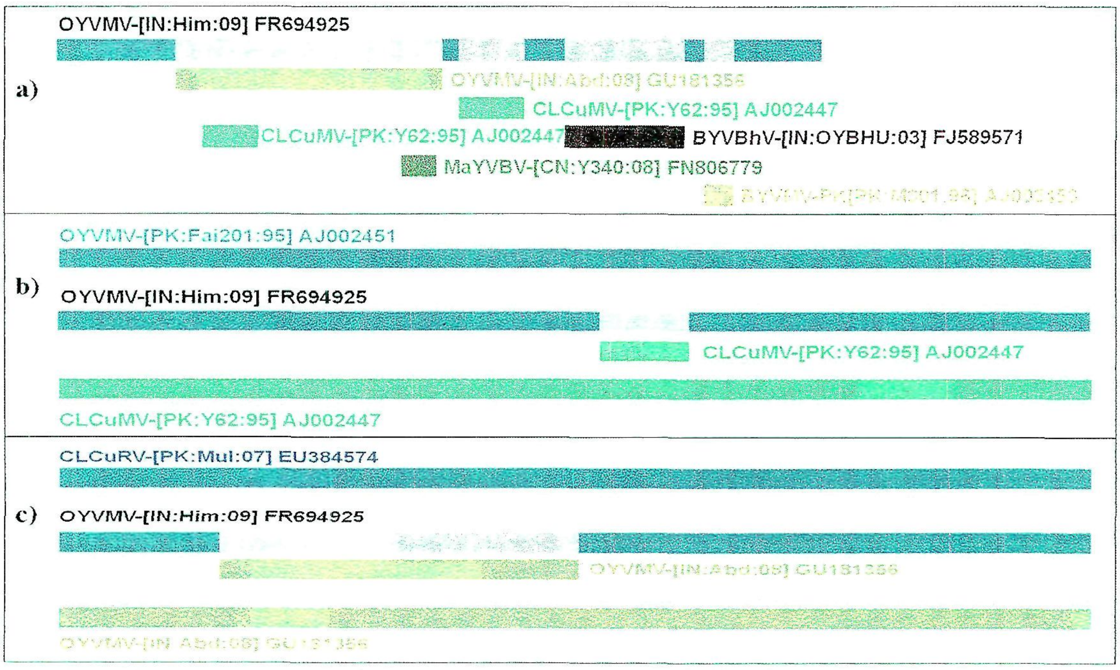


Figure 4.10. Best recombination events for OYVMV-[IN:Him:09] (a) and perfect recombination events for OYVMV-[IN:Him:09] (b and c) as detected by RDP3

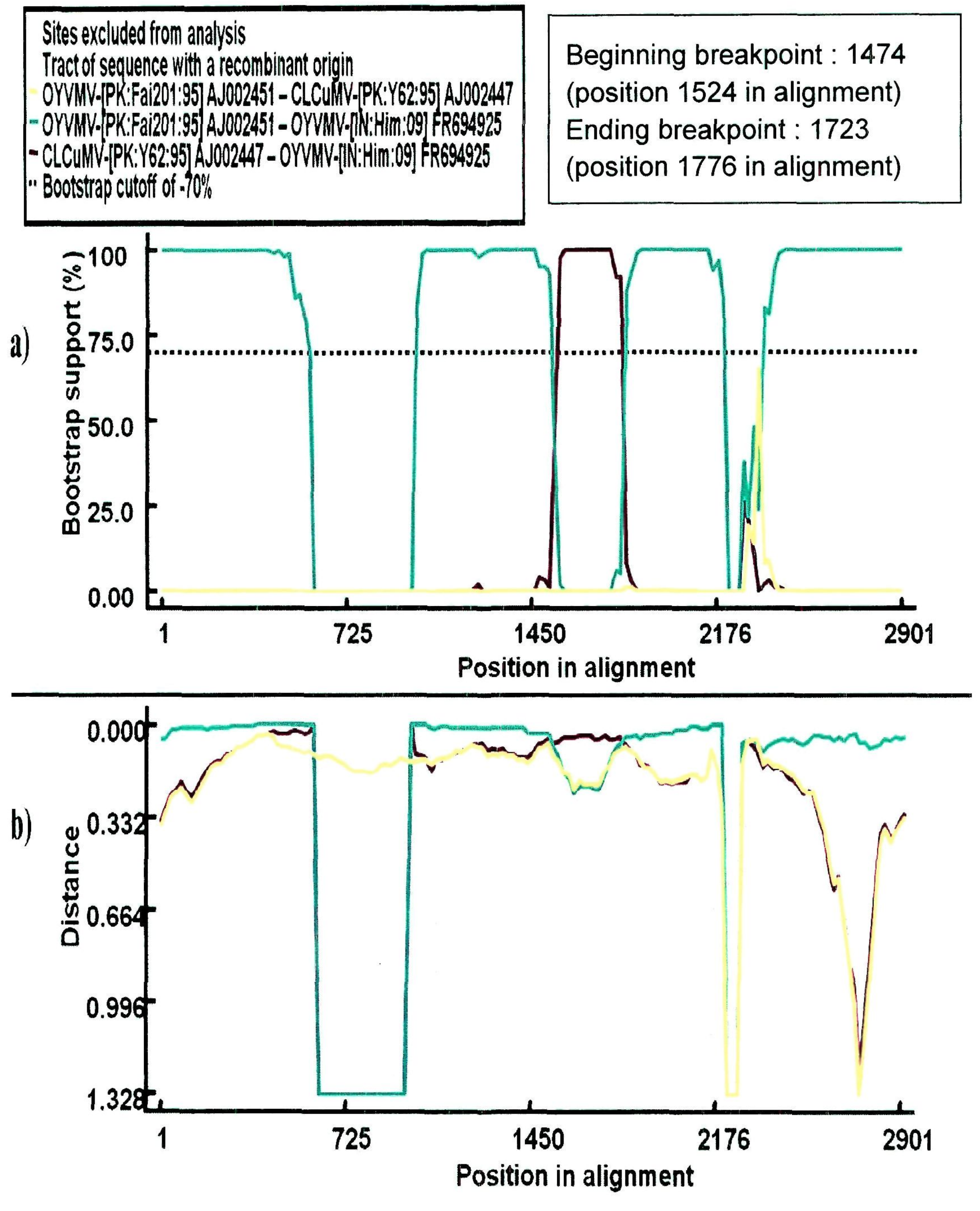
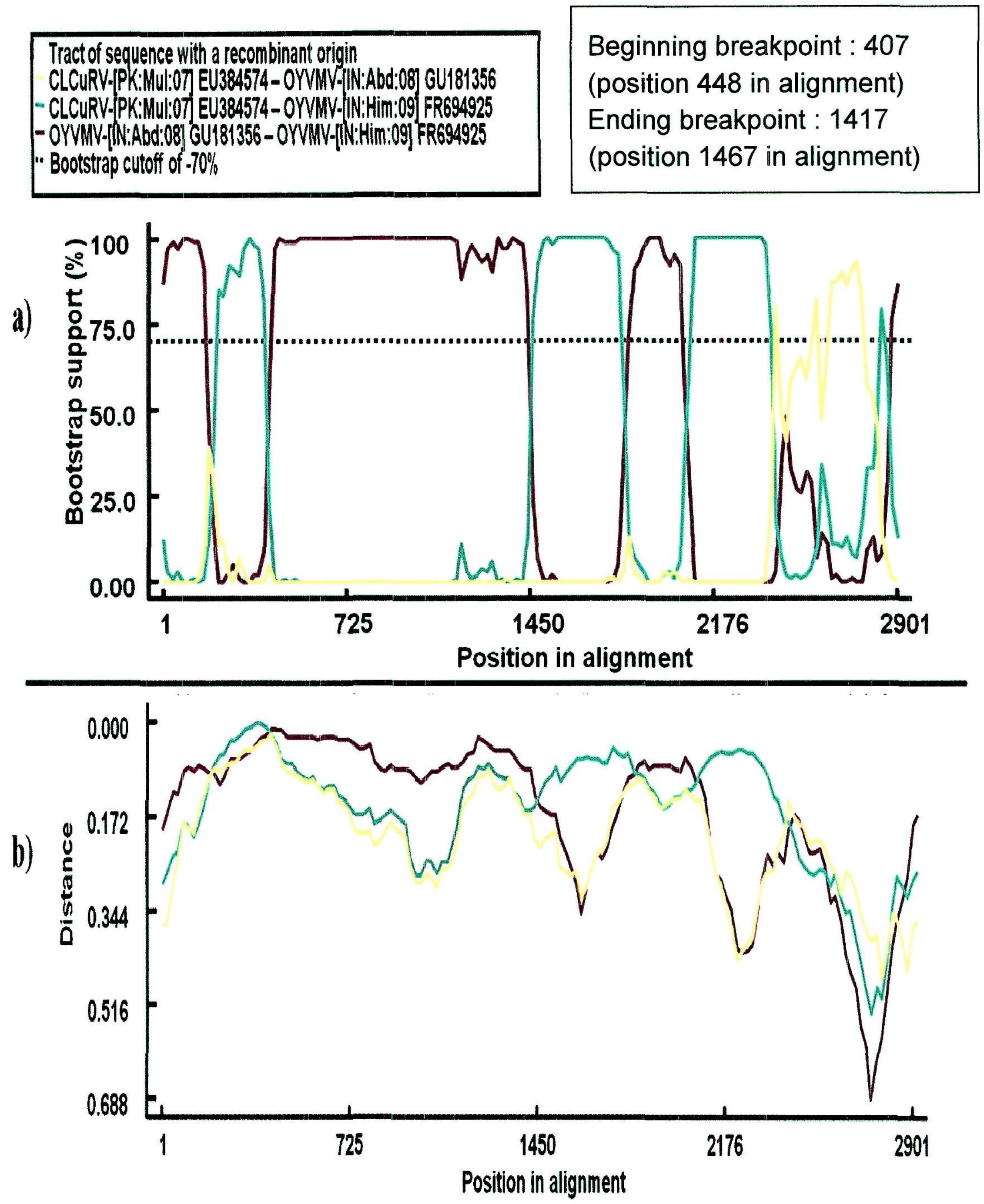


Plate 4.10. Plots showing recombination between OYVMV-[PK:Fai201:95] and CLCuMV-[PK:62:05] to give rise to OYVMV-[IN:Him:09]. (a) Bootscan plot and (b) distance plot showing recombination region



Beginning breakpoint : 407
(position 448 in alignment)
Ending breakpoint : 1417
(position 1467 in alignment)

Plate 4.11. Plots showing recombination between CLCuRV-[PK:Mul:07] and OYVMV-[IN:Abd:08] to give rise to the OYVMV-[IN:Him:09]. (a) Bootscan plot and (b) distance plot showing recombination region

4.9.2 Recombination analysis of betasatellite

The recombination analysis of betasatellite, with several closely related betasatellite sequences was carried out (Table 4.6). The recombination analysis was performed and three best recombination events were detected for present Betasatellite (Fig. 4.11a), with one major event. This showed that recombination had occurred between Betasatellites, OLCuIB-[IN:Mun:EL41:06] (major parent) and BYVB-[IN:Abd:OY164:06] (minor parent) to form a recombinant OYVB-[IN:Him:09] (present satellite DNA, Fig. 4.11b). In this recombination event, the beginning breakpoint is at position 213 (position 214 in alignment) and the ending breakpoint is at position 427 (position 471 in alignment) (Plate 4.12). The event was detected by seven different methods (Table 4.9)

This recombination analysis showed that the recombination between betasatellite was also interspecies, which leads to completely novel betasatellite OYVB-[IN:Him:09]. The recombination was independent of its helper virus as the recombination events detected in OYVMV-[IN:Him:09] were having totally different begomoviral isolates that that involved in betasatellite recombination (Fig. 4.10 and Fig. 4.11, respectively). The recombination event showed that recombination was occurred between OLCuB isolate and BYVB isolate to form OYVB-[IN:Him:09], which was confirming the result obtained after phylogenetic analysis (section 4.8.3), where the non coding region of OYVB-[IN:Him:09] was grouping with OLCuB isolates (Fig. 4.8) and coding region was grouping with BYVB isolates (Fig. 4.7).

Many recombination events were detected for both OYVMV-[IN:Him:09] and OYVB-[IN:Him:09] suggesting that the rate of recombination is higher in geminiviruses leading to great diversity among geminiviruses. There are various reasons that were reported be responsible for this higher rate of recombination between geminiviruses. One reason reported was the genome organization and rolling circle replication (RCR) which seems to increase the basal recombination rates in different parts of geminivirus genomes (Garcia-Andres *et al.* 2007; Jeske *et al.* 2001; Owor *et al.* 2007). The interspecies

Table 4.9. **Recombination confirmation table of OLCuIB-[IN:Mun:EL41:06] and BYVB-[IN:Abd:OY164:06] to form a recombinant OYVB-[IN:Him:09]. Analysis was conducted in RDP3**

Methods	Events	Average P-value
RDP	1	5.075×10^{-21}
GENECONV	1	5.771×10^{-22}
BootScan	1	2.284×10^{-21}
MaxChi	1	2.980×10^{-12}
Chimaera	1	1.337×10^{-12}
SiScan	1	4.786×10^{-17}
3Seq	1	1.122×10^{-19}

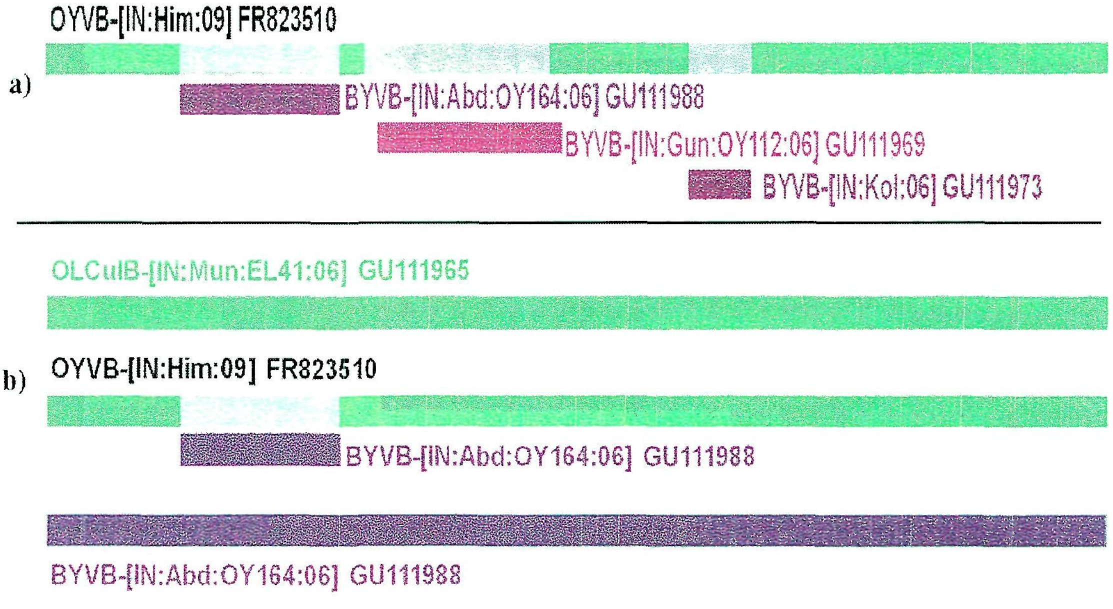


Figure 4.11 Best recombination events for OYVB-[IN:Him:09] (a) and perfect recombination event (b) as detected by RDP3

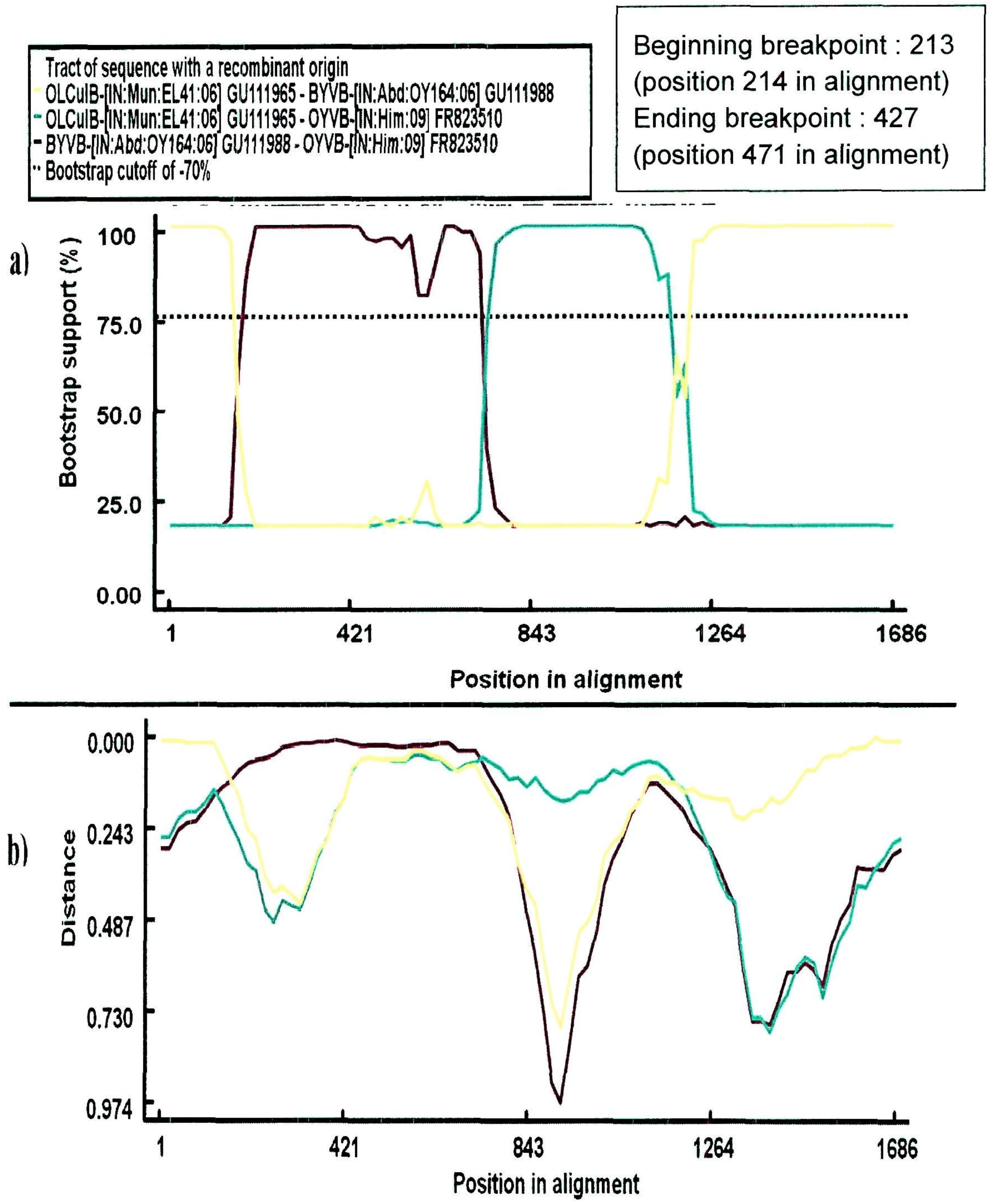


Plate 4.12. Plots showing recombination event between OLCuIB-[IN:Mun:EL41:06] and BYVB-[IN:Abd:OY164:06] to form a recombinant OYVB-[IN:Him:09]. (a) Bootscan plot and (b) distance plot showing recombination region

recombination events were detected for both OYVMV-[IN:Him:09] and OYVB-[IN:Him:09]. The both interspecies and interstrain recombination was found common in geminiviruses (Lefeuvre *et al.* 2009) leading to rapid variation in geminivirus groups.

4.10 INFECTIOUS CLONE CONSTRUCTION AND INFECTIVITY ASSAY

The construction of infectious clones was performed in two step cloning approach in pCambia 1300 and pCAM-OYVM-A was obtained as infectious clone. The orientation of the inserts was then monitored with end sequencing and restriction digestion with different enzymes that are having single restriction site in DNA-A, so that all these enzymes should yield 2.7 kb full length fragment (Plate 4.13). This confirmed that the inserts were in tandem orientation.

The infectious construct pCAM-OYVM-A was transformed into chemically competent cells of *Agrobacterium tumefaciens* strain EHA 105. Then the construct was agroinoculated into leaves, stems or petioles of okra and *N. benthamiana* plants at 4-6 leaf stage with a fine 1ml syringe. The inoculated plants were then checked after 35 days for infectivity assay. When the construct containing the viral genome was infiltrated alone, some yellowing of leaves and deep notches in leaf lamina were observed (Plate 4.14) but the typical yellow vein clearing symptoms were not observed.

The replication and movement of viral genome in plant was confirmed by isolating the DNA from newly emerged leaves. Presence of virus was confirmed by carrying out the PCR using geminivirus DNA-A specific AV494 and AC1048 (Wyatt and Brown 1996) degenerate primers and by southern hybridization using DNA-A specific probe (Plate 4.15). These results were found positive suggesting that the DNA-A component was successfully multiplying and moving across the plant. This confirmed the monopartite nature of virus as if it is a bipartite

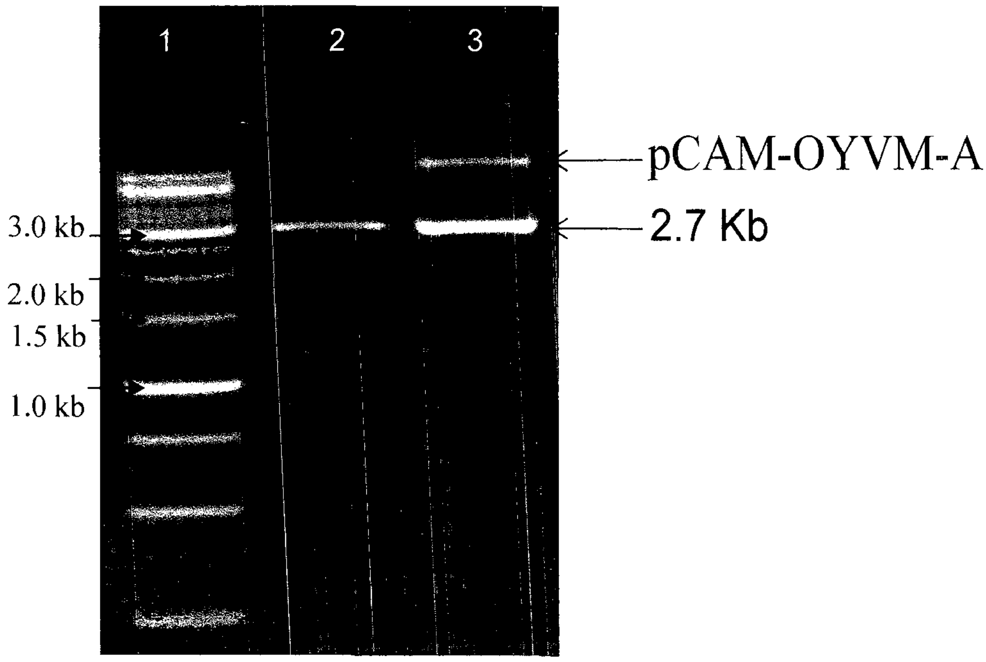


Plate 4.13. Restriction digestion of pCAM-OYVM-A with different enzymes to check the orientation of the inserts. Lane 1: 1 Kb marker, 2: *Hind*III digest, 3: *Sal*I digest

Plate 4.14. Agroinoculated plants photographed after 35dpi. A) Okra plants showing symptoms of yellowing and deep notches in leaf lamina on inoculated plant; B) *N. benthamiana* plants showing reduced growth (i) as compared with the control (ii).

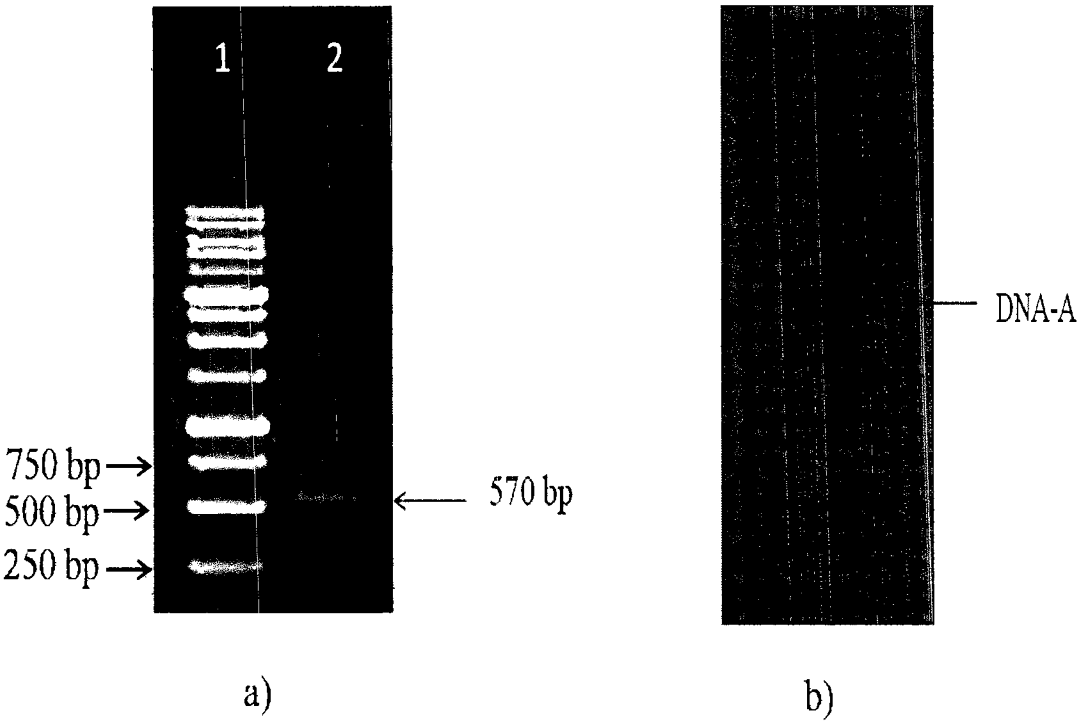


Plate 4.15. Confirmation of DNA-A multiplication in host plant. (a) PCR confirmation and (b) Confirmation by Southern hybridization with DNA-A specific probe.

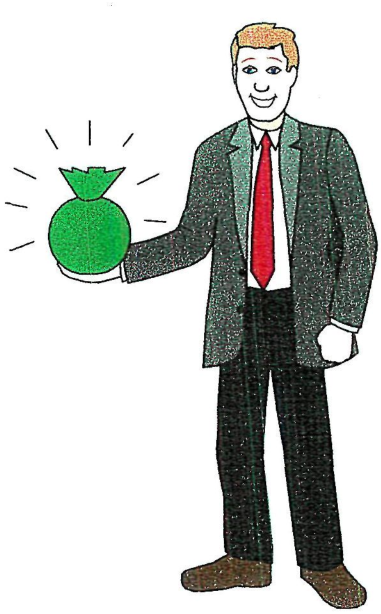
begomovirus then DNA-A alone cannot be able to move across the plant and DNA-B component is necessary for its systemic spread into the host plant. Also, as no typical vein clearing symptom was reported with DNA-A alone, supporting that the betasatellite might be necessary to produce typical symptoms in host plant. It has been reported erlier by Jose and Usha (2003) that the betasatellite component is necessary for producing typical vein clearing symptoms in okra host plant.

4.11 C, H, N, S ANALYSIS:

The C, H, N, S analysis of plants had shown that the leaves were showing less percentage of C, H, N, S as compared to fruits as fruits are being an important storage organ (Table 4.10). The analysis of C, H, N, S in infectious clone inoculated and non-inoculated leaves clearly shown that infectious clone inoculated samples has lesser amount of C, H, N, S (Table 4.10). This could be due to viral multiplication leading to degradation of chlorophyll in infectious clone inoculated plants.

Table 4.10. Analysis of CHNS in plant samples by using CHNS Analyzer

Sr. No.	Name	Weight (mg)	%N	%C	%H	%S	AbsN	AbsC	AbsH	AbsS
1	Blank	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Runln	10.42	15.99	39.69	4.80	24.44	1.69	4.14	0.50	2.55
3	Sulfanilamid	10.68	16.26	41.85	4.83	18.62	1.74	4.32	0.52	2.14
4	Non-inoculated Leave	10.32	2.50	30.29	5.71	0.52	0.24	2.87	0.66	0.04
5	Non-inoculated Fruits	10.61	4.29	39.70	7.45	0.66	0.47	3.97	0.77	0.04
6	Agroinoculated Leave	10.37	2.12	23.54	3.86	0.47	0.23	2.61	0.36	0.03
7	Agroinoculated Fruits	10.49	3.15	29.48	5.53	0.57	0.33	2.83	0.56	0.04



Summary and Conclusions

5. SUMMARY AND CONCLUSIONS

Okra is an economically important vegetable crop grown in tropical and sub-tropical parts of the world. It is also cultivated in sub-temperate region of Himachal Pradesh. The yellow vein mosaic disease, the most serious viral disease in okra, was observed in this area during last couple of years. The increase of this disease incidence in sub-temperate regions of Himachal Pradesh has resulted in need for accurate detection and characterization of the virus. Keeping this in mind, the present study entitled, "Genome organization and infectious clone construction of a geminivirus causing yellow vein mosaic in okra" was undertaken to detect and sequence the complete genome of the virus causing yellow vein mosaic in okra and to construct its infectious clone. For this, six samples were collected from different locations of Himachal Pradesh, Maharashtra and Jammu and Kashmir, based on the symptoms appeared on okra leaves. Out of these samples, the positive sample from Punder was then selected for further characterization and all other research work was then carried out on this sample.

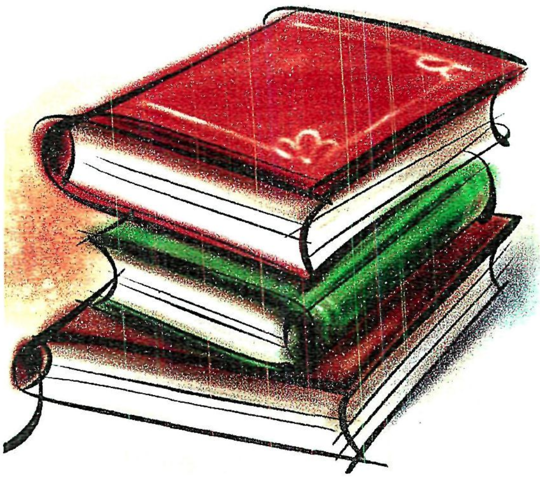
It is for the first time that a monopartite begomovirus species causing disease in okra was detected from sub-temperate region of Himachal Pradesh. The present viral species, *Okra yellow vein mosaic virus* (OYVMV-[IN:Him:09]), is also found to be associated with a distinct betasatellite and no DNA-B component was reported to be associated. The complete genome of the virus was sequenced with 2739 bases length and the complete nucleotide sequence of betasatellite was determined to be 1332 bases long. The genome organization study showed that six different ORFs are present on virion sense and complementary strand separated by one intergenic region (IR). The betasatellite associated with present begomovirus, *Okra yellow vein betasatellite* (OYVB-[IN:Him:09]), was also found to have one ORF and one A- rich region in its genome. OYVMV-[IN:Him:09] was found to be a isolate of OYVMV-

[PK:Fai201:95], as it showed 96% nucleotide identity with OYVMV-[PK:Fai201:95]. The betasatellite associated with present begomovirus was found to be novel as it showed highest identity (65%) with *Okra leaf curl betasatellite* (OLCuB-[PK:Goj:97]) reported from Pakistan which is less than the threshold value (78%) for species demarcation in begomovirus betasatellites, suggesting it to be a new betasatellite species associated with present begomovirus. Interestingly, both OYVMV-[IN:Him:09] and OYVB-[IN:Him:09] showed highest identity with the isolates of yellow vein mosaic and leaf curl virus of Okra reported from Pakistan, respectively. This indicated that the present begomovirus might have acquired the betasatellite component at some point due to disease complex in okra leading to a novel association. Keeping in view the geographical distribution, the present viral species could be introduced from Pakistan to Himachal Pradesh due to an increase in international commodity trade, intercontinental transportation networks and a changing global climate leading to spread of its vector.

In the betasatellite, the satellite conserved region (SCR) was absent and this corresponding region showed highest identity with IR of its helper virus. The phylogenetic analysis also showed that the SCR replaced region of present betasatellite was grouping with IR sequences of begomoviruses instead of SCR sequences of betasatellites. This indicates that there might be some recombination that had happened between DNA-A of OYVMV-[IN:Him:09] with some unknown betasatellite where the part of IR region of DNA-A was donated to the betasatellite replacing its SCR to form a present betasatellite i.e. OYVB-[IN:Him:09].

The recombination analysis showed that OYVMV-[IN:Him:09] was generated from recombination between OYVMV-[PK:Fai201:95] and CLCuMV[PK:62:05]. Similarly the Betasatellite was found to be recombinant generated from OLCuIB-[IN:Mun:EL41:06] and BYVB-[IN:Abd:OY164:06]. This analysis showed that both present viral DNA components were generated as a result of interspecies recombination leading to novel association. Betasatellite was found to be highly variable outside the gene coding region and showed highly conserved BetaC1, this indicate that the recombination events were happened outside the gene coding region during the stage of evolution, which leads to emergence of present distinct Betasatellite species.

Infectious clone of present begomovirus was constructed using two step cloning approach and agroinoculation studies with begomoviral DNA alone showed that this component is not sufficient to produce typical vein clearing symptoms in host plant although it was producing some yellowing in leaves and leaf lamina showed prominent notches. It was also found that the begomoviral DNA was replicating and moving across the different parts of plant, confirming the monopartite nature of virus. This also showed that betasatellite is not necessary for replication and movement of plant virus but it may involved in producing typical symptoms of vein clearing in host plant.



Literature Cited

LITERATURE CITED

Abouzid AM, Freitas-Astua J, Purcifull DE, Polston JE, Beckham KA, Crawford WE, Petersen MA, Peyser B, Patte C and Hiebert E. 2002. Serological studies using polyclonal antisera prepared against the viral coat protein of four begomoviruses expressed in Escherichia coli. *Plant Disease* 86(10): 1109-1114

Ach RA, Durfee T, Miller AB, Taranto P, Hanley-Browdoin L, Zambryski PC and Gruissem W. 1997. RRB1 and RRB2 encode maize retinoblastoma-related proteins that interact with a Plant D-type cyclin and geminivirus replication protein. *Molecular and Cellular Biology* 17(9): 5077- 5086

Ali S, Khan MA, Habib A, Rasheed S and Iftikhar Y. 2005. Correlation of environmental conditions with *Okra yellow vein mosaic virus* and *Bemisia tabaci* population density. *International Journal Of Agriculture & Biology* 7(1):142-144

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25(17):3389-3402

Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, and Daszak P. 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology Evolution* 19(10): 535-544

Anonymous. 2008. FAO estimates, FAOSTAT. http://faostat.fao.org/site/339/default.aspx [11th February, 2011]

Anonymous. 2010. National Horticulture Board (NHB), India. http://nhb.gov.in/statistics/area-production-statistics.html [15th March, 2011)

Arguello-Astorga GR, Guevara-Gonzalez RG, Herrera-Estrella LR and Rivera-Bustamante RF. 1994. Geminivirus replication origins have a group-specific organization of iterative elements - a model for replication. *Virology* 203(1): 90-100.

Baas PD and Jansz HS. 1988. Single-strand DNA phage origins. Current Topics in Microbiology and Immunology 136: 31-70

Behjatnia SAA, Dry IA and Rezaian MA. 1998. Identification of the replication-associated protein binding domain within the intergenic region of tomato leaf curl geminivirus. *Nucleic Acids Research* 26(4): 925-931

Bigarre L, Chazly M, Salah M, Ibrahim M, Padidam M, Nicole M, Peterschmitt M, Fauquet C and Thouvenel JC. 2001. Characterization of a new begomovirus from Egypt infecting hollyhock (*Althea rosea*). *European Journal of Plant Pathology* 107(7): 701-711

Bird J and Maramorosch K. 1978. Viruses and virus diseases associated with whiteflies. *Advances in Virus Research* 22: 55-110

Bock KR. 1982. Geminivirus diseases in tropical crops. *Plant disease* 66: 266-270.

Bosque-Perez NA. 2000. Eight decades of maize streak virus research. *Virus Research* 71(1-2): 107-21

Boulton MI. 2002. Functions and interactions of mastrevirus gene products. *Physiological and Molecular Plant Pathology* 60(5): 243-255.

Briddon RW, Brown JK, Moriones E, Stanley J, Zerbini M, Zhou X, Fauquet CM. 2008. Recommendations for the classification and nomenclature of the DNA-β satellites of begomoviruses. *Archives of Virology* 153(4): 763-781

Briddon RW, Bull SE, Amin I, Idris AM, Mansoor S, Bedford ID, Dhawan P, Rishi N, Siwatch SS and Abdel-Salam AM. 2003. Diversity of DNA ß, a satellite molecule associated with some monopartite begomoviruses. *Virology* 312(1): 106-121

Briddon RW, Bull SE, Amin I, Mansoor S, Bedford ID, Rishi N, Siwatch SS, Zafar Y, Abdel-Salam AM and Markham PG. 2004. Diversity of DNA 1: a satellite-like molecule associated with monopartite begomovirus-DNA ß complexes. *Virology* 324(2): 462-474.

Briddon RW, Bull SE, Mansoor S, Amin I, and Markham PG. 2002. Universal primers for the PCR-mediated amplification of DNA ß - a molecule associated with some monopartite begomoviruses. *Molecular Biotechnology* 20(3): 315-318.

Briddon RW, Mansoor S, Bedford ID, Pinner MS, Saunders K, Stanley J, Zafar Y, Malik KA and Markham PG. 2001. Identification of DNA components required for induction of cotton leaf curl disease. *Virology* 285(2): 234-243.

Briddon RW, Pinner MS, Stanley J and Markham PG. 1990. Geminivirus coat protein gene replacement alters insect specificity. *Virology* 177(1): 85-94

Bull SE, Briddon RW and Markham PG. 2003. Universal primers for the PCR-mediated amplification of DNA 1: A satellite-like molecule associated with begomovirus-DNA beta complexes. *Molecular Biotechnology* 23(1): 83-86

Bull SE, Tsai WS, Briddon RW, Markham PG, Stanley J and Green SK 2004. Diversity of begomovirus DNA beta satellites of non-malvaceous plants in east and south east Asia - Brief report. *Archives of Virology* 149(6): 1193-1200

Capoor SP and Varma PM. 1950. Yellow vein mosaic of *Hibiscus esculentus* L. *Indian Journal of Agricultural Sciences* 20: 217-230

Castillo AG, Collinet D, Deret S, Kashoggi A and Bejarano ER. 2003. Dual interaction of plant PCNA with geminivirus replication accessory protein (REn) and viral replication protein (Rep). *Virology* 312(02): 381-94

Chakraborty S, Pandey PK, Banerjee MK, Kalloo G and Fauquet CM. 2003. Tomato leaf curl Gujarat virus, a new Begomovirus species causing a severe leaf curl disease of tomato in Varanasi, India. *Phytopathology* 93(12): 1485-1495.

Challappan P, Vanitharani R and Fauquet CM. 2005. MicroRNA-binding viral protein interferes with *Arabidopsis* development. *Proceedings of the National Academy of Sciences* 102(29): 10381-10386

Chatterjee A and Ghosh SK. 2007. Association of a satellite DNA β molecule with mesta yellow vein mosaic disease. *Virus Genes* 35(03): 835-844.

Costa AS. 1965. Three whitefly-transmitted virus diseases of beans in São Paulo, Brazil. *FAO Plant Protection Bulletin* 13: 2-12.

Cromie GA, Connelly JC and Leach DR. 2001. Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. Molecular Cell 8(6):1163-1174.

Czosnek H and Laterrot H. 1997. A worldwide survey of tomato yellow leaf curl viruses. *Archives of Virology* 142(7): 1391-1406

Das S, Ghosh R, Paul S, Roy A and Ghosh SK. 2008. Complete nucleotide sequence of a monopartite begomovirus associated with yellow vein mosaic disease of mesta from north India. *Archives of Virology* 153(9): 1791-1796

Desbiez C, David C, Mettouchi A, Laufs J and Gronenborn B. 1995. Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. *Proceedings of the National Academy of Sciences* 92(12): 5640-5644

Doyle JJ and Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin* 19: 11-15

Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12:13-15

Dry IB, Krake LR, Rigden JE and Rezaian MA. 1997. A novel sub-viral agent associated with a geminivirus: The first report of a DNA satellite. *Proceedings of the National Academy of Sciences* 94(13): 7088-7093

Duan YP, Powell CA, Purcifull DE, Broglio P and Hiebert E. 1997. Phenotypic variation in transgenic tobacco expressing mutated geminivirus movement/pathogenicity (BC1) proteins. *Molecular Plant-Microbe Interactions* 10(9): 1065-1074

Duffy S and Holmes E. (2009). Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *Journal of General Virology* 90(6): 1539-1547

Escriu F, Fraile A and Garcia-Arenal F. 2007. Constraints to genetic exchange support gene coadaptation in a tripartite RNA virus. PLoS Pathogens 3(1):67-74

Etessami P, Watts J and Stanley J. 1989. Size reversion of African cassava mosaic virus coat protein gene deletion mutants during infection of *Nicotiana benthamiana*. *Journal of General Virology* 70(2): 277-289.

Fajinmi, A. A. and Fajinmi, O. B. 2010. Incidence of okra mosaic virus at different growth stages of okra plants (*Abelmoschus esculentus* (L.) Moench) under tropical condition. *Journal of General and Molecular Virology* 2 (1) : 28-31

Fauquet CM, Bisaro DM, Briddon RW, Brown J, Harrison BD, Rybicki EP, Stenger DC and Stanley J. 2003. Revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of begomovirus species. *Archives of Virology* 148(2): 405-421

Fauquet CM, Briddon RW, Brown JK, Moriones E, Stanley J, Zerbini M and Zhou X. 2008. Geminivirus strain demarcation and nomenclature. *Archives of Virology* 153(4): 783-821.

Fauquet CM, Maxwell DP, Gronenborn B and Stanley J. 2000. Revised proposal for naming geminiviruses. *Archives of Virology* 145(8):1743-1761

Ferreira TP, Lemos TO, Nagata T, Inoue-Nagata AK. 2008. One step cloning approach for construction of agroinfectious begomovirus clones. *Journal of Virological Methods* 147(2): 351-354

Fontes EP, Eagle PA, Spire PS, Luckow VA and Hanley-Bowdoin L. 1994(a). Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *The Journal of Biological Chemistry* 269(11): 8459-8465

Fontes EP, Gladfelter HJ, Schaeffer RL, Petty ITD and Hanley-Bowdoin L. 1994(b). Geminivirus replication origins have a modular organization. *Plant Cell* 6(3): 405-416

Gafni Y and Epel BL. 2002. The role of host and viral proteins in intra- and inter-cellular trafficking of geminiviruses. *Physiological and Molecular Plant Pathology* 60(5): 231-241

Garcia-Andres S, Tomas DM, Sanchez-Campos S, Navas-Castillo J, and Moriones E. 2007. Frequent occurrence of recombinants in mixed infections of tomato yellow leaf curl disease-associated begomoviruses. *Virology* 365(1): 210-219

Ge L, Zhang J, Zhou X and Li H. 2007. Genetic structure and population variability of *Tomato yellow leaf curl China virus*. *Journal of Virology* 81(11): 5902-5907

Ghanim M, Morin S, Zeidan M and Czosnek H. 1998. Evidence for transovarial transmission of tomato yellow leaf curl virus by its vector, the whitefly *Bemisia tabaci*. *Virology* 240(2): 295-303

Gilbertson RL, Hidayat SH, Martinez RT, Leong SA, Faria JC, Morales F and Maxwell DP. 1991. Differentiation of bean-infecting geminiviruses by nucleic-acid hybridization probes and aspects of bean golden mosaic in Brazil. *Plant Disease* 75(4): 336-342

Girish KR and Usha R. 2005. Molecular characterization of two soybean-infecting begomoviruses from India and evidence for recombination among legume-infecting begomoviruses from South-East Asia. *Virus Research* 108(1-2): 167-176

Goodman RM. 1977. Single-stranded DNA genome in a whitefly-transmitted virus. *Virology* 83(1): 171-179

Goodman RM. 1981. Geminiviruses. *Journal of General Virology* 54(1): 9-21

Guillemaut P and Maréchal-Drouard L. 1992. Isolation of plant DNA: a fast, inexpensive, and reliable method. *Plant Molecular Biology Reporter* 10: 60-65.

Guo X and Zhou X. 2005. Molecular characterization of Alternanthera yellow vein virus: a new Begomovirus species infecting Alternanthera philoxeroides. *Journal of Phytopathology* 153(11-12): 694-696

Gutierrez C. 1999. Geminivirus DNA replication. Cellular and Molecular Life Sciences 56(3-4): 313-329

Haber S, Ikegami M, Bajet NB & Goodman RM. 1981. Evidence for a divided genome in bean golden mosaic virus, a geminivirus. *Nature* 289: 324 - 326

Haible D, Kober S and Jeske H. 2006. Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *Journal of Virological Methods* 135(1): 9-16

Hallan V and Gafni Y. 2001. Tomato yellow leaf curl virus (TYLCV) capsid protein (CP) subunit interactions: implications for viral assembly. *Archives of Virology* 146(9): 1765-1773

Thompson HC and Kelly WC. 1957. *Vegetable Crops*. McGraw Hill Book Company, Inc. New York. pp 611

Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S and Robertson D. 2000. Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. *Critical Reviews in Biochemistry and Molecular Biology* 35(2): 105-140.

Harper K and Creamer R. 1995. Hybridization detection of insect-transmitted plant-viruses with digoxigenin-labeled probes. *Plant Disease* 79(6): 563-567

Harrison B and Robinson D. 1999. Natural genomic and antigenic variation in whitefly-transmitted geminiviruses (begomoviruses). *Annual Review of Phytopathology* 37: 369-398

Harrison BD, Barker H, Bock KR, Guthrie EJ, Meredith G, Atkinson M. 1977. Plant viruses with circular single-stranded DNA. *Nature* 270(5639): 760-762

Harrison BD, Swanson MM and Fargette D. 2002. Begomovirus coat protein: serology, variation and functions. *Physiological and Molecular Plant Pathology* 60(5): 257-271

Hartitz MD, Sunter G and Bisaro DM. 1999. The tomato golden mosaic virus transactivator (TrAP) is a single-stranded DNA and zinc-binding phosphoprotein with an acidic activation domain. *Virology* 263(1): 1-14

Hatta T and Francki RIB. 1979. The fine structure of chloris striate mosaic virus. *Virology* 92(2): 428-435

Henson JM and French R. 1993. The polymerase chain-reaction and plant-disease diagnosis. *Annual Review of Phytopathology* 31: 81-109

Henson JM and French R. 1993. The polymerase chain-reaction and plant-disease diagnosis. *Annual Review of Phytopathology* 31: 81-109.

Heyraud-Nitschke F, Schumacher S, Laufs J, Schaefer S, Schell J and Gronenborn B. (1995). Determination of the origin cleavage and joining domain of geminivirus Rep proteins. *Nucleic Acids Research* 23(6): 910-916

Hofer P, Bedford ID, Markham PG, Jeske H and Frischmuth T. 1997. Coat protein gene replacement results in whitefly transmission of an insect nontransmissible geminivirus isolate. *Virology* 236(2): 288-295

Holmes DS and Quingley M. 1981. A rapid boiling method for the preparation of bacterial plasmids. *Annals of Biochemistry* 114: 193-197

Homs M, Kober S, Kepp G and Jeske H. 2008. Mitochondrial plasmids of sugar beet amplified via rolling circle method detected during curtovirus screening. *Virus Research* 136(1-2): 124-129

Horns T and Jeske H. 1991. Localization of abutilon mosaic virus (AbMV) DNA within leaf tissue by in situ hybridization. *Virology* 181(2): 580-588

Hou YM, Sanders R, Ursin VM and Gilbertson RL. 2000. Transgenic plants expressing geminivirus movement proteins: abnormal phenotypes and delayed infection by Tomato mottle virus in transgenic tomatoes expressing the Bean dwarf mosaic virus BV1 or BC1 proteins. *Molecular Plant-Microbe Interactions* 13(3): 297-308

Hull R. 1993. Nucleic acid hybridization procedures. In: *Diagnosis of Plant Virus Diseases* (Matthews REF, eds), CRC Press, Boca Raton, FL. pp 253-271

Hull R. 2009. Plant Virus Origins and Evolution. In: *Comparitive Plant Virology.* Second edition, Elsevier Academic Press, USA. pp 63-80.

Inoue-Nagata AK, Albuquerque LC, Rocha WB and Nagata T. 2004. A simple method for cloning the complete begomovirus genome using the bacteriophage φ29 DNA polymerase. *Journal of Virological Methods* 116(2): 209-211

Isnard M, Granier M, Frutos R ,Reynaud B and Peterschmitt M. 1998. Quasispecies nature of three Maize streak virus isolates obtained through different modes of selection from a population used to assess response to infection of maize cultivars. *Journal of General Virology* 79(12): 3091-3099

Jain R, Rivera MC, and Lake JA. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proceedings of National Academy of Science* 96(7): 3801- 3806

Jeske H, Lutgemeier M and Preiss W. 2001. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. The European Molecular Biology Organization *Journal* 20(21): 6158-6167

Johne R, Muller H, Rector A, van Ranst M and Stevens H. 2009. Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends in Microbiology* 17(5): 205-211

Jones DR. 2003. Plant viruses transmitted by whiteflies. *European Journal of Plant Pathology* 109(3): 195-219

Jose J and Usha R. 2003. Bhendi yellow vein mosaic disease in India is caused by association of a DNA β satellite with a begomovirus. *Virology* 305(2): 310-317

Jupin I, De Kouchkovsky F, Jouanneau F and Gronenborn B. 1994. Movement of *Tomato yellow leaf curl geminivirus* (TYLCV): involvement of the protein encoded by ORF C4. *Virology* 204(1): 82-90

Jupin I, Hericourt F, Benz B and Gronenborn B. 1995. DNA replication specificity of TYLCV geminivirus is mediated by the amino-terminal 116 amino acids of the Rep protein. Federation of European Biochemical Societies *Letters* 362(2): 116-120

Katterman FRH and Shattuck VL. 1983. An effective method of DNA isolation from the mature leaves of *Gossypium* species that contain large amounts of phenolic terpenoids and tannins. *Preparative Biochemistry* 13: 347-359.

Knierim D and Maiss E. 2007. Application of phi29 DNA polymerase in identification and full-length clone inoculation of *Tomato yellow leaf curl Thailand virus* and *Tobacco leaf curl Thailand virus. Archives of Virology* 152(5): 941-954

Kulkarni GS. 1924. Mosaic and other related diseases of crops in the Bombay Presidency. *Poona Agricultural College Magazine* 6: 12

Kumar Y, Hallan V and Zaidi AA. 2008. Molecular characterization of a distinct bipartite begomovirus species infecting tomato in India. *Virus Genes* 37(3): 425-431

Kumari SG, Makkouk KM and Attar N. 2006. An improved antiserum for sensitive serologic detection of Chickpea chlorotic dwarf virus. *Journal of Phytopathology* 154(3): 129-133

Kunik T, Palanichelvam K, Czosnek H, Citovsky V and Gafni Y. 1998. Nuclear import of the capsid protein of tomato yellow leaf curl virus (TYLCV) in plant and insect cells. *Plant Journal* 13(3): 393-399

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21): 2947-2948

Laufs J, Jupin I, David C, Schumacher S, Heyraud-Nitschke F and Gronenborn B. 1995(a). Geminivirus replication: genetic and biochemical characterization of Rep protein function, a review. *Biochimie* 77(10): 765-773

Laufs J, Traut W, Heyraud F, Matzeit V, Rogers SG, Schell J and Gronenborn B. 1995(b). In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proceedings of National Academy of Science* 92(9): 3879-3883

Lazarowitz SG, Wu LC, Rogers SG and Elmer JS. 1992. Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. *Plant Cell* 4: 799-809

Lazarowitz SG. 1992. Geminiviruses: genome structure and gene function. *Critical Reviews in Plant Sciences* 11(4): 327-349

Liu H, Boulton MI, Thomas CL, Prior DA, Oparka KJ and Davies JW. 1999. Maize streak virus coat protein is karyophyllic and facilitates nuclear transport of viral DNA. *Molecular Plant-Microbe Interactions* 12(10): 894-900

Lodhi MA, Ye GN, Weeden NF, Reisch BI. 1994. A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter* 12: 6-13

Loomis WD. 1974. Overcoming problems of phenolics and quinones in the isolation of plant enzymes and organelles. *Methods in Enzymology* 31: 528-544.

Lotrakul P, Valverde RA, Clark CA, Sim J and Torre RDL. 1998. Detection of a geminivirus infecting sweet potato in the United States. *Plant Disease* 82(11): 1253-1257

Lyttle DJ and Guy PL. 2004. First record of Geminiviruses in New Zealand: Abutilon mosaic virus and Honeysuckle yellow vein virus. *Australasian Plant Pathology* 33(2): 321-322

Mandal B, Varma A and Malathi VG. 1997. Systemic Infection of *Vigna mungo* using the cloned DNAs of the blackgram isolate of *Mungbean yellow mosaic geminivirus* through agroinoculation and transmission of the progeny virus by whiteflies. *Journal of Phytopathology* 145(11-12): 505-510

Mansoor S, Amin I, Hussain M, Zafar Y, Bull S, Briddon RW and Markham PG. 2001. Association of a disease complex involving a begomovirus, DNA 1 and a distinct DNA β with leaf curl disease of okra in Pakistan. *Plant Disease* 85(8): 922

Mansoor S, Briddon RW, Zafar Y and Stanley J. 2003. Geminivirus disease complexes: an emerging threat. *Trends in Plant Science* 8(3): 128-134

Mansoor S, Briddon RW, Zafar Y and Stanley J. 2003. Geminivirus disease complexes: an emerging threat. *Trends in Plant Science* 8(3): 128-134

Mansoor S, Khan SH, Bashir A, Saeed M, Zafar Y, Malik KA, Briddon R, Stanley J and Markham PG. 1999. Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. *Virology* 259(1): 190-199

Martin DP, van der Walt E, Posada D and Rybicki EP. 2005. The evolutionary value of recombination is constrained by genome modularity. *PLoS Genetics* 1(4):475-479

Martin RR, James D and Levesque CA. 2000. Impacts of molecular diagnostic technologies on plant disease management. *Annual Review of Phytopathology* 38: 207-239

Maruthi MN, Czosnek H, Vidavski F, Tarba SY, Milo J, Leviatov S, Venkatesh HM, Padmaja AS, Kulkarni RS and Muniyappa V. 2003. Comparison of resistance to *Tomato leaf curl virus* (India) and *Tomato yellow leaf curl virus* (Israel) among Lycopersicon wild species, breeding lines and hybrids. *European Journal of Plant Pathology* 109(1): 1-11

Matthews RE. 1979. The classification and nomenclature of viruses Summary of results of meetings of the International Committee on Taxonomy of Viruses in The Hague, September 1978. *Intervirology* 11(3): 133-135

Michel B, Flores MJ, Viguera E, Grompone G, Seigneur M and Bidnenko V. 2001. Rescue of arrested replication forks by homologous recombination. *Proceedings of National Academy of Science* 98(15): 8181-8188

Moffat A. 1999. Geminiviruses emerge as serious crop threat. *Science* 286 (5446): 1835

Monci F, Sanchez-Campos S, Navas-Castillo J and Moriones E. 2002. A natural recombinant between the geminiviruses *Tomato yellow leaf curl Sardinia virus* and *Tomato yellow leaf curl virus* exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. *Virology* 303(2): 317-326

National Horticulture Board. 2008. Provisional Area and Production Estimates for 2007-08 as approved by Review Committee for comments by States http://nhb.gov.in/

Noris E, Vaira AM, Caciagli P, Masenga V, Gronenborn B and Accotto G P. 1998. Amino acids in the capsid protein of tomato yellow leaf curl virus that are crucial for systemic infection, particle formation, and insect transmission. *Journal of Virology* 72: 10050-10057

Noueiry AO, Lucas WJ and Gilbertson RL. 1994. Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. *Cell* 76(5): 925-932

Novick RP. 1998. Contrasting lifestyles of rolling-circle phages and plasmids. Trends BiochemSci. 23(11): 434-438

Orozco B and Hanley-Bowdoin L. 1996. A DNA structure is required for geminivirus replication origin function. *Journal of Virology* 70: 148-158

Orozco BM, Kong LJ, Batts LA, Elledge S and Hanley-Bowdoin L. 2000. The multifunctional character of a geminivirus replication protein is reflected by its complex oligomerization properties. *Journal of Biological Chemistry* 275(9): 6114-6122

Orozco BM, Miller AB, Settlage SB and Hanley-Bowdoin L. 1997. Functional domains of a geminivirus replication protein. *Journal of Biological Chemistry* 272(15): 9840-9846

Owor BE, Martin DP, Shepherd DN, Edema R, Monjane AL, Rybicki EP, Thomson JA and Varsani A. 2007. Genetic analysis of maize streak virus isolates from Uganda reveals widespread distribution of a recombinant variant. *Journal of General Virology* 88(11):3154-3165

Padidam M, Beachy RN and Fauquet CM. 1995(a). Classification and identification of geminiviruses using sequence comparisons. *Journal of General Virology* 76 (2): 249-263

Padidam M, Beachy RN and Fauquet CM. 1995(b). Tomato leaf curl geminivirus from India has a bipartite genome and coat protein is not essential for infectivity. *Journal of General Virology* 76 (1): 25-35

Padidam M, Beachy RN and Fauquet CM. 1996. The role of AV2 ("precoat") and coat protein in viral replication and movement in tomato leaf curl geminivirus. *Virology* 224(2): 390-404

Padidam M, Sawyer S and Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* 265(2): 218-225

Page RD. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Computer Application in Biosciences* 12(4):357-358

Palmer KE and Rybicki EP. 1998. The molecular biology of Mastreviruses. *Advances in Virus Research* 50: 183-234

Pascal E, Sanderfoot AA, Ward BM, Medville R, Turgeon R and Lazarowitz SG. 1994. The geminivirus BR1 movement protein binds single-stranded DNA and localizes to the cell nucleus. *Plant Cell* 6(7): 995-1006

Paximadis M, Idris AM, Torres-Jerez I, Villarreal A, Rey MEC and Brown JK. 1999. Characterization of tobacco geminiviruses in the Old and New World. *Archives of Virology* 144(4): 703-717

Peterschmitt M, Reynaud B, Sommermeyer G and Baudin P. 1991. Characterization of maize streak virus isolates using monoclonal and polyclonal antibodies and by transmission to a few hosts. *Plant Disease* 75(1): 27-32

Petty ITD, Carter SC, Morra MR, Jeffrey JL and Olivey HE. 2000. Bipartite geminivirus host adaptation determined cooperatively by coding and noncoding sequences of the genome. *Virology* 277(2): 429-438

Pico B, Diez MJ and Nuez F. 1996. Viral diseases causing the greatest economic losses to the tomato crop. II. The *Tomato yellow leaf curl virus* - a review. *Scientia Horticulturae* 67(3): 151-196

Pico B, Diez MJ and Nuez F. 1999. Improved diagnostic techniques for tomato yellow leaf curl virus in tomato breeding programs. *Plant Disease* 83: 1006-1012

Pinner MS and Markham PG. 1990. Serotyping and strain identification of maize streak virus isolates. *Journal of General Virology* 71(8): 1635-1640

Pinner MS, Markham PG, Rybicki EP and Greber RS. 1992. Serological relationships of geminivirus isolates from Gramineae in Australia. *Plant Pathology* 41(5): 618-625

Polston JE, Dodds JA and Perring TM. 1989. Nucleic-acid probes for detection and strain discrimination of cucurbit geminiviruses. *Phytopathology* 79: 1123-1127.

Polston JE, McGovern RJ and Brown LG. 1999. Introduction of tomato yellow leaf curl virus in Florida and implications for the spread of this and other geminiviruses of tomato. *Plant Disease* 83: 984-988

Pooma W, Gillette WK, Jeffrey JL and Petty ITD. 1996. Host and viral factors determine the dispensability of coat protein for bipartite geminivirus systemic movement. *Virology* 218(1): 264-268

Potter JL, Nakha MK, Mejia L and Maxwell DP. 2003. PCR and DNA hybridization methods for specific detection of bean-infecting begomoviruses in the Americas and Caribbean. *Plant Disease* 87(10): 1205-1212

Pun KB and Doraiswamy S. 1999. Effect of age of okra plants on susceptibility to *Okra yellow vein mosaic virus*. *Indian Journal of Virology* 15(1): 57-58

Rampersad SN and Umaharan P. 2003. Detection of begomoviruses in clarified plant extracts: A comparison of standard, direct-binding, and immunocapture polymerase chain reaction techniques. *Phytopathology* 93(9): 1153-1157

Rector A, Tachezy R and Van Ranst M. 2004. A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. *Journal of Virology* 78(10): 4993-4998

Rhee Y, Gurel F, Gafni Y, Dingwall C and Citovsky V. 2000. A genetic system for detection of protein nuclear import and export. *Nature Biotechnology* 18(4): 433-437

Rigden JE, Krake LR, Rezaian MA and Dry IB. 1994. ORF C4 of tomato leaf curl geminivirus is a determinant of symptom severity. *Virology* 204(2): 847-850

Rojas MR, Jiang H, Salati R, Xoconostle-Cazares B, Sudarshana MR, Lucas WJ and Gilbertson RL. 2001. Functional analysis of proteins involved in movement of the monopartite begomovirus, tomato yellow leaf curl virus. *Virology* 291(1): 110-125

Rojas MR, Noueiry AO, Lucas WJ and Gilbertson RL. 1998. Bean dwarf mosaic geminivirus movement proteins recognize DNA in a form- and size-specific manner. *Cell* 95(1): 105-113

Rubio L, Herrero JR, Sarrio J, Moreno P and Guerri J. 2003. A new approach to evaluate relative resistance and tolerance of tomato cultivars to begomoviruses causing the tomato yellow leaf curl disease in Spain. *Plant Pathology* 52(6): 763-769

Rybicki EP. 1994. A phylogenetic and evolutionary justification for 3 genera of Geminiviridae. *Archives of Virology* 139(1-2): 49-77

Sambrook J, Fritsch EF and Maniatis T. 1989. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA. pp 6.3-6.13

Sanderfoot AA, Ingham DJ and Lazarowitz SG. 1996. A viral movement protein as a nuclear shuttle (The geminivirus BR1 movement protein contains domains essential for interaction with BL1 and nuclear localization). *Plant Physiology* 110(1): 23-33

Sastry KSM and Singh SJ. 1974. Effect of yellow vein mosaic virus infection on growth and yield of okra crop. *Indian Phytopathology* 27: 294-297

Saunders K and Stanley J. 1999. A nanovirus-like DNA component associated with yellow vein disease of Ageratum conyzoides: evidence for interfamilial recombination between plant DNA viruses. *Virology* 264(1): 142-152

Saunders K, Bedford ID, Briddon RW, Markham PG, Wong SM and Stanley J. 2000. A unique virus complex causes ageratum yellow vein disease. *Proceedings of National Academy of Science* 97(12): 6890-6895

Saunders K, Salim N, Mali VR, Malathi VG, Briddon R, Markham PG, Stanley J. 2002. Characterization of Sri Lankan cassava mosaic virus and Indian cassavamosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology* 293(1): 63-74

Schubert J, Habekuss A, Kazmaier K and Jeske H. 2007. Surveying cereal-infecting geminiviruses in Germany - diagnostics and direct sequencing using rolling circle amplification. *Virus Research* 127(1): 61-70

Seal SE, van den Bosch F and Jeger MJ. 2006. Factors influencing begomovirus evolution and their increasing global significance: Implications for sustainable control. *Critical Reviews in Plant Sciences* 25(1): 23-46

Settlage SB, Miller AB and Hanley-Bowdoin L. 1996. Interactions between geminivirus replication proteins. *Journal of Virology* 70(10): 6790-6795

Settlage SB, Miller AB, Gruissem W and Hanley-Bowdoin L. 2001. Dual interaction of a geminivirus replication accessory factor with a viral replication protein and a plant cell cycle regulator. *Virology* 279(2): 570-576

Shepherd DN, Martin DP, Lefeuvre P, Monjane AL, Owor BE, Rybicki EP, and Varsani A. 2008. A protocol for the rapid isolation of full geminivirus genomes from dried plant tissue. *Journal of Virological Methods* 149(1): 97-102

Singh SJ. 1980. Studies on Epidemiology of yellow vein mosaic virus of Okra. *Indian Journal of Mycology and Plant Pathology* 10: 35

Stanley J and Townsend R. 1985. Characterisation of DNA forms associated with cassava latent virus infection. *Nucleic Acids Research* 13(7): 2189-2206

Stanley J, Bisaro DM, Briddon RW, Brown JK, Fauquet CM, Harrison BD, Rybicki EP and Stenger DC. 2005. Geminiviridae. In: Virus Taxonomy, VIII Report of the International Committee on Taxonomy of Viruses (CM Fauquet *et al.*, eds), Elsevier Academic Press, London. pp 301-326

Stenger DC, Revington GN, Stevenson MC and Bisaro DM. 1991. Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. *Proceedings of National Academy of Science* 88(18): 8029-8033

Stenger DC, Revington GN, Stevenson MC, and Bisaro DM. 1991. Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. *Proceedings of the National Academy of Sciences* 88(18): 8029-8033

Stonor J, Hart P, Gunther M, DeBarro P and Rezaian MA. 2003. Tomato leaf curl geminivirus in Australia: occurrence, detection, sequence diversity and host range. *Plant Pathology* 52: 379-388

Sunter G and Bisaro DM. 1992. Transactivation of geminivirus AR1 and BR1 gene expression by the viral AL2 gene product occurs at the level of transcription. *Plant Cell* 4(10): 1321-1331

Sunter G, Hartitz MD, Hormuzdi SG, Brough CL and Bisaro DM. 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

Sunter G, Sunter JL and Bisaro DM. 2001. Plants expressing tomato golden mosaic virus AL2 or beet curly top virus L2 transgenes show enhanced susceptibility to infection by DNA and RNA viruses. *Virology* 285(1): 59-70

Sunter, G and Bisaro DM. 1997. Regulation of a geminivirus coat protein promoter by AL2 protein (TrAP): evidence for activation and derepression mechanisms. *Virology* 232(2): 269-280

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22(22):4673-4680

Ueda S, Kimura T, Onuki M, Hanada K and Iwanami T. 2004. Three distinct groups of isolates of Tomato yellow leaf curl virus in Japan and construction of an infectious clone. *Journal of General Plant Pathology* (2004) 70(4):232-238

Unseld S, Hohnle M, Ringel M and Frischmuth T. 2001. Subcellular targeting of the coat protein of African cassava mosaic geminivirus. *Virology* 286(2): 373- 383

Vanitharani R, Challappan P, Pita JS and Fauquet CM. 2004. Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of post-transcriptional gene silencing. *Journal of Virology* 78(17): 9487-9498

Vincze T, Posfai J and Roberts RJ. 2003. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Research* 31(13): 3688-3691

Voinnet O, Pinto YM and Baulcombe DC. 1999. Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. *Proceedings of the National Academy of Sciences* 96(24): 14147-14152

Wang B, Potter SJ, Lin Y, Cunningham AL, Dwyer DE, Su Y, Ma X, Hou Y and Saksena NK. 2005. Rapid and sensitive detection of severe acute respiratory syndrome coronavirus by rolling circle amplification. *Journal of Clinical Microbiology* 43(5): 2339-2444

Ward BM and Lazarowitz SG. 1999. Nuclear export in plantsUse of geminivirus movement proteins for a cell-based export assay. *Plant Cell* 11(7): 1267-1276

Ward BM, Medville R, Lazarowitz SG and Turgeon R. 1997. The geminivirus BL1 movement protein is associated with endoplasmic reticulum-derived tubules in developing phloem cells. *Journal of Virology* 71(5): 3726-3733

Wartig L, Kheyr-Pour A, Noris E, De Kouchkovsky F, Jouanneau F, Gronenborn B and 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

Werner P and Jeske H. 2003. Multitasking in replication is common among geminiviruses. *Journal of Virology* 77(5): 2972-2980

Wu C, Lai Y, Lin N, Hsu Y, Tsai H, Liao J and Hu C. 2008. A simplified method of constructing infectious clones of begomovirus employing limited restriction enzyme digestion of products of rolling circle amplification. *Journal of Virological Methods* 147(2): 355-359

Wyatt SD and Brown JK. 1996. Detection of subgroup III geminivirus isolates in leaf extracts by degenerate primers and polymerase chain reaction. *Phytopathology* 86(12): 1288-1293

Wyatt SD and Brown JK. 1996. Detection of subgroup III geminivirus isolates in leaf extracts by degenerate primers and polymerase chain reaction. Disease Detection and Losses 86(12): 1288-1293

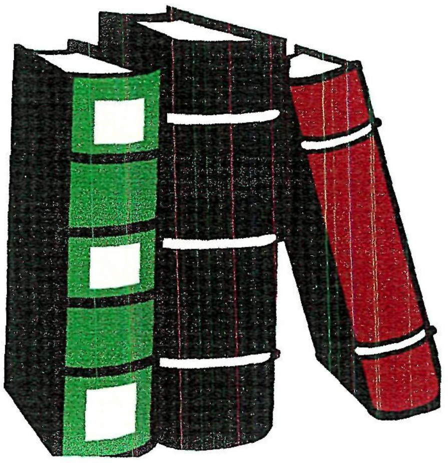
Yadava P, Suyal G and Mukherjee SK. 2010. Begomovirus DNA replication and pathogenicity. *Current Science* 98(3): 360-368

Zhang W, Olson NH, Baker TS, Faulkner L, Agbandje-McKenna M, Boulton MI, Davies JW and McKenna R. 2001. Structure of the maize streak virus geminate particle. *Virology* 279(2): 471-477

Zhou X, Liu Y, Calvert L, Munoz C, Otim-Nape GW, Robinson DJ and Harrison BD. 1997. Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *Journal of General Virology* 78(8): 2101-2111

Zhou X, Liu Y, Robinson DJ and Harrison BD. 1998. Four DNA-A variants among Pakistani isolates of cotton leaf curl virus and their affinities to DNA-A of geminivirus isolates from okra. *Journal of General Virology* 79(4): 915-923

Zhou X, Xie Y, Tao X, Zhang Z, Li Z and Fauquet CM. 2003. Characterization of DNA β associated with begomoviruses in China and evidence for co-evolution with their cognate viral DNA-A. *Journal of General Virology* 84(1): 237-247



Appendices

APPENDIX I

1. Nucleotide sequences submitted in EMBL database

A) The complete DNA-A sequence (2739 bases) of present begomovirus (OYVMV-[IN:Him:09]) (Accession No. FR694925):

```
ACCGGATGGCCGCGCGATTTTTTTTAGTGGTGGGTCCAGAACGCACGACGATGCAGACTCAAAGC
TTAGATAACGCTCCTTTGGCTATAAGTACTTGCGCACTAAGTTTAAATTCAAAACATGTGGGATC
CACTATTAAACGAGTTCCCGGATACGGTTCACGGGTTTCGTTGTATGCTATCTGTAAAATATTTG
CAACTTTTGTCGCAGGATTATTCTCCAGATACGCTTGGGTACGAGTTAATACGGGATTTAATTTG
TATTTTACGCTCCCGTAATTATGTCGAAGCGAGCTGCCGATATCGTCATTTCTACGCCCGCGTCG
AAAGTACGCCGGCGTCTGAACTTCGGCAGCCCATACACCAGCCGTGCTGCTGCCCCCATTGTCCG
CGTCACAAAACAACAGGCATGGACAAACAGGCCTATGAACAGGAAACCCAGAATGTACCGGATGT
ACAGAAGTCCGGATGTTCCACGGGGATGTGAGGGTCCCTGTAAGGTACAGTCGTTTGAATCTCGA
CACGATGTCGTTCATATTGGTAAGGTAATGTGTATTTCGGATGTTACGCGTGGAGTCGGTTTGAC
CCATCGTATAGGTAAGCGTTTTTGTGTCAAGTCAGTTTATGTTTTAGGTAAAATATGGATGGACG
AGAACATCAAGACCAAGAACCATACGAATTCGGTGATGTTTTTCCTTGTTCGTGATCGACGACCG
GTAGATAAACCACAAGATTTTGGTGAAGTATTTAATATGTTTGATAACGAGCCTAGTACGGCGAC
CGTGAAGAACATGCATAGGGACCGGTACCAGGTGTTGAGGAAATGGCATGCAACCGTTACTGGTG
GACAATATGCGAGTAAGGAGCACGCTTTGGTCAAGAAGTTTGTTAGGGTTAACAACTACGTTGTT
TACAACCAGCAGGAAGCAGGAAAATACGAGAATCACACCGAGAATGCATTGATGCTTTACATGGC
TTGTACTCATGCTAGCAACCCAGTGTATGCTACTCTTAAGATTAGAATATATTTTTATGACTCTG
TAACGAACTAATATTAATAAAGTTTAAATTGTATATCTGAATATTGGTCTACATACATTGTTTTA
TTAATTACATTGTACAATACATGTTCAACGGCTTTAATAACTAAATTAATTGAGATTACACCTAG
ATTGTTGAGATATTTGAGGACTTGGGTTTTGAATACCCTTAAGAAAAGACCAGTCGGAGGGTGTA
AGGTCGTCCAGATTCGGAAGGTTAGAAAACACTTGTGTATTCCCAGAGCTTTCCGTAGGTTGTAG
TTGAAATGGATCCTGATTGTTATTATGTCCATGTTCGTCGTGAATGGACGGTTGTCGTGGTTGAG
GATCTTGAAATAGAGGGGATTTGGAACCTTCCAGATATAGACGCCATTCTTTGCCTGAGCTGCAG
TGATGCGTTCCCCTGTGCGAGAATCCATGGTTGTGGCAGTTAATTGAAAGATAATAAGAACACCC
GCATTCAAGATCTACTCTCCTCCTCCTGTTGCGTCTCTTCGCTTCCCTGTGCTGTACTTTGATTG
GTACCTGAGTACAGTGGTTCTTCGAGGGTGATGAAGATCGCATTCTTGACTGCCCAGTTCTTTAG
TGCGTTGTTCTTTTCCTCGTCCAGGAATTCTTTATAACTGCTGTTGGGACCAGGATTGCAGAGGA
AGATTGTTGGTATTCCGCCTTTAATTTGAACTGGCTTCCCGTACTTTGTATTAGATTGCCAGTCC
CTTTGGGCCCCCATGAACTCTTTAAAGTGCTTGAGGAAGTGCGGATCAACGTCATCAATGACGTT
ATACCAAGCGTCGTTACTGTATACCTTTGGACTTAGATCTAAATGCCCACATAAATAGTTATGTG
```

GGCCTAAAGACCTAGCCCACATTGTTTTCCCAGTACGACTGTCACCCTCAATTACTATACTTTGA
GGTCTCAGGGGCCGCGCAGCGGCGTCGACAACATTCTCAGACACCCACTCTTCAAGTTCTTCTGG
AACTTGATCGAAAGAAGAAGAGGAAAAAGGAGAAACATAAGGAGCTGGAGGCTCCTGAAAGATCC
TGTCTAGATTTGCATTTAAATTATGAAATTGTAGTACAAAATCTTTAGGAGCTAGTTCCTTAATG
ACTCTAAGAGACTCTGACTTACTTCCCGCGTTAAGTGCTGCGGCGTAAGCGTCGTTGGCTGTTTG
TTGCCCTCCTCTTGCTGATCTTCCGTCGATCTGAAATTCGCCCCAGTCGAGAATGTCCCCGTCCT
TAGCGATGTAGGACTTGACGTCGGAGCTGGATTTAGCTCCCTGAATGTTTGGATGGAAATGTGCT
GACCTGGTTGGGGATATGAGGTTGAAGAATCGGTTATTTTTGCATTGGTATTTGCCCTCGAACTG
GATGAGCACGTGAAGATGAGGTTCCCCATTTTCATGAAACTCTCTGCAGATTCTAATGTATTTTT
TGTTTACTGGTGTCTGTAGGTTTTGTAATTGGGAAAGTGCTTCATCTTTAGTAAGAGAGCAAGTG
GGATAAGTGAGGAAATAATTTTTGGCATATATCTGAAATTGTTTGGGAGGAGCCATTGACTTGGT
CAATCGGTACCCAGATCTAGTCTTATGTCAATTGGTGAACGGTACCCTATATATAGTGGGTACTG
AATGGCAATATTTGTAATTATGAAAATAAATTCAAAATCCTCACGCTCCAAAAAGCGGCCATCCG
TATAATATT

B) The complete betasatellite sequence (1332 bases) associated with present begomovirus (OYVB-[IN:Him:09]) (Accession No. FR823510):

ACCGGATGGCCGCGCGATTTTTTTAAGTGGTTGGTCCAGATCTCGCGCGATTTTTTCCCAGAACT
CGCGCGATTCTCTTAAGTGGTGGGTCCAGAACTCGCGCGATTTTTTTCCAACACTCGCGCGATTT
CTTTAAGTGGTGGGTCCAGAACTCACCACGATGCAGACCACTTATGAATGAAGTTTATGGGTGAT
TTCAATTATATGGAGGAAATTGTGGATGAGAAAAATGAATCATGTTTGTTTTTTGCAAATTATTT
TAGATAACAGTCTCCTAATAATAATTAATATGCAAACATATTACTAACAAAAATTAAATTATTAT
CTTATTATCAATAGTTACTGGTTCGTTTACATCCATTCCCATAATTTCTGGGTTTTCAATAATAA
TAATATCCACCATATGTATGATGTCTTCTTCTAATATTTCTTCTGCTTTTGAACCCCAGTAATGA
AATTTGAAGGTTGATGTGATCGTTCCTTCTAAACCATTGAAGTCGAATGGAATGTGTAATTGTTG
GTATGTGTAGTCGATGATGAATGTTTTTGTTGATATTGCTGGAGATCTTGTTGAACTGATCTGCA
TCAGGACTGATATCCTTTGTTGTTGCATGAGCCTGACATCTACTTTGAAGACGATCCCCTCCCTG
TTTTTGCTGAATCTTGTCATTTTGATTGTGCGTCTAATGTATATGTTTGAAATGAATATATAGAT
GAAATGTATAGATATTTTCATATAATGTGTGGATATCATTTTAAATGATATATATGGTTCTTGTT
TGTTGGTACATAATATCTGTTTACTAATGGAGAATAAGGTATTATATATTTATATGTAGGTGTAT
GTATAATTTGAAGAAAGAAAAATATTATGTTATAGATATACTAGTATATTTTTAGGATAAGAAAA
AGAAAAGAGGAAATAGAGGAAAAATGAAACCAGAAAGAAATGAAATATAGATAAACTATGAAAAA
AGAGAAGGGAGCGCAGCGAGAAAAAAAGAATCGGAAAGGAAAAAATAACAAAATCAAATAAAAAA
AATTCACAAAGAAAAAAAAGGAAAATCCGACGTCAAAATGAAGCCGTTTTGAGGAGGAAGAGGAA
AGAAAACGATAAATTTAATACTCTCTGTATATTTTAATATATTTTTTGATTATTGGGTTTATTTT
GTAATTGGGAAAGTGCCTCCTTGACTTGGTCAATCGGTACCCTATATATAGTGGGTACTGAATGG
CAATATTGTAATTATATAAATAAAATCAAAATCCTCACGCTCCAAAAAGTAAAATCAAAATCCTC
ACGCTCCAAAAAGCGGCCATCCGTATAATATT

2. ORF sequences of present begomovirus submitted in EMBL database

A) V1 (Accession no. CBX25663):

```
MSKRAADIVISTPASKVRRRLNFGSPYTSRAAAPIVRVTKQQAWTNRPMNRKPRMYRMYRSPDVPRGCEG
PCKVQSFESRHDVVHIGKVMCISDVTRGVGLTHRIGKRFCVKSVYVLGKIWMDENIKTKNHTNSVMFFLV
RDRRPVDKPQDFGEVFNMFDNEPSTATVKNMHRDRYQVLRKWHATVTGGQYASKEHALVKKFVRVNNYVV
YNQQEAGKYENHTENALMLYMACTHASNPVYATLKIRIYFYDSVTN
```

B) V2 (Accession no. CBX25662):

```
MWDPLLNEFPDTVHGFRCMLSVKYLQLLSQDYSPDTLGYELIRDLICILRSRNYVEASCRYRHFYARVES
TPASELRQPIHQPCCCPHCPRHKTTGMDKQAYEQETQNVPDVQKSGCSTGM
```

C) C1 (Accession no. CBX25666):

```
MAPPKQFQIYAKNYFLTYPTCSLTKDEALSQLQNLQTPVNKKYIRICREFHENGEPHLHVLIQFEGKYQC
KNNRFFNLISPTRSAHFHPNIQGAKSSSDVKSYIAKDGDILDWGEFQIDGRSARGGQQTANDAYAAALNA
GSKSESLRVIKELAPKDFVLQFHNLNANLDRIFQEPPAPYVSPFSSSSFDQVPEELEEWVSENVVDAAAR
PLRPQSIVIEGDSRTGKTMWARSLGPHNYLCGHLDLSPKVYSNDAWYNVIDDVDPHFLKHFKEFMGAQRD
WQSNTKYGKPVQIKGGIPTIFLCNPGPNSSYKEFLDEEKNNALKNWAVKNAIFITLEEPLYSGTNQSTAQ
GSEETQQEEESRS
```

D) C2 (Accession no. CBX25665):

```
MRSSSPSKNHCTQVPIKVQHREAKRRNRRRRVDLECGCSYYLSINCHNHGFSHRGTHHCSSGKEWRLYLE
GSKSPLFQDPQPRQPSIHDEHGHNNNQDPFQLQPTESSGNTQVFSNLPNLDDLTPSDWSFLKGIQNPSPQ
ISQQSRCNLN
```

E) C3 (Accession no. CBX25664):

```
MDSRTGERITAAQAKNGVYIWKVPNPLYFKILNHDNRPFTTNMDIITIRIHFNYNLRKALGIHKCFLTFR
IWTTLHPPTGLFLRVFKTQVLKYLNNLGVISINLVIKAVEHVLYNVINKTMYVDQYSDIQFKLY
```

F) C4 (Accession no. CBX25667):

```
MGNLIFTCSSSSRANTNAKITDSSTSYPQPGQHISIQTFRELNPAPTSSPTSLRTGTFSTGANFRSTEDQ
QEEGNKQPTTLTPQHLTREVSQSLLESLRN
```

G) BetaC1 (Accession no. CBZ47002):

```
MISTHYMKISIHFIYIFISNIYIRRTIKMTRFSKNREGIVFKVDVRLMQQQRISVLMQISSTRSPAISTKTFI
IDYTYQQLHIPFDFNGLEGTITSTFKFHYWGSKAEEILEEDIIHMVDIIIIENPEIMGMDVNEPVTIDNKIII
```

APPENDIX II

1. Multiple Alignments of all ORF sequences of present begomovirus with the other closely related begomoviruses from Indian sub-continent

A) Multiple alignment of AC1 ORF sequences. The conserved amino acid residues were represented by ‘*’.

```
OYVMV-[IN:Him:09]_CBX25666_AC1       MAPPKQFQIYAKNYFLTYPTCSLTKDEALSQLQNLQTPVNKKYIRICREF 50
OYVMV-[PK:Fai201:95]_CAA05440_AC1    MAPPKQFQIYAKNYFITYPTCSLTKKEALSQLQNLQTPVNKKYIRICREF 50
OYVMV-[IN:Abd:08]_ACY78674_AC1       MAAPHRFKINAKIYFLTYPKCSLRKEKALSQFLNLQTPVNKKSIRICREL 50
BYVMV-[IN:OY177:07]_ACN51283_AC1     MAPPKKFQINAKNYFITYPTCSLTKEEALSQLQNLQTPVNKKYIKICREF 50
CLCuMV-[IN:Lud:04]_AAX84764_AC1      --MPRDLKTNAKNYFLTYPKCSLTKEETLSQLLNLHTPVNKKYIKICREL 48
BYVMV-[IN:THD:05]_ACN42646_AC1       -MPPKRFVINSKNYFLTYPQCSLNKEEALSQLQNIQTPTNKKYIKICREL 49
BYVMV-[IN:PAN:05]_ACN42653_AC1       -MPSKRFQIYSKNYFLTYPKCSLTKEEALSQLQNLQTPANKKFIRICREL 49
OkLCuIV-[IN:ND:08]_ADG29347_AC1      -MPSKRFQIYSKNYFLTYPKCSLTKEEALSQIQNLQTPTNKKFIKICKEL 49
BYVMV-PK[PK:M301:96]_CAA05452_A      -MPPKRFLINSKNYFLTYPKCSLTKEEALSQIQNLQTPVNKKYIKICREF 49
BYVMV-IN[IN:Mad:01]_AAF63746_AC1     -MPPRRFVINSKNYFLTYPQCSLNKEEALSQLQNIQTPTNKKYIKICSEL 49

OYVMV-[IN:Him:09]_CBX25666_AC1       HENGEPHLHVLIQFEGKYQCKNNRFFNLISPTRSAHFHPNIQGAKSSSDV 100
OYVMV-[PK:Fai201:95]_CAA05440_AC1    HENGEPHLHVLIQFEGKYKCKNNRFFDLVSPTRSAHFHPNIQGVKSSSDV 100
OYVMV-[IN:Abd:08]_ACY78674_AC1       HEDGNPQLHVLIQFEGKFQCRNNRFFDLTSPTRSAHFHPNIQGAKSSSDV 100
BYVMV-[IN:OY177:07]_ACN51283_AC1     HENGEPHLHVLIQFEGKFKCKNYRFFDLVSPTRSAHFHPNIQGAKSSSDV 100
CLCuMV-[IN:Lud:04]_AAX84764_AC1      HQNGEPHLHVLIQFEGKYQCTNKRFFDLVSPTRTAHFHPNIQGAKSSSDV 98
BYVMV-[IN:THD:05]_ACN42646_AC1       HENGEPHLHVLIQFEGKYKCQNQRFFDLVATNRSAHFHPNIQGAESSSDV 99
BYVMV-[IN:PAN:05]_ACN42653_AC1       HEDGDPHLHVLIQFEGKYQCRNNRFFDLTPPSRSAHFHPNIQGAKSSSDV 99
OkLCuIV-[IN:ND:08]_ADG29347_AC1      HENGEPHLHVLIQFEGKYKCQNQRFFDLVSPSRSAHFHPNIQGAKSSSDV 99
BYVMV-PK[PK:M301:96]_CAA05452_A      HEDGEPHLHVLIQFEGKYKCQNQRFFDLVSPTRSAHFHPNIQGAKSSSDV 99
BYVMV-IN[IN:Mad:01]_AAF63746_AC1     HENGEPHLHVLIQFEGKYKCQNQRFFDLVATNRSAHFHPNIQGAKSSSDV 99

OYVMV-[IN:Him:09]_CBX25666_AC1       KSYIAKDGDILDWGEFQIDGRSARGGQQTANDAYAAALNAGSKSESLRVI 150
OYVMV-[PK:Fai201:95]_CAA05440_AC1    KSYIAKDGDILDWGEFQIDGRSARGGQQTANDAYAAALNAGSKSEALRVI 150
OYVMV-[IN:Abd:08]_ACY78674_AC1       KTYIEKDGDILDHGVFQVDARSARGGCQTANDAYAEAINAGSKAQALNIL 150
BYVMV-[IN:OY177:07]_ACN51283_AC1     KSYIDKDGDILEWGEFQIDGRSARGGQQTANDAYAAALNAGSKSEALRVI 150
CLCuMV-[IN:Lud:04]_AAX84764_AC1      KAYIDKDGDTLGWGEFQIDGRSARGGRQTANDAYAAALNAGNKSEALRVI 148
BYVMV-[IN:THD:05]_ACN42646_AC1       KSYIDKDGDTLEWGGFQIDGRSARGGQQSANDAYAAALNTGSKAEALRVI 149
BYVMV-[IN:PAN:05]_ACN42653_AC1       KSYIDKDGDTLEWGEFQIDGRSARGGQQSANDAYAAALNTGSKAEALRVL 149
OkLCuIV-[IN:ND:08]_ADG29347_AC1      KSYIDKDGDTLEWGEFQIDGRSARGGQQTANDAYAAALNAGSKSEALRVI 149
BYVMV-PK[PK:M301:96]_CAA05452_AC1    KSYLEKDGDTIDWGGFQIDGRSARGGQQSANDAYAAALNAGSKSEALRVI 149
BYVMV-IN[IN:Mad:01]_AAF63746_AC1     KSYIDKDGDTLEWGEFQIDGRSARGGQQSANDAYAAALNTGSKAEALRVL 149

OYVMV-[IN:Him:09]_CBX25666_AC1       KELAPKDFVLQFHNLNANLDRIFQEPPAPYVSPFSSSSFDQVPEELEEWV 200
OYVMV-[PK:Fai201:95]_CAA05440_AC1    KELAPKDFVLQFHNLNANLDRIFQEPPAPYVSPFSSSSFDQVPEELEEWV 200
OYVMV-[IN:Abd:08]_ACY78674_AC1       KEKAPRDFLLQFHNLNSNLDRYFQEPPAPYVSPFSSSSFDQVPEELEEWA 200
BYVMV-[IN:OY177:07]_ACN51283_AC1     KELAPKDFVLQFHNLNANLDRIFQEPPAPYVSPFSSSSFDQVPEELEEWA 200
CLCuMV-[IN:Lud:04]_AAX84764_AC1      KELAPKDFVLQFHNLNANLDRIFQEPPAPYISPFSSSSFDQVPEELEVWA 198
BYVMV-[IN:THD:05]_ACN42646_AC1       KEIAPTDFQPQFHDLDPNLHRYFQEPPAPYVTPFSSSSFDQVPEELEEWA 199
BYVMV-[IN:PAN:05]_ACN42653_AC1       KELAPKDYVLQYHNLNSNLDRIFTPPQEVYVCPFLFSSFDQVSEEPEEWA 199
OkLCuIV-[IN:ND:08]_ADG29347_AC1      RELAPKDYVLQFHNLNANLDRIFTPPLEVYVSPFLSSSFDQVPEELEEWV 199
BYVMV-PK[PK:M301:96]_CAA05452_AC1    KELAPKDYVLQFHNLNANLDRIFTPPMEVYVSPFSSSSFDQVPEELEEWA 199
BYVMV-IN[IN:Mad:01]_AAF63746_AC1     KELAPKDYVLQYHNLNSNLDRIFTPPQEVYVCPFLSSSFDQVPEELEEWA 199

OYVMV-[IN:Him:09]_CBX25666_AC1       SENVVDAAARPLRPQSIVIEGDSRTGKTMWARSLGPHNYLCGHLDLSPKV 250
OYVMV-[PK:Fai201:95]_CAA05440_AC1    SENVVDAAARPLRPQSIVIEGDSRTGKTMWARSLGPHNYLCGHLDLSPKV 250
```

```
OYVMV-[IN:Abd:08]_ACY78674_AC1        CENVVDAAARPLRPQSIVIEGDSRTGKTMWARSLGPHNYLCGHLELSPKV 250
BYVMV-[IN:OY177:07]_ACN51283_AC1      CENVVAAAARPLRPQSIVIEGDSRTGKTMWARSLGPHNYLCGHLDLSPKA 250
CLCuMV-[IN:Lud:04]_AAX84764_AC1       IDNVVDPAARPLRPRSIVIEGDSRTGKTMWARSLGPHNYLCGHLDLSPKV 248
BYVMV-[IN:THD:05]_ACN42646_AC1        CKNFVNAAARPLRPQSIAIEVDTRTGKTMWARSLGPHNYLCGHLDLSPKV 249
BYVMV-[IN:PAN:05]_ACN42653_AC1        SKNVMDAAARPQRPLSLVLEGDSRTGKTMWARSLGPHNYLCGHLDLSPKV 249
OkLCuIV-[IN:ND:08]_ADG29347_AC1       SENVMDAAARPLRPLSLVLEGDSRTGKTMWARSLGPHNYLCGHLDLSPKV 249
BYVMV-PK[PK:M301:96]_CAA05452_AC1     AENVVSAAARPLRPLSLVLEGESRTGKTQWARSLGPHNYLCGHLDLSPKV 249
BYVMV-IN[IN:Mad:01]_AAF63746_AC       SENVMDAAARPQRPLSLVLEGESRTGKTQWARSLGPHNYLCGHLDLSPKV 249

OYVMV-[IN:Him:09]_CBX25666_AC1        YSNDAWYNVIDDVDPHFLKHFKEFMGAQRDWQSNTKYGKPVQIKGGIPTI 300
OYVMV-[PK:Fai201:95]_CAA05440_AC1     YSNDAWYNVIDGVDPHFLKHFKEFMGAQRDWQSNTKYGKPVQIKGGIPTI 300
OYVMV-[IN:Abd:08]_ACY78674_AC1        YSNDAWYNVIDDVDPHFLKHFKEFMGAQRDWQSNTKYGKPVQIKGGIPTI 300
BYVMV-[IN:OY177:07]_ACN51283_AC1      YSNDAWYNVIDDVDPHFLKHFKEFMGAQRDWQSNTKYGKPVQIKGGIPTI 300
CLCuMV-[IN:Lud:04]_AAX84764_AC1       YSNDAWYNVIDDVDPHFLKHFKEFMGAQRDWQSNTKYGKPVQIKGGIPTI 298
BYVMV-[IN:THD:05]_ACN42646_AC1        YSNDAWFNIIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPVQIKGGIPTI 299
BYVMV-[IN:PAN:05]_ACN42653_AC1        YSNDAWFNIIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPVQIKGGIPTI 299
OkLCuIV-[IN:ND:08]_ADG29347_AC1       YSNDAWFNIIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPVQIKGGIPTI 299
BYVMV-PK[PK:M301:96]_CAA05452_AC1     YSNDAWFNIIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPVQIKGGIPTI 299
BYVMV-IN[IN:Mad:01]_AAF63746_AC1      YSNDAWFNIIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPVQIKGGIPTI 299

OYVMV-[IN:Him:09]_CBX25666_AC1        FLCNPGPNSSYKEFLDEEKNNALKNWAVKNAIFITLEEPLYSGTNQSTAQ 350
OYVMV-[PK:Fai201:95]_CAA05440_AC1     FLCNPGPNASYKEYLDEEKNSALKAWALKNAAFVFLTQPLYSGTNQSTAQ 350
OYVMV-[IN:Abd:08]_ACY78674_AC1        FLCNPGPNASYKEFLDEDKNSALKSWALKNATFVFLTQPLYSGTNQSSTQ 350
BYVMV-[IN:OY177:07]_ACN51283_AC1      FLCNPGPNASYKEFLDEEKNSALKAWALKNAAFVFLTEPLYSGTNQSTAQ 350
CLCuMV-[IN:Lud:04]_AAX84764_AC1       FLCNPGPNSSYKEFLDEERNTALTNWAVKNAIFITLEGPLYSGTNQSTAQ 348
BYVMV-[IN:THD:05]_ACN42646_AC1        FLCNPGPNASYKEFLDEEKNSALKAWALKNAAFVFLTEPLYSGTNQSTAQ 349
BYVMV-[IN:PAN:05]_ACN42653_AC1        FLCNPGPNASYKEFLDEEKNSALKAWALKNAAFVFLTEPLYSGTNQSTAQ 349
OkLCuIV-[IN:ND:08]_ADG29347_AC1       FLCNPGPNASYKEFLEEEKNSALKAWALKNAAFVFLTQPLYSGTNQSSTQ 349
BYVMV-PK[PK:M301:96]_CAA05452_AC1     FLCNPGPNASYKEFLDEEKNSALKAWTLKNATFVFLTQPLYSGTNQSSTQ 349
BYVMV-IN[IN:Mad:01]_AAF63746_AC1      FLCNPGPNASYKEFLDEEKNSALKAWALKNAAFVFLTQPLYSGTNQSSTQ 349

OYVMV-[IN:Him:09]_CBX25666_AC1        GSEETQQEEESRS 363
OYVMV-[PK:Fai201:95]_CAA05440_AC1     GSEETQQEEESRS 363
OYVMV-[IN:Abd:08]_ACY78674_AC1        GSEESAQEETSGP 363
BYVMV-[IN:OY177:07]_ACN51283_AC1      GSEETHQEEESRS 363
CLCuMV-[IN:Lud:04]_AAX84764_AC1       GSEETQQEEESRS 361
BYVMV-[IN:THD:05]_ACN42646_AC1        GSEETHQEEESRS 362
BYVMV-[IN:PAN:05]_ACN42653_AC1        GSEETHQEEESRS 362
OkLCuIV-[IN:ND:08]_ADG29347_AC1       GSEESAQEETSGP 362
BYVMV-PK[PK:M301:96]_CAA05452_AC1     GSEESAQEETSGP 362
BYVMV-IN[IN:Mad:01]_AAF63746_AC1      GSEASAQEETSGP 362
```

B) Multiple alignment of AC2 ORF sequences. The conserved amino acid residues were represented by '*'.

```
OYVMV-[IN_Him_09]_CBX25665_AC2        MRSSSPSKNHCTQVPIKVQHREAKRRNRRRRVDLECGCSYYLSINCHNHG 50
OYVMV-[PK:Fai201:95]_CAA05441_AC2     MQHSSFSPNRCTQVPIKVQHREAKRRNRRRRVDLECGCFYYLSINCHNHG 50
OYVMV-[IN:Abd:08]_ACY78673_AC2        MQHSSFSPNRSTPVPIKVQHREAKRAHRRKRVDLECGCSYYLSINCHNHG 50
BYVMV-[IN:OY177:07]_ACN51282_AC2      MQHSSFSPNRCTRVPIKVQHREAKRRIRRRRVDLECGCSYYLSINCHNHG 50
CLCuMV-[IN:Lud:04]_AAX84765_AC2       MRSSSLSKDPCTQVPIKVQHREAKRRNRRRRVDLECGCSYYLSINCHNHG 50
BYVMV-[IN:THD:05]_ACN42645_AC2        MQHSSFSPNRCTRVPIKVQHREAKRRIRRRRVDLECGCSYYLSINCHNHG 50
BYVMV-[IN:PAN:05]_ACN42652_AC2        MQHSSFSPNRCTRVPIKVQHREAKRRIRRRRVDLECGCSYYLSINCHNHG 50
OkLCuIV-[IN:ND:08]_ADG29348_AC2       MQHSSSSPNRCTRVPIKVQHREAKRAHRRKRVDLECGCSYYLSINCHNHG 50
BYVMV-PK[PK:M301:96]_CAA05453_AC2     MQHSSFSPNRSTPVPIKVQHREAKRAHRRKRVDLECGCSYYLSINCHNHG 50
BYVMV-IN[IN:Mad:01]_AAF63747_AC2      MQHSSFSPNRCTRVPIKVQHKEAKRVHRRRRVDLECGCSYYLSINCHNHG 50

OYVMV-[IN_Him_09]_CBX25665_AC2        FSHRGTHHCSSGKEWRLYLEGSKSPLFQDPQPRQPSIHDEHGHNNNQDPF 100
OYVMV-[PK:Fai201:95]_CAA05441_AC2     FSHRGTHHCSSSKEWRLYLEGSKSPLFQDPQPRQPSIHDEHGHNNTQDPF 100
OYVMV-[IN:Abd:08]_ACY78673_AC2        FPHRGTHHCSSFNEWRLYLGGSKSALFQNPQPRQPSIHDEHGHHNNEDPI 100
BYVMV-[IN:OY177:07]_ACN51282_AC2      FSHRGTHHCSSSKEWRLYLEGSKSPLFQDPQPRQPSIHDEHGHNNTQDPF 100
CLCuMV-[IN:Lud:04]_AAX84765_AC2       FTHRGTHHCSSSREWRIYLGGSKSPLFQDHQPRQPSIHDEYGYTLDPDPV 100
BYVMV-[IN:THD:05]_ACN42645_AC2        SSHRGTHHCSSSKEWRLYLEGSKSPLFQDPQPRQPSIHDEHGHNNTQDPF 100
BYVMV-[IN:PAN:05]_ACN42652_AC2        FSHRGTHHCSSSKEWRLYLEGSKSPLFQDPQPRQPSIHDEHGHNNTQDPF 100
OkLCuIV-[IN:ND:08]_ADG29348_AC2       FSHRGTHHCSSFNEWRLYLGGSKSPLFQNPQPRQPSIHDEHGHHNNEDPI 100
BYVMV-PK[PK:M301:96]_CAA05453_AC2     FSHRGTHHCSSFNEWRLYLGGSKSPLFQNPQPRQPSIHDEHGHHNNEDPI 100
BYVMV-IN[IN:Mad:01]_AAF63747_AC2      FSHRGTHHCSSSKEWRLYLEGSKSPLFQDPQPRQPSIHDEHGHNNTQDPF 100
```

```
OYVMV-[IN_Him_09]_CBX25665_AC2        QLQPTESSGNTQVFSNLPNLDDLTPSDWSFLKGIQNPSPQISQQSRCNLN 150
OYVMV-[PK:Fai201:95]_CAA05441_AC2     QLQPTESSGNTQVFSNLPNLDDLTPSDWSFLKGIQNPSPQTSQQSRCNLN 150
OYVMV-[IN:Abd:08]_ACY78673_AC2        QLQPSESSGSAQVFSNLPNLDDLTPSDWSFLKGIQNPSTQVSQQSKCNLT 150
BYVMV-[IN:OY177:07]_ACN51282_AC2      QLQPTESSGNTQVFSNLPYLDDLTPSDWSFLKGIQNPSPQISQQSRCSFN 150
CLCuMV-[IN:Lud:04]_AAX84765_AC2       QLQHSESSGTAQVFSNLPNLDDLTPSDWSFLKGIQNQSPQISEQSRCNFN 150
BYVMV-[IN:THD:05]_ACN42645_AC2        QLQPTESSGNTQVFSNLPNLDDLTPSDWSFLKGVQNPSPQISQQSRCSFN 150
BYVMV-[IN:PAN:05]_ACN42652_AC2        QLQPTESSGNTQVFSNLPNLDDLTPSDWSFLKGIQNPSPQISQQSRCSFN 150
OkLCuIV-[IN:ND:08]_ADG29348_AC2       QLQPSESSGSAQVFSNLPNLDDLTPSDWSFLKGIQNPSPQISQQSRCNLN 150
BYVMV-PK[PK:M301:96]_CAA05453_AC2     QLQPSESSGSAQVFSNLPNLDDLTPSDWSFLKGIQNPSPQISQQSRCNLT 150
BYVMV-IN[IN:Mad:01]_AAF63747_AC2      QLQPTESSGNTQVFSNLPNLDDLTPSDWSFLKGIQNPSPQISQ------- 143
                                      *** :****.:******* **************:** *.* *:
```

C) Multiple alignment of AC3 ORF sequences. The conserved amino acid residues were represented by '*'.

```
OYVMV-[IN_Him_10]_CBX25664_AC3        MDSRTGERITAAQAKNGVYIWKVPNPLYFKILNHDNRPFTTNMDIITIRI 50
OYVMV-[PK:Fai201:95]_CAA05442_AC3     MDSRTGERITAAQAKNGVYIWKVPNPLYFKILNHDNRPFTTNMDIITLRI 50
OYVMV-[IN:Abd:08]_ACY78672_AC3        MDSRTGERITAAHSTNGVFIWEVPNPLYFKILSHDNRPFTTNMDIITMRI 50
BYVMV-[IN:OY177:07]_ACN51281_AC3      MDSRTGARITAAQAKNGVYIWKVPNPLYFKILNHDNRPFTTNMDIITLRI 50
CLCuMV-[IN:Lud:04]_AAX84766_AC3       MDSRTGEPITAAQAGNGAYIWEVPNPLYFKIISHVNRPFTTNMDILSIRI 50
BYVMV-[IN:THD:05]_ACN42644_AC3        MDPRTGARITAAQAKNGVYIWKVPNPLYFKILNHDNRPFTTNMDIITLRI 50
BYVMV-[IN:PAN:05]_ACN42651_AC3        MDSRTGARITAAQAKNGVYIWKVPNPLYFKILNHDNRPFTTNMDIITLRI 50
OkLCuIV-[IN:ND:08]_ADG29349_AC3       MDSRTGERITAAHSTNGVFIWEVPNPLYFKILNHDNRPFTTNMDIITMRI 50
BYVMV-PK[PK:M301:96]_CAA05454_AC3     MDSRTGERITAAHSTNGVFIWEVPNPLYFKILSHDNRPFTTNMDIITMRI 50
BYVMV-IN[IN:Mad:01]_AAF63748_AC3      MDSRTGERITAAQAKNGVYIWKVPNPLYFKILNHDNRPFTTNMDIITLRI 50
                                      **.***  ****:: **.:**:*********:.* **********:::**

OYVMV-[IN_Him_10]_CBX25664_AC3        HFNYNLRKALGIHKCFLTFRIWTTLHPPTGLFLRVFKTQVLKYLNNLGVI 100
OYVMV-[PK:Fai201:95]_CAA05442_AC3     HFNYNLRKALEIHKCFLTFRIWTTLHPPTGLFLRVFKTQVLKHLNNLGVI 100
OYVMV-[IN:Abd:08]_ACY78672_AC3        QFNYNLRKALGVHKCFLTFRIWTTLHPPTGLFLRVFKTQVLKYLNNLSVI 100
BYVMV-[IN:OY177:07]_ACN51281_AC3      HFNYNLRKALGIHKCFLTFRIWTTLHPQTGLFLRVFKTQVLKYLNNLGVV 100
CLCuMV-[IN:Lud:04]_AAX84766_AC3       QFNYNTRKALGLHKCFLTFRIWTTLHPQTGLFLRVFKTKVLKYLNNLGVI 100
BYVMV-[IN:THD:05]_ACN42644_AC3        HFNYNLRKALGIHKCFLTFRIWTTLHPPTGLFLRVFKTQVLKYLNNLGVV 100
BYVMV-[IN:PAN:05]_ACN42651_AC3        HFNYNLRKALGIHKCFLTFRIWTTLHPQTGLFLRVFKTQVLKYLNNLGVV 100
OkLCuIV-[IN:ND:08]_ADG29349_AC3       QFNYNLRKALGVHKCFLTFRIWTTLHPPTGLFLRVFKTQVLKYLNNLGVI 100
BYVMV-PK[PK:M301:96]_CAA05454_AC3     QFNYNLRKALGVHKCFLTFRIWTTLHPPTGLFLRVFKTQVLKYLNNLGVI 100
BYVMV-IN[IN:Mad:01]_AAF63748_AC3      HFNYNLRKALGIHKCFLTFRIWTTLHPPTGLFLRVFKTQVLKYLNNLGVI 100
                                      :**** ****.:*************** **********:***:****.*:

OYVMV-[IN_Him_10]_CBX25664_AC3        SINLVIKAVEHVLYNVINKTMYVDQYSDIQFKLY 134
OYVMV-[PK:Fai201:95]_CAA05442_AC3     SINLVIKAVEHVLYNVINQTMYVDQYSDIQFKLY 134
OYVMV-[IN:Abd:08]_ACY78672_AC3        SLNLVIKAVEHVLYNVINQTMYVDQYSDIKLKL- 133
BYVMV-[IN:OY177:07]_ACN51281_AC3      SINLVIKAVEHVLYNVINQTMYVDQYSDIKFNLY 134
CLCuMV-[IN:Lud:04]_AAX84766_AC3       SINLVIKAVEHVLYNLIQQTMYVDQYSDIKFKLY 134
BYVMV-[IN:THD:05]_ACN42644_AC3        SINLVIKAVEHVLYNVINQTMYVDQYSDIKFNLY 134
BYVMV-[IN:PAN:05]_ACN42651_AC3        SINLVIKAVEHVLYNVINQTMYVDQYSDIKFNLY 134
OkLCuIV-[IN:ND:08]_ADG29349_AC3       SINLVIKAVEHVLYNVINQTMYVDQYSDIKFKLY 134
BYVMV-PK[PK:M301:96]_CAA05454_AC3     SLNLVIKAVEHVLYNVINQTMYVDQYSDIKFKLY 134
BYVMV-IN[IN:Mad:01]_AAF63748_AC3      SLNLVIKAVEHVLYNVVNQTMYVDQYSDIKFKLY 134
                                      *:*************::::**********:::*
```

D) Multiple alignment of AC4 ORF sequences. The conserved amino acid residues were represented by '*'.

```
OYVMV-[IN_Him_09]_CBX25667_AC4       --MGNLIFTCSSSSRANTNAKITDSSTSYPQPGQHISIQTFRELNPAPTS 48
OYVMV-[PK:Fai201:95]_CAA05443_AC4    --MGNLIFTCSSSSKGNTSAKITDSSTSYPQPGQHISIRTFRELNPAPTS 48
OYVMV-[IN:Abd:08]_ACY78675_AC4       --MGILNCMFSSSSKESSSAGTIDSSTSPPQLGQHISIRTFRELKAQAMS 48
BYVMV-[IN:OY177:07]_ACN51284_AC4     --MGNLIFTCSSSSRGNSSAKITDSSTSYPQPGQHISIRTYRELNPAPTS 48
CLCuMV-[IN:Lud:04]_AAX84767_AC4      --MGNLIFTCSSSSRGSTNARIKDSSTWYPQPGQHISIRTFRELNQAQMS 48
BYVMV-[IN:THD:05]_ACN42647_AC4       --MGNLIFTCSSSSKENTSARIRDSSTWSPQTGQHISIRTFRELKAQAMS 48
BYVMV-[IN:PAN:05]_ACN42654_AC4       --MGILICMFSSSSRASTSAGTIDSSTSLPQAGQHISIQTFRELNPAPTS 48
OkLCuIV-[IN:ND:08]_ADG29350_AC4      MRMGNLIFTCSSNSRGNTSARISDSSTWSPQAGQHISIRTFRELNQAQMS 50
BYVMV-PK[PK:M301:96]_CAA05455_AC4    --MGNLIFTCSSSSRENTSAKINDSSTWSPQPDQHISIRTFRELNQVPTS 48
BYVMV-IN[IN:Mad:01]_AAF63749_AC4     MKMGNLIFTCSSSSKENTSARIRDSSTWSPQTGQHISIQTFRELNPAPTS 50
                                       ** *    **.*: .:.*   ****  ** .*****:*:***:    *

OYVMV-[IN_Him_09]_CBX25667_AC4       SPTSLRTGTFSTGANFRSTEDQQEEGNKQPTTLTPQHLTREVSQSLLESL 98
OYVMV-[PK:Fai201:95]_CAA05443_AC4    SPTSPRTGTFSTGGNFRSTEDQQEEGNKQPTTLTPQHLTREVSQRLLESS 98
OYVMV-[IN:Abd:08]_ACY78675_AC4       KPTSRKTETSLIMEFSKSMQDQLEEVAKLPTTHMPRQSTQGPKLRPSIY- 97
BYVMV-[IN:OY177:07]_ACN51284_AC4     SPTSTRTETFSNGGNFRSMEDLLEEGNRQPTTLTRQHLTREVSQRLLESL 98
CLCuMV-[IN:Lud:04]_AAX84767_AC4      RPTSTRTGTLSGGESFRSMEDQQEEDDRQPTTLTPQHLTQAISQRLLESL 98
BYVMV-[IN:THD:05]_ACN42647_AC4       SPTSTRTETPLNGEDSRSMEDLLEEDNNQPMTLTRQHLTQEVRQRLLELL 98
BYVMV-[IN:PAN:05]_ACN42654_AC4       SPISTRTEIPLNGGNSRSTADLQGEDNNQPMTLTRQHLTQEVRQRLLEFL 98
OkLCuIV-[IN:ND:08]_ADG29350_AC4      SPTSTRTETPSNGESFRSMDDLQEGDSRQPTTLTPQHLTQAVSQRLLE-  98
BYVMV-PK[PK:M301:96]_CAA05455_AC4    SLIWRRTETPSTGEGFRSMDDLHEEDNNQPMTLTPQRLTQVVSQRLLE-- 96
BYVMV-IN[IN:Mad:01]_AAF63749_AC4     SPISTRTEIPLNGGNSRSTADLQGEDNNQPMTLTRQHLTQEVRQRLLEFL 100
                                          :*         :*  *      . * *   :: *:

OYVMV-[IN_Him_09]_CBX25667_AC4       RN 100
OYVMV-[PK:Fai201:95]_CAA05443_AC4    RS 100
OYVMV-[IN:Abd:08]_ACY78675_AC4
BYVMV-[IN:OY177:07]_ACN51284_AC4     RN 100
CLCuMV-[IN:Lud:04]_AAX84767_AC4      RN 100
BYVMV-[IN:THD:05]_ACN42647_AC4       RR 100
BYVMV-[IN:PAN:05]_ACN42654_AC4       KN 100
OkLCuIV-[IN:ND:08]_ADG29350_AC4      --
BYVMV-PK[PK:M301:96]_CAA05455_AC4    --
BYVMV-IN[IN:Mad:01]_AAF63749_AC4     KN 102
```

E) Multiple alignment of AV1 ORF sequences. The conserved amino acid residues were represented by '*'

F)

```
OYVMV-[IN_Him_09]_CBX25663_AV1       MSKRAADIVISTPASKVRRRLNFGSPYTSRAAAPIVRVTKQQAWTNRPMN 50
OYVMV-[PK:Fai201:95]_CAA05438_AV1    MSKRAADIVISTPASKVRRRLNFGSPYTSRAAAPIVRVTKQQAWTNRPMN 50
OYVMV-[IN:Abd:08]_ACY78671_AV1       MSKRAADIVISTPASKVRRRLNFGSPYTSRAAAPIVRVTKQQAWTNRPMN 50
BYVMV-[IN:OY177:07]_ACN51279_AV1     MSKRAADIVIFTPASKVRRRLNFGSPYTSRAAAPIFRVTKQQAWTNRPMN 50
CLCuMV-[IN:Lud:04]_AAX84763_AV1      MSKRAADIVISTPASKVRRRLNFGSPYTSRAAAPIVRVTKQQAWTNRPMN 50
BYVMV-[IN:THD:05]_ACN42642_AV1       MSKRAADIVISTPASKVRRRLNFGSPYTSRAAAPIVRVTKQQAWTNRPMN 50
BYVMV-[IN:PAN:05]_ACN42649_AV1       MSKRAADIVISTPASKVRRRLNFGSPYTSRAAAPIVRVTKQQAWTNRPMN 50
OkLCuIV-[IN:ND:08]_ADG29345_AV1      MSKRAADIVISTPASKVRRRLNFGSPYSSRVAAPIVRVTKQQAWTNRPMN 50
BYVMV-PK[PK:M301:96]_CAA05450_AV1    MSKRAADIVISTPASKVRRRLNFGSPYTSRAAAPIVRVTKQQAWTNRPMN 50
BYVMV-IN[IN:Mad:01]_AAF63751_AV1     MSKRAADIVISTPASKVRRRLNFGSPYTSRAAAPIVRVTKQQAWTNRPMN 50
                                     ********** ****************:**.****.**************

OYVMV-[IN_Him_09]_CBX25663_AV1       RKPRMYRMYRSPDVPRGCEGPCKVQSFESRHDVVHIGKVMCISDVTRGVG 100
OYVMV-[PK:Fai201:95]_CAA05438_AV1    RKPRMYRMYRSPDVPRGCEGPCKVQSFESRHDVVHIGKVMCISDVTRGVG 100
OYVMV-[IN:Abd:08]_ACY78671_AV1       RKPRMYRMYRSPDVPRGCEGPCKVQSFESRHDVVHIGKVMCISDVTRGVG 100
BYVMV-[IN:OY177:07]_ACN51279_AV1     RKPRMYRMYRSPDVPRGCEGPCKVQSFESRHDVVHIGKVMCISDVTRGVG 100
CLCuMV-[IN:Lud:04]_AAX84763_AV1      RKPRMYRMYRSPDVPRGCEGPCKVQSFESRHDVVHIGKVMCISDVTRGVG 100
BYVMV-[IN:THD:05]_ACN42642_AV1       RKPRMYRMYRSPDVPRGCEGPCKVQSFESRHDVVHIGKVMCISDVTRGVG 100
BYVMV-[IN:PAN:05]_ACN42649_AV1       RKPRMYRMYRSPDVPRGCEGPCKVQSFESRHDFVSFGKVMCISDVTRGVG 100
OkLCuIV-[IN:ND:08]_ADG29345_AV1      RKPRMYRMYRSPDVPRGCEGPCKVQSFESRHDVVHIGKVMCISDVTRGVG 100
BYVMV-PK[PK:M301:96]_CAA05450_AV1    RKPRMYRMYRSPDVPRGCEGPCKVQSFESRHDVVHIGKVMCISDVTRGVG 100
BYVMV-IN[IN:Mad:01]_AAF63751_AV1     RKPRMYRMYRSPDVPRGCEGPCKVQSFESRHDVVHIGKVMCISDVTRGVG 100
                                     ********************************.* :**************

OYVMV-[IN_Him_09]_CBX25663_AV1       LTHRIGKRFCVKSVYVLGKIWMDENIKTKNHTNSVMFFLVRDRRPVDKPQ 150
OYVMV-[PK:Fai201:95]_CAA05438_AV1    LTHRIGKRFCVKSVYVLGKIWMDENIKTKNHTNSVMFFLVRDRRPVDKPQ 150
OYVMV-[IN:Abd:08]_ACY78671_AV1       LTHRIGKRFCVKSVYVLGKIWMDENIKSKNHTNSVMFFLVRDRRPTDKPR 150
BYVMV-[IN:OY177:07]_ACN51279_AV1     LTHRIGKRFCVKSVYVLGKIWMDENIKTKNHTNSVMFFLVRDRRPTDKPQ 150
CLCuMV-[IN:Lud:04]_AAX84763_AV1      LTHRIGKRFCVKSVYVLGKIWMDENIKTKNHTNSVMFFLVRDRRPTDKPQ 150
BYVMV-[IN:THD:05]_ACN42642_AV1       LTHRIGKRFCVKSVYVLGKIWMDENNKTKNHTNSVMFFLVRDRRPTDKPQ 150
BYVMV-[IN:PAN:05]_ACN42649_AV1       LTHRIGKRFCVKSVYVLGKIWMDENIKTKNHTNSVMFFLVRDRRPTDKPQ 150
OkLCuIV-[IN:ND:08]_ADG29345_AV1      LTHRIGKRFCVKSVYVLGKIWMDENIKTKNHTNSVMFFLVRDRRPVDKPQ 150
```

```
BYVMV-PK[PK:M301:96]_CAA05450_AV1     LTHRIGKRFCVKSVYVLGKIWMDENIKTKNRTNSVMFFLVRDRRPTDKPQ 150
BYVMV-IN[IN:Mad:01]_AAF63751_AV1      LTHRIGKRFCVKSVYVLGKIWMDENIKTKNHTNSVMFFLVRDRRPVDKPQ 150
                                      ************************* *:**:***************.**:

OYVMV-[IN_Him_09]_CBX25663_AV1        DFGEVFNMFDNEPSTATVKNMHRDRYQVLRKWHATVTGGQYASKEHALVK 200
OYVMV-[PK:Fai201:95]_CAA05438_AV1     DFGEVFNMFDNEPSTATVKNMHRDRYQVLRKWHATVTGGQYASKEQALVK 200
OYVMV-[IN:Abd:08]_ACY78671_AV1        DFGEVFNMFDNEPSTATVKNMHRDRYQVLRKWHATVTGGQYAAREQALVK 200
BYVMV-[IN:OY177:07]_ACN51279_AV1      DFGEVFNMFDNEPSTATVKNMHRDRYQVLSKWHATVTGGQYASKEQALVK 200
CLCuMV-[IN:Lud:04]_AAX84763_AV1       DFGEVFNMFDNEPSTATVKNMHRDRYQVLRKWHATVTGGQYASKEQALVK 200
BYVMV-[IN:THD:05]_ACN42642_AV1        DFGEVFNMFDNEPSTATVKNMHRDRYQVLRKWHATVTGGQYASKEQALVK 200
BYVMV-[IN:PAN:05]_ACN42649_AV1        DFGEVFNMFDNEPSTATVKNMHRDRYQVLRKWHATVTGGQYASKEQALVK 200
OkLCuIV-[IN:ND:08]_ADG29345_AV1       DFGEVFYMFDNEPSTATVKNSHRDRYQVLRKWHATVTGGQYASKEQALVK 200
BYVMV-PK[PK:M301:96]_CAA05450_AV1     DFGEVFNMFDNEPSTATVKNMHRDRYQVLRKWHATVTGGQYASKEQALVK 200
BYVMV-IN[IN:Mad:01]_AAF63751_AV1      DFGEVFNMFDNEPSTATVKNMHRDRYQVLRKWHATVTGGQYAAREQALVK 200
                                      ****** ************* ******** ************::*:****

OYVMV-[IN_Him_09]_CBX25663_AV1        KFVRVNNYVVYNQQEAGKYENHTENALMLYMACTHASNPVYATLKIRIYF 250
OYVMV-[PK:Fai201:95]_CAA05438_AV1     KFVRVNNYVVYNQQEAGKYENHTENALMLYMACTHASNPVYATLKIRIYF 250
OYVMV-[IN:Abd:08]_ACY78671_AV1        KFVRVNNYVVYNQQEAGKYENHTENALMLYMACTHASNPVYATLKIRIYF 250
BYVMV-[IN:OY177:07]_ACN51279_AV1      KFVRVNNYVVYNQQEAGKYENHTENALMLYMACTHASNPVYATLKIRIYF 250
CLCuMV-[IN:Lud:04]_AAX84763_AV1       KFIRVNNYVVYNQQEAGKYENHTENALMLYMACTHASNPVYATLKIRIYF 250
BYVMV-[IN:THD:05]_ACN42642_AV1        KFVRVNNYVVYNQQEAGKYENHTENALMLYMACTHASNPVYATLKIRIYF 250
BYVMV-[IN:PAN:05]_ACN42649_AV1        KFVRVNNYVVYNQQEAGKYENHTENALMLYMACTHASNPVYATLKIRIYF 250
OkLCuIV-[IN:ND:08]_ADG29345_AV1       KFVKVNNYVVYNQQEAGKYENHTENALMLYMACTHASNPVYATLKIRIYF 250
BYVMV-PK[PK:M301:96]_CAA05450_AV1     KFIRVNNYVVYNQQEAGKNENHTENALMLYMACTHASNPVYATLKIRIYF 250
BYVMV-IN[IN:Mad:01]_AAF63751_AV1      KFVRVNNYVVYNQQEAGKYENHTENALMLYMACTHASNPVYATLKIRIYF 250
                                      **::************** *******************************

OYVMV-[IN_Him_09]_CBX25663_AV1        YDSVTN---   256
OYVMV-[PK:Fai201:95]_CAA05438_AV1     YDSVTN---   256
OYVMV-[IN:Abd:08]_ACY78671_AV1        YDSVTNHMI   259
BYVMV-[IN:OY177:07]_ACN51279_AV1      YDSVTN---   256
CLCuMV-[IN:Lud:04]_AAX84763_AV1       YDSVTN---   256
BYVMV-[IN:THD:05]_ACN42642_AV1        YDSVTN---   256
BYVMV-[IN:PAN:05]_ACN42649_AV1        YDSVTN---   256
OkLCuIV-[IN:ND:08]_ADG29345_AV1       YDSVTN---   256
BYVMV-PK[PK:M301:96]_CAA05450_AV1     YDSVTN---   256
BYVMV-IN[IN:Mad:01]_AAF63751_AV1      YDSVTN---   256
                                      ******
```

G) Multiple alignment of AV2 ORF sequences. The conserved amino acid residues were represented by '*'.

```
OYVMV-[IN_Him_09]_CBX25662_AV2        MWDPLLNEFPDTVHGFRCMLSVKYLQLLSQDYSPDTLGYELIRDLICILR 50
OYVMV-[PK:Fai201:95]_CAA05439_AV2     MWDPLVNEFPDTVHGFRCMLSVKYLQLLSQDYSPDTLGYELIRDLICILR 50
BYVMV-[IN:OY177:07]_ACN51278_AV2      MWDPLLNEFPDTVHGFRCMLSVKYLQFLSQDYSPDTLGYELIRDLICILR 50
CLCuMV-[IN:Lud:04]_AAX84762_AV2       MWDPLLNEFPDTVHGFRCMLSVKYLQLLSQDYSPDTLGYELIRDLICILR 50
BYVMV-[IN:THD:05]_ACN42641_AV2        MWDPLLNEFPDTVHGFRCMLSVKYLQLLSQDYSPDTLGYELIRDLICILR 50
BYVMV-[IN:PAN:05]_ACN42648_AV2        MWDPLVNEFPDTVHRFRCMLSVKYLQLLSQDYSPDTLGYELIRDLICILR 50
OkLCuIV-[IN:ND:08]_ADG29346_AV2       MWDPLLNEFPDTVHGFRCMLSVKYLQLLSQDYSPDTLGYDLIRDLICILR 50
BYVMV-PK[PK:M301:96]_CAA05451_AV2     MWDPLVNEFPDTVHRFRCMLSVKYLQLLSQDYSPDTLGYELIRDLICILR 50
BYVMV-IN[IN:Mad:01]_AAF63752_AV2      MWDPLLNEFPDTVHGFRCMLSLKYLQLLSQDYSPDTLGYELIRDLICILR 50
CLCuAV-[PK:K802a:96]_CAA05457_AV2     MWDPLVNEFPDTVHGFRCMLSVKYLQLLSQDYSPDTLGYELIRDLICILR 50
                                      *****:******** ******:****:************:**********

OYVMV-[IN_Him_09]_CBX25662_AV2        SRNYVEASCRYRHFYARVESTPASELRQPIHQPCCCPHCPRHKTTGMDKQ 100
OYVMV-[PK:Fai201:95]_CAA05439_AV2     SRNYVEASCRYRHFYARVESTPASELRQPIHQPCCCPHCPRHKTTGMDKQ 100
BYVMV-[IN:OY177:07]_ACN51278_AV2      SRNYVEASCRYRHFYARVESTPAFELRQPIHQPCCCSHFPRHKTTGMDKQ 100
CLCuMV-[IN:Lud:04]_AAX84762_AV2       SRNYVEASCRYRHFYARVESTPASELRQPIHQPCCCPHCPRHKTTGMDKQ 100
BYVMV-[IN:THD:05]_ACN42641_AV2        SRNYVEASCRYRHFYARVESTPASELRQPIHQPCCCPHCPRHKTTGMDKQ 100
BYVMV-[IN:PAN:05]_ACN42648_AV2        SRNYVEASCRYRHFYARVESTPASELRQPLHQPCCCPHCPRHKTTGMDKQ 100
OkLCuIV-[IN:ND:08]_ADG29346_AV2       SRNYVEASCRYRHFYARVESTSASELRQPIQQPCCCAHCPRHKTTGMDKQ 100
BYVMV-PK[PK:M301:96]_CAA05451_AV2     SRNYVEASCRYRHFYARVESTPASELRQPLHQPCCCPHCPRHKTTGMDKQ 100
BYVMV-IN[IN:Mad:01]_AAF63752_AV2      SRNYVEASCRYRHFYARVESTPASELRQPIHQPCCCPHCPRHKTTGMDKQ 100
CLCuAV-[PK:K802a:96]_CAA05457_AV2     SRNYVEASCRYRHFYARVESTPASELRQPIHQPCCCPHCPRHKTTGMDKQ 100
                                      *********************.* *****::*****.* ***********

OYVMV-[IN_Him_09]_CBX25662_AV2        AYEQETQNVPDVQKSGCSTGM 121
OYVMV-[PK:Fai201:95]_CAA05439_AV2     AYEQETQNVPDVQKSGCSKGM 121
BYVMV-[IN:OY177:07]_ACN51278_AV2      AYEQETQNVPDVQKSGCSKGM 121
CLCuMV-[IN:Lud:04]_AAX84762_AV2       AYEQEAQNVSDVQKSGCSKGM 121
BYVMV-[IN:THD:05]_ACN42641_AV2        AYEQETQNVPDVQKSGCSKGM 121
BYVMV-[IN:PAN:05]_ACN42648_AV2        AYEQETQNVPDVQKSGCPKGM 121
OkLCuIV-[IN:ND:08]_ADG29346_AV2       AYEQEAQNVSDVQKSGCSKGL 121
BYVMV-PK[PK:M301:96]_CAA05451_AV2     AYEQETQNVPDVQKSGCSKGM 121
BYVMV-IN[IN:Mad:01]_AAF63752_AV2      AYEQETQNVPDVQKSGCSKGM 121
CLCuAV-[PK:K802a:96]_CAA05457_AV2     AYEQETQNVPDVQKSGCSKGM 121
                                      *****:***.*******..*:
```

H) Multiple alignment of BetaC1 ORF sequences of betasatellites. The conserved amino acid residues were represented by '*'.

```
OYVB-[IN:Him:09]_CBZ47002_BetaC1             MISTHYMKISIHFIYIFISNIYIRRTIKMTRFSKNREGIVFKVDVRLMQQ 50
OLCuB-[PK:Goj:97]_CAC87055_BetaC1            MISTHYLKISIHFIYIFISNIHIRRTIKMTRFSRNREGIVFKVDVRLMQQ 50
BYVB-[IN_Gun_OY112_06]_ADQ55866_BetaC1       ------MKISIHFIYIFTSNIYIRRTIKMTRFSKNREGIVFKVDVKLMQQ 44
OYVB-[IN:Abd:08]_ADA60716_BetaC1             MISTHYMKISIHFIYKFISSIYIRRTIKMTRFSKNREGIVFKVDVKLMQQ 50
OLCuB-[IN:Chn:OY177:06]_ADQ55874_BetaC1      ----------------------------MTRISKNREGIVFKVDVRLMQQ 22
BYVB-[IN:Abd:OY164:06]_ADQ55881_BetaC1       ------MKISNHFIHIFISSIYIRRTIKMTRFSKNREGIVFKVDVRLMQQ 44
BYVB-[IN:Mut:01]_CAC28127_BetaC1             ------MKISIHFIYIFISNIYIKRTIKMTRFSKNREGIVFKVDVRLMQQ 44
BYVB-[IN:Bar:06]_ABN80228_BetaC1             ----------------------------MTRFSKNREGIVFKVDVKLMQQ 22
OLCuIB-[IN:ND:08]_ADG29352_BetaC1            -----------------------------------------------MQQ 3
BYVB-[IN:Rai:OY56B:05]_ADQ55865_BetaC1       ----------------------------MTRFSKTREGIVFKVDVRLMQQ 22
                                                                                              **

OYVB-[IN:Him:09]_CBZ47002_BetaC1             QRISVLMQISSTRSPAISTKTFIIDYTYQQLHIPFDFNGLEGTITSTFKF 100
OLCuB-[PK:Goj:97]_CAC87055_BetaC1            QRISVHMQISSTKSPAISTRKFIIDYTYQQLHIPFDFNGLEGTITSTFKF 100
BYVB-[IN_Gun_OY112_06]_ADQ5586_BetaC1        QRISVQMQISSTRSPAISTKTFIIDYTYQQLHIPFDFNGLEGTITSTSKF 94
OYVB-[IN:Abd:08]_ADA60716_BetaC1             QRISVQMQISSTRSPAISTKTFIIDYTYQQLHIPFDSNGLEGTITSTFKF 100
OLCuB-[IN:Chn:OY177:06]_ADQ55874_BetaC1      QRISVHMQISSTRSPAVSTRTFIIDYTYQQLHIPFDFNGLEGTITSTFKF 72
BYVB-[IN:Abd:OY164:06]_ADQ55881_BetaC1       QRISVQMQISSTRSPAISTKTFIIDYTYQQLHIPFDFNGLEGTITSTFKF 94
BYVB-[IN:Mut:01]_CAC28127_BetaC1             QRISVHMQISSTRSPAISTRTFIIDYTYQQLHIPFDFNGLEGTITSTFKF 94
BYVB-[IN:Bar:06]_ABN80228_BetaC1             QRISVQMQISSTRSPAISTKTFIIDYTYQQLHIPFDFNGLEGTITSTFKF 72
OLCuIB-[IN:ND:08]_ADG29352_BetaC1            QRISVQMQISSTRSPAISTKTFIIDYTYQQLHIPFDFNGLEGTITSTFKF 53
BYVB-[IN:Rai:OY56B:05]_ADQ55865_BetaC1       QRISVHMQISSTKSPAISTKTFIIDYTYQQLHIPFDFNGLEGTITSTFKF 72
                                             ***** ******:***:**: **************.*********** **

OYVB-[IN:Him:09]_CBZ47002_BetaC1             HYWGSKAEEILEEDIIHMVDIIIIENPEIMGMDVNEPVTIDNKIII 146
OLCuB-[PK:Goj:97]_CAC87055_BetaC1            HYWGSKAEEILDEDIIHMVDIIIIENPDIMGMDVNEPVTIDNKIII 146
BYVB-[IN_Gun_OY112_06]_ADQ5586_BetaC1        HYWGSKAEEILEEDIIHMVDIIIIENPEIMGMDVNEPVTIDNKIII 140
OYVB-[IN:Abd:08]_ADA60716_BetaC1             HYWGSKAEEILEEDIIHMVDIIIIENPEIMGMDVNEPVTIDNKIII 146
OLCuB-[IN:Chn:OY177:06]_ADQ55874_BetaC1      HYWGSKAEEILDEDIIHMVDLIIIENPDIMGMDVNEPVTLDNKIII 118
BYVB-[IN:Abd:OY164:06]_ADQ55881_BetaC1       HYWGSKAEEILEEDIIHMVDIIIIENPEIMGMDVNEPVTIDNKIII 140
BYVB-[IN:Mut:01]_CAC28127_BetaC1             HYWGSKAEEILEEDIIHMVDIIIIENPDIMGMDVNEPVTIDNKIII 140
BYVB-[IN:Bar:06]_ABN80228_BetaC1             HYWGSKAEEILEEDIIHMVDIIIIENPEIMGMDVNEPVTIDNKIII 118
OLCuIB-[IN:ND:08]_ADG29352_BetaC1            HYWGSKAEEILEEDIIHMVDIIIIENPEIMGMDVNEPVTIDNKIII 99
BYVB-[IN:Rai:OY56B:05]_ADQ55865_BetaC1       HYWGSKAEEILEEDIIHMVDIIIIENPDIMGMDVNEPVTIDNKIII 118
                                             ***********:********:******:************:******
```

2. Multiple Alignments of non coding regions of present begomovirus with the other closely related begomoviruses from Indian sub-continent

A) Multiple alignment of intergenic region (IR) sequences. The conserved amino acid residues were represented by '*'.

```
OYVMV-[IN:Him:09]_FR694925_IR       --TGACTTGGTCAATCGGTACC---CAGATCTAGTCTTATGTCAATTGGT 45
OYVMV-[PK:Fai201:95]_AJ002451_IR    --TGACTTGGTCAATCGGTACC---CAGATCTAATCCTATGTCAATTGGT 45
OYVMV-[IN:Abd:08]_GU181356_IR       GTTGACTAAGTC------AATC----CGGTGTGCTCTT-TAACTCTCTGC 39
BYVMV-[IN:OY177:07]_FJ179371_IR     --TGACTTGGTCAATCGGTACC---CAGATCTAATCCTATGTCAATTGGG 45
CLCuMV-[IN:Lud:04]_AY765257_IR      TTTGGC--AATCGGT--GTAC------ACTCTAATTCTATGGCAATCGGT 40
BYVMV-[IN:THD:05]_FJ176235_IR       GTTGACTATGTTTG--AGACCC-GATTGAC-CGCTCTTACAACTCTCCCC 46
BYVMV[IN:PAN:05]_FJ176236_IR        GTTGACTAAAATTG--AGTACC-GATTGACTCGCTCTGGCAACTCTCTCT 47
OkLCuIV-[IN_ND_08]_GQ245760_IR      GTTGACTAAATTTG--AGTACC-GATTGACTCGCTCTGGCAACTCTCCCC 47
BYVMV-PK[PK:M301:96]_AJ002453_IR    GTTGACCAAGTGAG--AGGACCCGATTGAC-CGCTCTTGCAACTCTCCCC 47
BYVMV-IN[IN:Mad:01]_AF241479_IR     GTTGACTATTTTTG--AGACCC-GATTGAC-CGCTCTTACAACTCTCCCC 46
                                      ** *                            *       *  *

OYVMV-[IN:Him:09]_FR694925_IR       G-----------AACGGTACCCTATATATAG-TGGGTACTGAATGGCAAT 83
OYVMV-[PK:Fai201:95]_AJ002451_IR    G-----------AACGGTGCCCTATATATAG-TGGGTACTGAATGGCATT 83
OYVMV-[IN:Abd:08]_GU181356_IR       ATGTATCGGTGTTTTGGAGTCCTATATATAT-GGAGACTCTAATGGCATT 88
BYVMV-[IN:OY177:07]_FJ179371_IR     G-----------AACGGTACCCTATATATAG-TGGGTACCAAATGGCAAT 83
CLCuMV-[IN:Lud:04]_AY765257_IR      G----------TAACGGGGTGCGATATATAGGTGTACCCCAAATGGCATT 80
BYVMV-[IN:THD:05]_FJ176235_IR       A------G-TATATCGGGTCCCTATATATAG-TGAGACCCAAATGGCAT- 87
BYVMV[IN:PAN:05]_FJ176236_IR        G------G-TATATCGGTACCCAATATATAG-TGAGTACCAAATGGCAT- 88
OkLCuIV-[IN_ND_08]_GQ245760_IR      T------GGTATATCGGTACCCAATATATAG-TGAGTACCAAATGGCAT- 89
BYVMV-PK[PK:M301:96]_AJ002453_IR    T------G-TATATTGGGTCTCAATATATAG-TGAGACCCAAATGGCAT- 88
BYVMV-IN[IN:Mad:01]_AF241479_IR     A------G-TATATCGGGTCCCTATATATAG-TGAGACCCAAATGGCAT- 87
                                                   **    * *******   *       *******

OYVMV-[IN:Him:09]_FR694925_IR       -ATTTGTAAT---TATGAAAA--TAAATTCAAAATCCTCACGCTCCAA-- 125
OYVMV-[PK:Fai201:95]_AJ002451_IR    -ATTTGTAAT---TACAAAAG--GAAATTCAAAATCTACACGCTCCAA-- 125
OYVMV-[IN:Abd:08]_GU181356_IR       AAAATGTAAAATAAAAGAAAACACCTATTTGAATTCC---------AAGG 129
BYVMV-[IN:OY177:07]_FJ179371_IR     -ATTTGTAAT---TATGAAAG--GAAATTCAAAATCCTCACGCTCCAA-- 125
CLCuMV-[IN:Lud:04]_AY765257_IR      --ATCGTAAT---TTGAGAAA--TCATTTCAAAATCCTCACGCTTCAA-- 121
BYVMV-[IN:THD:05]_FJ176235_IR       -AATTGTAA----TAAAACAACTTTAATTTGAAATTCAAACGA--AAAGG 130
BYVMV[IN:PAN:05]_FJ176236_IR        -AAGTGTAAA---TAAGAGAACTTTAACTTGAACCTCACACGA--CAAGG 132
OkLCuIV-[IN_ND_08]_GQ245760_IR      -ATGTGTAAT---TTTG-TAA-TGAAATTCAAAATCCTCACGCTCCAA-- 131
BYVMV-PK[PK:M301:96]_AJ002453_IR    -TATCGTAAT---TTGGACAA-T-AGATTCAAAATCCTCACGCTCCAA-- 130
BYVMV-IN[IN:Mad:01]_AF241479_IR     -AATTGTAA----TAAAACAACTTTAATTTGAAATTCAAACGA--AAAGG 130
                                         ****          *        *  **             **

OYVMV-[IN:Him:09]_FR694925_IR       --AAAGCGGCCATCCGTATAATATTACCGGATGGCCGCGCGATTTTTTTT 173
OYVMV-[PK:Fai201:95]_AJ002451_IR    --AAAGCGGCCATCCGTATAATATTACCGGATGGCCGCGCGATTTTTTTT 173
OYVMV-[IN:Abd:08]_GU181356_IR       --AAAGCGGCCATCCGTCTAATATCACGGGATGGCCGCGCGATTTTTT-A 176
BYVMV-[IN:OY177:07]_FJ179371_IR     --AAAGCGGCCATCCGTATAATATTACCGGATGGCCGCGGTATTTTTTTA 173
CLCuMV-[IN:Lud:04]_AY765257_IR      --AAAGCGGCCATCCGTTTAATATTACCGGATGGCCGCGCGATTTTTT-- 167
BYVMV-[IN:THD:05]_FJ176235_IR       CTAAAGCGGCCATCCGTATAATATTACCGGATGGCCGCGCGATTTTTTTA 180
BYVMV[IN:PAN:05]_FJ176236_IR        CTAGCGCGGCCATCCATCTAACATTACCGGATGGCCGCGCGATTTTTTTA 182
OkLCuIV-[IN_ND_08]_GQ245760_IR      --AAAGCGGCCATCCGTATAATATTACCGGATGGCCGCGCGATTTTTTAA 179
BYVMV-PK[PK:M301:96]_AJ002453_IR    --AAAGCGGCCATCCGACTAATATTACCGGATGGCCGCGCGATTTTTTTA 178
BYVMV-IN[IN:Mad:01]_AF241479_IR     CTAAAGCGGCCATCCGTATAATATTACCGGATGGCCGCGCGATTTTTT-A 179
                                      *  **********   *** ** ** ***********  *******

OYVMV-[IN:Him:09]_FR694925_IR       AGTGGTGGGTCCAGAACGCACGACGATGCAGACTCAAAGCTTAGATAACG 223
OYVMV-[PK:Fai201:95]_AJ002451_IR    AGTGGTGGGTCCAGAACGCACGACGATGCAGACTCAAAGCTTAGATAACG 223
OYVMV-[IN:Abd:08]_GU181356_IR       AGTGGTGGTTCCAGAACGCACGACTATGCAGACTCAAAGCTTAGATAACG 226
BYVMV-[IN:OY177:07]_FJ179371_IR     AGTGGGGGGTCCAGAACGCACCACGATGCAGACTCAAAGCTTAGATAACG 223
CLCuMV-[IN:Lud:04]_AY765257_IR      --TTGTGGGCCCCCGATTTATGAGATTGCTCCCTCAAAGCT-AAATAACG 214
BYVMV-[IN:THD:05]_FJ176235_IR       AGTGGTGGGTCCAGAACGCACGACGATGCAGACTCAAAGCTTAGATAACG 230
BYVMV[IN:PAN:05]_FJ176236_IR        AGTGGTGGGTCCAGAACGCACGACGATGCAGACTCAAAGCTTAGATAACG 232
OkLCuIV-[IN_ND_08]_GQ245760_IR      AGTGGTGGGTCCAGAACGCACGACGATGCAGACTCAAAGCTTAGATAACG 229
BYVMV-PK[PK:M301:96]_AJ002453_IR    AGTGGTGGGTCCAGAACGCACGACGATGCAGACTCAAAGCTTAGATAACG 228
BYVMV-IN[IN:Mad:01]_AF241479_IR     AGTGGTGGGTCCAGAACGCACGACGATGCAGACTCAAAGCTTAGATAACG 229
                                      * * **  **   *   *  *   ***   ********* * ******
```

```
OYVMV-[IN:Him:09]_FR694925_IR       CTCCTTTG-GCTATAAGTA-CTTGCGCACTAAGTTTAAATTCAAAAC 268
OYVMV-[PK:Fai201:95]_AJ002451_IR    CTTCTTTG-GCTATAAGTA-CTTGCGCACTAAGTTTAAATTCAAAAC 268
OYVMV-[IN:Abd:08]_GU181356_IR       CTCCTTCG-GCTATAAGTAACGTGCGCACTAAGTTTCAAATTCAAAA 272
BYVMV-[IN:OY177:07]_FJ179371_IR     CTCCTTCG-GATAAAAGTA-CTTGCGCATCAAGTTTCAATTTAAAAA 268
CLCuMV-[IN:Lud:04]_AY765257_IR      CTCCCGCACACTATAAGTA-CTTGCGCACTAAGTTTCAAATTCAAAC 260
BYVMV-[IN:THD:05]_FJ176235_IR       CTCCTTCG-GCTATAAGTA-CGTGCGCACTAAGTTTCAATTCAAAAA 275
BYVMV[IN:PAN:05]_FJ176236_IR        CTCCTTCG-GCTATAAGTA-CTTGCGCACTAAGTTTCAAATTGAAAC 277
OkLCuIV-[IN_ND_08]_GQ245760_IR      CTCCTTCG-GCTATAAGTA-CTTGCGCACTAAGTTTCAATTCAAAAA 274
BYVMV-PK[PK:M301:96]_AJ002453_IR    CTCCTTCG-GCTATAAGTA-CTTGCGCACTAAGTTTCAAATTGAAAC 273
BYVMV-IN[IN:Mad:01]_AF241479_IR     CTCCTTCG-GCTATAAGTA-CGTGCGCACTAAGTTTCAATTCAAAAA 274
                                    ** *       ** ***** * ******  ****** ** *  ***
```

105188

Brief *Resume* of student

Name	:	More Prashant Digambar
Father's Name	:	Sh. More Digambar Tukaram
Mother's Name	:	Smt. More Asha Digambar
Date of Birth	:	6th April 1987
Permanent Address	:	Village: Jadhavwadi, PO: Modnimb, Teh: Madha, Distt: Solapur (Maharashtra), India (413301)

Academic Qualification:

Qualification	Year	School/ Board/University	Marks (%)	Division	Major Subject
High school (Matriculation)	2002	MSBSHSE, Pune, (Maharashtra)	76.80%	Ist	Marathi, Hindi, English, Science, Mathematics
10+2	2004	MSBSHSE, Pune, (Maharashtra)	79.00%	Ist	Biology, Chemistry, Physics, Mathematics, English
B.Sc. (Agri. Biotech.)	2008	Marathwada Agriculture University, Parbhani (M.S.)	81.5%	Ist	Plant Biotech, Animal Biotech, Biochemistry, Food technology, Microbiology and Molecular biology,

Fellowships/Scholarships/Gold Medals/Awards/Any Other Distinction:

2008-2010: Recipient of merit Fellowship from Department of Biotechnology, Govt. of India.