DEVELOPMENT OF RECOMBINANT COAT PROTEIN FOR IMMUNODIAGNOSIS OF BANANA BUNCHY TOP AND BRACT MOSAIC DISEASES

DARSANA DILIP K.C.

2016-21-024

DEPARTMENT OF PLANT PATHOLOGY

COLLEGE OF AGRICULTURE

VELLANIKKARA, THRISSUR - 680 656

KERALA, INDIA

2021

DEVELOPMENT OF RECOMBINANT COAT PROTEIN FOR IMMUNODIAGNOSIS OF BANANA BUNCHY TOP AND BRACT MOSAIC DISEASES

by

DARSANA DILIP K.C.
2016-21-024

THESIS

Submitted in partial fulfillment of the

requirement for the degree of

DOCTOR OF PHILOSOPHY IN AGRICULTURE

Faculty of Agriculture

Kerala Agricultural University

DEPARTMENT OF PLANT PATHOLOGY

COLLEGE OF AGRICULTURE

VELLANIKKARA, THRISSUR - 680 656

KERALA, INDIA

2021

DECLARATION

I, **Darsana Dilip K. C.** (2016-21-024) hereby declare that this thesis entitled **"Development of Recombinant Coat Protein for Immunodiagnosis of Banana Bunchy Top and Bract Mosaic Diseases"** is a bonafide record of research work done by me during the course of research and the thesis has not been previously formed for the award of any degree, diploma, associateship, fellowship or other similar titles, of any other University or Society.

Vellanikkara,

Date: 20.5.2021

Darsana Dilip K.C.

(2016 -21- 024)

CERTIFICATE

Certified that this thesis entitled **"Development of Recombinant Coat Protein for Immunodiagnosis of Banana Bunchy Top and Bract Mosaic Diseases"** is a record of research work done independently by **Ms. Darsana Dilip K.C. (2016-21-024)** under my guidance and supervision and that it has not previously formed the basis for the award of any degree, diploma, fellowship or associateship to her.

20/5/2021

Kannara
Date: 20/05/2021

Dr. Vimi Louis
Professor (Plant Pathology)
Banana Research Station
Kannara, Thrissur - 680 656

CERTIFICATE

We, undersigned members of the advisory committee of **Ms. Darsana Dilip K. C**. (2016-21-024) a candidate for the degree of **Doctor of Philosophy in Agriculture** with major in Plant Pathology, agree that this thesis entitled **"DEVELOPMENT OF RECOMBINANT COAT PROTEIN FOR IMMUNODIAGNOSIS OF BANANA BUNCHY TOP AND BRACT MOSAIC DISEASES"** may be submitted by Ms. Darsana Dilip K.C. (2016-21-024), in partial fulfilment of the requirement for the degree.

20/5/2021

Dr. Vimi Louis
(Chairman, Advisory Committee)
Professor (Plant Pathology)
Banana Research Station
Kannara, Thrissur

20.5.2021

Dr. Anita Cherian K.
(Member, Advisory Committee)
Professor & Head
Dept. Plant Pathology
CoA, Vellanikkara

20/5/2021

Dr. Abida P.S.
(Member, Advisory Committee)
Professor & Head
CPBMB, COA, Vellanikkara

20/05/2021.

Dr. Smita Nair
(Member, Advisory Committee)
Assistant Professor
CPBMB, COA, Vellanikkara

20/05/2021

Dr. M. Mini
(Member, Advisory Committee)
Professor & Head
Dept. Veterinary Microbiology
CoVAS, Mannuthy

20.5.2021

Dr.A. ISHWARA BHAT. Ph.D
Principal Scientist (Plant Pathology)
Indian Institute of Spices Research
Kozhikode-673012, Kerala

EXTERNAL EXAMINER

ACKNOWLEDGMENTS

During these 5 years, not less than a roller coaster ride, there have been many people who stood by me and influenced me to reach my goal. When it's finally time to thank them all, I feel words aren't enough.

I would like to express my sincere gratitude to my guide, Dr. Vimi Louis for believing in me and giving me a very challenging problem as my Ph.D work. Apart from giving me the freedom and space to design my experiments, her valuable suggestions helped me shape my project and to a great extend explore my scientific abilities.

I would like to thank Prof. H.S. Savithri, my mentor and an inspiration to budding scientists in the country. Working in one of the oldest and most prestigious "HSS lab" in the Biochemistry department, IISc was more than anything I have ever wanted. I am highly indebted to Mam for admitting me to her lab and for those long and fruitful discussions during the work which helped me to understand and love my subject more than ever. It has been an immense pleasure and pride to have been associated with Prof. N. Appaji Rao who apart from giving me some important suggestions in my work, narrated anecdotes that was worth listening and absorbing from.

My sincere gratitude all my advisory committee members Dr. Anita Cherian, Dr. M. Mini, Dr. Abida P.S., Dr. D. Girija and Dr. Smita Nair not just for their valuable suggestions during the review presentations at various stages of my work but also for keeping my morale high whenever I had to face a new hurdle. I would also like to express my gratitude to Dr. Babu T.D (Amala Cancer Research Institute) for his valuable to help during the most crucial time. I thank all teachers of Department of Plant Pathology Dr. Beena, Dr. Reshmy, Dr. Deepa, Dr. Sainamole Kurien, Dr. Gleena, and Dr. Rashmi for being the source of constant support during these 5 years. I am greatly obliged to all the non-teaching staffs of Banana Research Station, Kannara and Dpt. Of Plant Pathology, COA for they have always encouraged me to stay cheerful and courageous. The confidence they all have shown in me was the fuel for me to go on. I like to thank Namitha chechi for the guidance throughout; all the scientific and non-scientific discussions with her have been very useful. I thank all my BRS Kannara labmates especially Vinny, Litty and Athira for being wonderful friends always and helping me in every aspect.

It was Dr. Jayanta Saha and Dr. Birendranath Panja who influenced me the most and evoked a scientific temper during my M.Sc which greatly helped in shaping decision making and analytical skills in me. They will always be a vital part of my career.

I am grateful to all the wonderful members of HSS lab, Dr. Pallavi, Dr. Preeti, Dr. Arindam, Dr. Gowri, Dr. Vishnu, Rajani Mam and Sushma for investing their

valuable time in teaching me new techniques and for those long and fruitful discussions. If not for their support my thesis would be incomplete. I particularly thank Alan, Midhuna, Alinda and Sowmya for successful collaborations during my Ph.D which was a tremendous learning experience too.

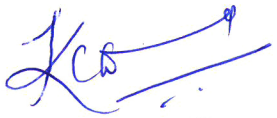
Special thanks are due to the all my batchmates, seniors and juniors and for their cooperation and for keeping my spirits high. I particularly thank Sumbula chechi and Sanju chechi for being so supportive. Thanks to Shobha, Ranji, Manohar, Indhu, Srilakshmi chechi, Smitha Miss for their friendships, personal and professional support. I thank Shakila Mam and her team for all the administrative help throughout the period of my Ph.D. I thank all my other friends for all their love and support.

I have to thank all the teaching and non-teaching staffs of the department of Microbiology and CPBMB, where I completed most of my minor and supporting courses. The practical exposure that I obtained from these courses were immense I am highly indebted to each one of the teachers for helping me while analysis and thesis writing.

I am deeply obliged to my parents forever, for not just giving me the freedom to dream but also for supporting to achieve it, whatever it takes. Thank you Achan and Amma for teaching me the difference between good and bad at the same time leaving the choices to my discretion. You are my role models and this thesis is dedicated to both of you.

My sister, she channelized my energy throughout this exercise. Thank you for being the most supportive sibling at the same time a constructive critic. I am grateful to my husband, Aravind, for being the source of my energy, for holding me up and being so appreciative during every stage of my work. I thank my brother-in-law and my extended family for their love and encouragement.

I bow my head before the Almighty for all the opportunities given to me. Success encouraged me; obstacles made me stronger. Moreover, I cannot thank him enough for introducing me to the best people wherever I go.



Darsana Dilip K.C.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

ABBREVIATIONS

%	:	Percentage
ACN	:	Acetonitrile
ASB	:	Acid salt buffer
bp	:	base pair(s)
BSA	:	Bovine serum albumin
CBB	:	Coomassie brilliant blue
cDNA	:	Complementary deoxyribonucleic acid
CP	:	Coat protein
Da	:	Dalton(s)
DAC	:	Direct antigen coating
DMSO	:	Dimethyl sulfoxide
DNA	:	Deoxyribonucleic acid
dNTPs	:	deoxy Nucleotide triphosphates
dsRNA	:	Double stranded RNA
DTT	:	Dithiothreitol
E. coli	:	*Escherichia coli*
ECL	:	Electrochemiluminescence
EDTA	:	Ethylenediaminetetraacetic acid
ELISA	:	Enzyme-linked immunosorbent assay
EtBr	:	Ethidium bromide
G	:	gram
GST-	:	Glutathione-S-transferase-tagged
h	:	hour(s)
HEPES	:	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-	:	Histidine-tagged
HRP	:	Horse-radish peroxidase

ICTV	:	International Committee on Taxonomy of Viruses
IgG	:	Immunoglobulin G
IPTG	:	Isopropyl-ß-D-thiogalactopyranoside
K	:	kilo - (103)
l	:	litre
LB	:	Luria broth
m	:	milli - (10-3)
M	:	Molar
mAbs	:	Monoclonal antibodies
MBP	:	Maltose binding protein
min	:	minute
Ni^{2+}-NTA	:	Nickel-nitrilotriacetic acid
NIb	:	Nuclear inclusion protein-b
nt	:	nucleotide(s)
ºC	:	Degree celsius
OD	:	Optical density
ORF	:	Open reading frame
PAGE	:	Polyacrylamide gel electrophoresis
PBS	:	Phosphate buffer saline
PCR	:	Polymerase chain reaction
pI	:	Isoelectric point
PIPO	:	Pretty Interesting Potyviridae ORF
PISPO	:	Pretty interesting sweet potato potyvirus ORF
Poly (A)	:	Polyadenylated
RdRp	:	RNA-dependent RNA polymerase
RMSD	:	Root mean square deviation
RNA	:	Ribonucleic acid
Rpm	:	revolutions per minute
RT	:	Room temperature

S	:	second
SDS	:	Sodium dodecyl sulphate
TEM	:	Transmission electron microscopy
TEMED	:	N,N,N',N'-tetramethylethylenediamine
TFA	:	Trifluoroacetic acid
Trx-	:	Thioredoxin-S-His-tagged
Tween-20	:	Polyoxyethylene (20) sorbitan monolaurate
UTR	:	Untranslated region
UV	:	Ultraviolet
V	:	Volts
v/v	:	volume/volume
VLP	:	Virus like particles
w/v	:	weight/volume
λmax	:	Maximum wavelength
α	:	Alpha
β	:	Beta
βME	:	β-Mercaptoethanol
μ	:	micro - (10^{-6})

Introduction

1. INTRODUCTION

Banana (*Musa paradisiaca*) is the most consumed and produced fruit in India. The yield of this highly remunerative fruit crop is challenged by various pests and diseases. Among the devastating diseases infecting banana, viral diseases caused by *Banana bunchy top virus* (BBTV), *Banana bract mosaic virus* (BBrMV), *Cucumber mosaic virus* (CMV), and *Banana streak virus* (BSV) are economically important. Yield loss ranging from 40- 100 per cent, are incurred annually directly or indirectly due to viral infections. Direct effects of viruses include stunting of growth, reduction in number of fingers and bunch weight. Nevertheless, viruses predispose plants to various other biotic and abiotic stresses thus indirectly affecting the yield (Estelitta, 1998). Since these viruses are transmitted in vegetative propagules or micropropagules and there is a latent phase during initial stage of infection cycle, it restricts germplasm exchange between countries. Considering the above reasons, undoubtedly viral infections result in long-term devastating effects since the clonal progeny of infected plants are automatically infected by the virus.

Hitherto, banana cultivars that are resistant or tolerant to these viruses have not been reported. Considering the fact that management strategies have not been devised for these viral diseases, early detection and roughing of infected plants are the possible solutions to reduce economic loss especially when micropropagated.

Among the four viruses infecting banana, BBTV and BBrMV are the most invasive. BBTV infection have reduced area under cultivation of banana in the "Old World" drastically in the past decade. It belongs to the genus *Babuvirus* in the family *Nanoviridae*. This icosahedral virus packs six circular ssDNA segments in separate capsids of 18-20 nm diameter (Wanitchakorn et al., 1997). BBrMV, a potyvirus, consist of ssRNA genome encapsidated in a flexuous rod like particle arranged in a helical symmetry. The (+) ssRNA virus encodes for a polyprotein which is modified after translation into peptides of various function by its own proteases (Rodoni et al., 1997; Urcuqui-Inchima et al., 2001). This virus which was only prevalent in the "Old World" countries, recently have been reported in Columbia and Costa Rica (Kumar et al., 2015a). Reports suggest that the virus has not only spread to "New World" but also expanded its host range to cardamom and flowering

ginger (Siljo et al., 2012; Wang et al., 2010). Both these viruses are the most prevalent in Kerala.

Lately, many nucleic acid based detection methods have been developed for detection of BBTV and BBrMV. Although these methods are rapid and more sensitive than serological assays, the requirement of skilled personnel and sophisticated infrastructure for the same makes them less preferable for diagnosis. Nevertheless, according to WHO a perfect diagnostic technique would be "ASSURED" *ie.*, Affordable, Sensitive, Specific, User friendly, Rapid and robust, Equipment free, and Deliverable. Molecular based techniques are indeed specific, sensitive and rapid but definitely cannot be user friendly as they require sophisticated equipment and demand expertise. Serological based assays can be perfected to fit in all the parameters of good diagnostic assay must have, provided, it is made more specific and sensitive.

In case of banana, when antiserum is raised against viruses partially purified from infected plant samples, phenols and secondary metabolites present in the host give rise to a lot of interference (Selvarajan et al., 2010a). Inconsistent and low concentration of partially purified virus preparation is another problem encountered. This notably challenges the specificity of the antisera as well as considerably reduces sensitivity of the assay (Hema et al., 2003). Development of recombinant protein for immunising the animal to raise antiserum greatly solves the above problem and help to develop more reliable and sensitive serological assays.

The DNA-S segment of BBTV consisting of a single Open Reading Frame (ORF) coding for the capsid protein and the CP gene at the 3' end of the BBrMV ssRNA coding for 34 kDa coat protein are targeted for detection in the present study. Compared to DNA viruses, RNA viruses are mutated rapidly due to the lack of proof reading mechanism in RNA dependent RNA polymerase replicating the genome (Domingo, 1997). Potyviruses are no exception (Nigam et al., 2019). However, a new isolate of BBTV was identified and characterised from Umiam, Meghalaya which was apparently a recombinant (Banerjee et al., 2014). In this context it is inevitable to analyse BBTV and BBrMV population in Kerala and understand the driving forces behind their evolution and devise methods for detection taking into account evolution of viruses.

The experiments were laid out systematically according to the objectives enlisted below

- To develop nucleic-acid based detection of the viruses considering their evolution
- To develop recombinant coat protein for immunodetection of BBTV and BBrMV
- To produce antiserum against any one of the recombinant coat proteins and demonstrate the efficacy when compared to the commercially available antiserum in various serological assays

Review of Literature

2. REVIEW OF LITERATURE

Viruses infecting banana

India contributes 29 per cent of total world production of banana which is most remunerative fruit crop grown in tropical and subtropical regions of the country (FAOSTAT, 2018). Banana is grown in 8.3 lakh ha with a productivity of 35.9 mt/ ha. It is also the most consumed fruit in India. Tamil Nadu is the leading producer of banana in India followed by Maharashtra, Gujarat, Andhra Pradesh and Karnataka. In Kerala, 28 thousand ha are under banana cultivation with a productivity of 14 mt/ ha, Nendran being the most popular variety (NHB, n.d.). Other varieties like Poovan, Red banana, Njalipoovan are also preferred in Kerala domestic markets.

A major challenge faced by banana growers in Kerala are the viral diseases that causes considerable yield losses in addition to increase the cost of cultivation. *Banana bunchy top virus* (BBTV), *Banana bract mosaic virus* (BBrMV), *Cucumber mosaic virus* (CMV) and *Banana streak virus* (BSV) are major causes of economic loss (Kumar et al., 2015a; Tripathi et al., 2016). These viruses which are primarily transmitted through suckers are not indexed before it is distributed to the farmers at various outlets. Even with the advent of tissue culture (TC) for micropropagation of banana, virus indexing of the planting materials are not popularised among the farming community. Although, the quality of the TC material is assured, the cost is more than double that of the suckers. However, the awareness that buying quality planting material can reduce the cost of cultivation considerably have to be advocated among the farmers.

This chapter discuss BBTV and BBrMV in detail under various sub headings.

2.1 BANANA BUNCHY TOP VIRUS

2.1.1 History and epidemiology of banana bunchy top disease

Banana bunchy top virus (BBTV) causing banana bunchy top disease (BBTD) was first reported in Fiji in 1889 threatening its banana export industry (Magee, 1927; Simmonds, 1935; Taylor, 1969). It is one of the most devastating diseases in banana especially in tropical regions especially South Pacific, Asia and Australia (Dale, 1987; Elayabalan et al., 2015). Countries like India, Sri Lanka, Bangladesh, Myanmar, China, Indonesia, Japan, Philippines, Thailand Pakistan and Malaysia in Asia (Vakili, 1969; Wardlaw, 1972; Dale,

1987; Panhwar, 1991; Kawano and Su, 1993; Khalid et al., 1993), Vietnam, Fiji, Australia, Samoa, Hawaii, Kiribati and Tuvalu in Pacific region (Magee, 1927; Eastwood, 1946; Vakili, 1969; Dale, 1987) and African countries like Cameroon, Angola, Rwanda, Burundi, Congo and Egypt (Magee, 1927; Wardlaw et al., 1961; Diekmann and Putter, 1996; Sebasigari and Stover, 1988; Oben et al., 2009; Kumar et al., 2009) are most effected by the disease.

Hitherto, the disease is not reported from Central and South America or Western Australia (Thomas et al., 1994; Diekmann and Putter, 1996; Kenyon et al., 1997; Thomas and Iskra-Caruana, 2000; Kagy et al., 2001). Also, confirmation of BBTV infection in Caribbean, Papua New Guinea and few other banana growing countries in Asia and Africa has not been established (Karan et al., 1994; Singh, 2003; Mandal, 2010).

In India, the disease is widespread in Kerala, Gujarat, West Bengal and Tamil Nadu (Rao et al., 2002). It has also been reported from different regions of Uttar Pradesh and recently in Meghalaya (Vishnoi et al., 2009; Banerjee et al., 2014) (Plate 1). In Kerala, BBTV infects banana throughout the state irrespective of the variety cultivated causing yield loss even up to 100 %. There have been no reports of natural resistance in any of the cultivars to BBTD till date (Magnaye and Valmayor, 1995). Banana bunchy top virus is deemed as 'quarantine pest' of very high importance due to its huge destructive potential. Moreover, it is listed in the 100 of the World's Worst Invasive Alien Species by Invasive Species Specialist Group (ISSG) of the World Conservation Union-International Union for conservation of nature (IUCN) (Lowe et al., 2000).

2.1.2 Symptoms

Typical symptom of the BBTD is the rosetting of the leaves that are smaller in size than normal, giving a bunchy appearance, from where the disease gains its name. Dark green broken streaks on leaf veins, petiole, midribs are observed. These dot dash like patterns sometimes resembles Morse code and are seldom seen on pseudostem (Thomas and Iskra-Caruana, 2000). The short and narrow leaves become chlorotic and brittle which tends to show epinasty, finally succumbs to necrosis. Banana infected at a later stage of development produces malformed male bud or sometimes bears no fruit at all (Elayabalan et al., 2015; Qazi, 2016). Latent infection of BBTD have been reported in Taiwan and in Hill banana grown in Palani hills of Tamil Nadu (Diekmann and Putter, 1996; Elayabalan et al., 2015).

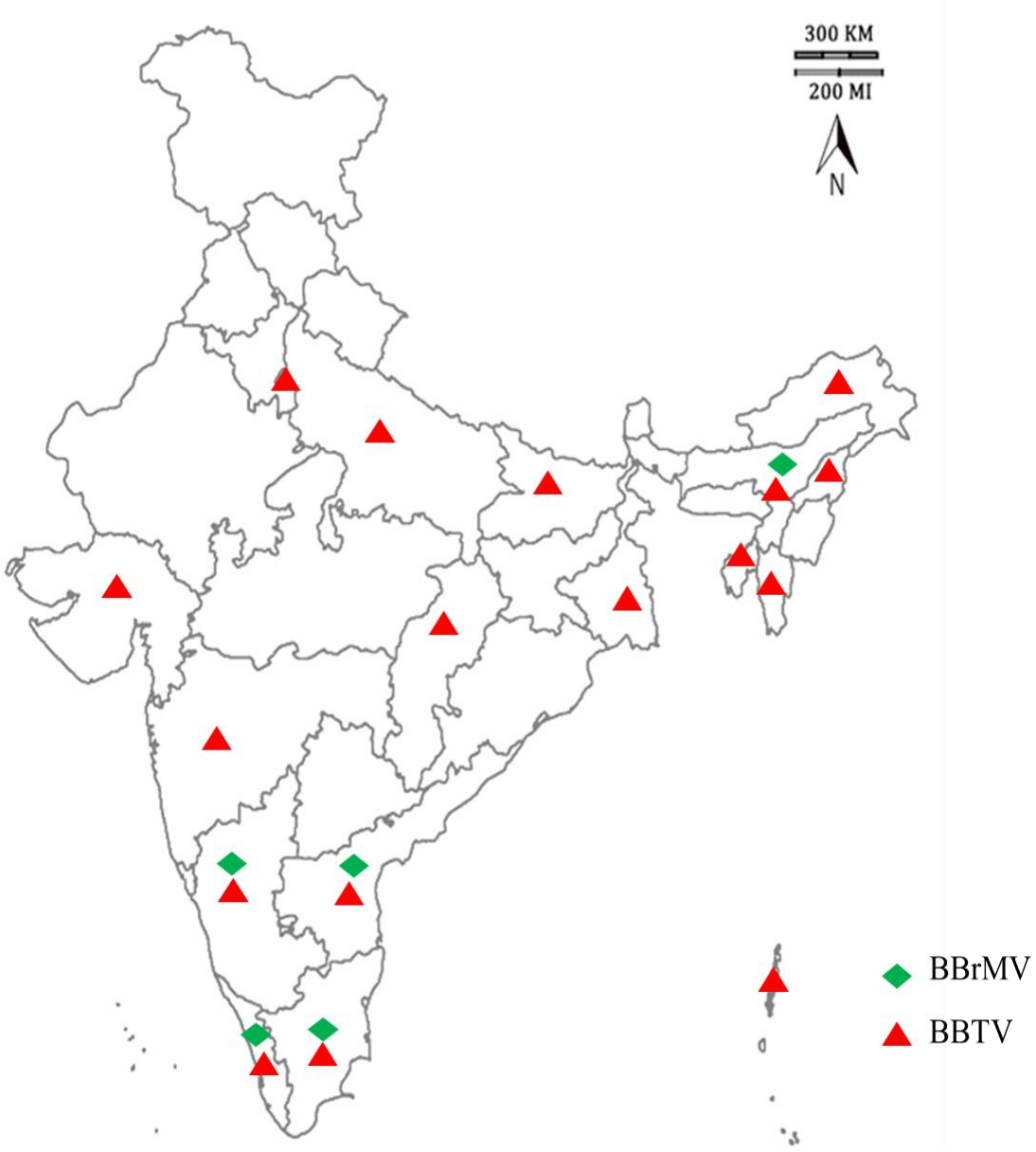


Plate 1 Distribution of BBTV and BBrMV in India

2.1.3 Disease transmission

The BBTV is transmitted by banana aphids, *Pentalonia nigronervosa* Coquerel (*Hemiptera*: *Sternorrhyncha*: *Aphididae*) persistently but is non-propagative in the aphid vector which requires a latent period of 20-28 h before transmission (Magee, 1927; Anhalt and Almeida, 2008). According to Hu et al., (1996) acquisition feeding period of at least 4 h and inoculation feeding of period of more than 10 min are required for successful transmission with an efficiency of up to 76 %. Although, depending on temperature and age of plants, susceptibility to BBTV varies and symptoms start to appear after 25 days following the transmission (Allen, 1978). Long distance spread of the virus is by the planting materials like suckers, corms and also through tissue-cultured plants (Drew et al., 1989). However, mechanical transmission of BBTV from banana to banana or banana to any other indicator plant has been unsuccessful (Magee, 1967; Hafner et al., 1995; Thomas et al., 2003).

2.1.4 Host range

BBTV is highly specialised and reported to cause typical symptoms in most of the cultivated varieties of banana (*Musa paradisiaca*) and other species in genus *Musa* including *M. balbisiana*, *M. jackeyi*, *M. acuminata*, *M. textilis*, *M. coccinea*, *M. velutina*, *M. ornata* (Magee, 1927, 1948; Thomas and Dietzgen, 1991; Espino et al., 1993; Thomas and Iskra-Caruana, 2000; Furuya et al., 2003) and closely related *Ensete ventricosum* (Wardlaw, 1961). However, in the account of banana aphids colonising many other plants belonging to *Araceae*, *Commelinaceae*, *Musaceae* and *Zingiberaceae* (Blackman and Eastop, 1984) stimulated various scientists to determine whether BBTV can infect other hosts too. The reports of BBTV infecting *Canna indica* (*Cannaceae*) and *Hedychium coronarium* (*Zingiberaceae*) in Taiwan was result of such studies (Geering and Thomas, 1997; Yasmin et al., 2001). Chronicles of alternative hosts of this virus require further confirmation as crops proven to be hosts of BBTV in a region was proved to be a non host in some other regions (Ram and Summanwar, 1984; Hu et al., 1996; Geering and Thomas, 1997).

2.1.5 Genome organisation of BBTV

Banana bunchy top virus belongs to the genus *Babuvirus* in the family *Nanoviridae* (Harding et al., 1991; Vetten et al., 2005). It has multipartite genome of six single stranded

circular DNA (ssDNA), similar in organization and size (approximately 1 kb) designated as DNA-R, DNA-U3, DNA-S, DNA-M, DNA-C and DNA-N (Harding et al., 1991; Burns et al., 1995; Wanitchakorn et al., 2000a; Vetten et al., 2005) (Fig 1). Each of the six genomic components encompasses an intergenic region (IR) and one ORF transcribed into structural and non-structural proteins of various functions.

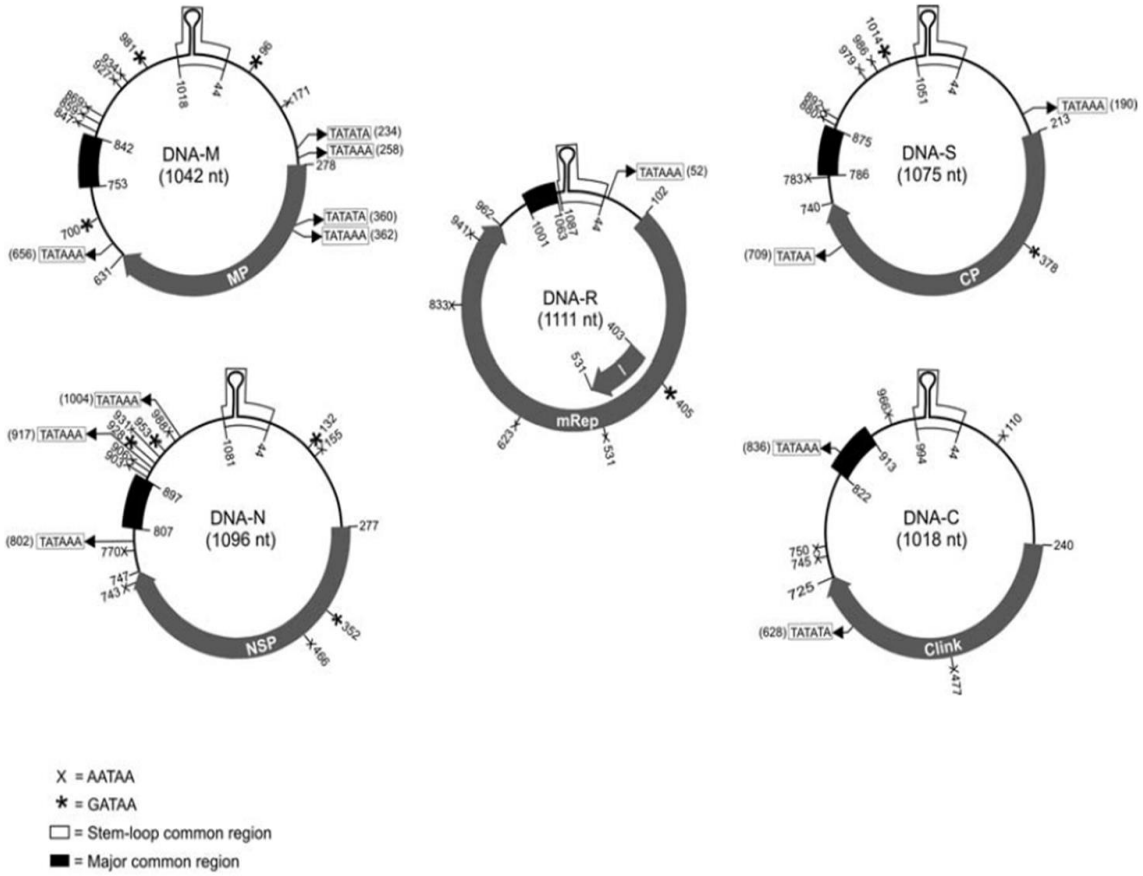


Fig 1. Genome organisation of BBTV. Genomic components except DNA-U is represented in the figure. The ORF is denoted as grey arrow representing its orientation. DNA-R encoding an additional gene is represented as I, function of which is unknown. (Source: Amin et al., 2008)

The IR of each genomic component comprises of a stem-loop common region (SL-CR), a major common region (CR-M), a TATA box and a polyadenylation signal (Burns et al., 1995). Role of CR-M in second strand synthesis of circular ssDNA has been established (Hafner et al., 1997a). It consists of three relatively conserved domains *viz.*, I, II and III and short primer sequences. IR also comprises of promoter and terminator for the ORFs (Dugdale et al., 1998).

DNA-R encodes the master replication initiation protein or Rep protein which helps in replication of the BBTV genomic components (Horser et al., 2001a). It comprises of domains for origin specific endonucleolytic, nucleotidyl transferase, helicase and ATPase activities. The closed circular ssDNA of the virus is first converted to supercoiled dsDNA. Rep protein binds to the hairpin structure at the origin of replication (ori) which is single stranded and introduces an endonucleolytic nick in the AT rich conserved nanonucleotide sequence (5'-AGTTATTAC-3') within the SL-CR of IR thereby initiating rolling circle replication (RCR) (Horser et al., 2001). Rep protein binds to the 5' end resulting in a free 3' end acting as a primer for cellular DNA polymerase. Replication is terminated when a Rep-catalysed nucleotidyl transfer reaction releases a circular single-stranded virus genome (Hafner et al., 1997b).

DNA-S codes for approximately 20 kDa coat protein (CP) which assembles to isometric capsid that encapsidates the viral genome (Wanitchakorn et al., 1997). DNA-M and DNA-N encode the movement protein (MP) and nuclear shuttle proteins (NSP) which localises in cell cytoplasm and nucleus, respectively. However, it has been demonstrated that when both these proteins are expressed together NSP moves to periphery of the nucleus by the interaction of 29 amino acid N-terminal hydrophobic region of the movement protein. The NSP protein does not essentially react with the coat protein (Wanitchakorn et al., 2000a).

DNA-C is translated to the cell cycle link protein (C-link) that binds to plant retinoblastoma-like protein (Rb) through its LXCXE motif to manipulate the cell cycle. This protein-Rb complex has the ability to switch cell in any stage to S phase in order to enhance viral DNA replication (Aronson et al., 2000; Wanitchakorn et al., 2000a). The Rb binding motif in mastrevirus is located in the Rep protein where as in nanovirus, it is present in a protein encoded by another genomic component. However, the Rep protein encoded by AC1 gene of begomovirus binds to Rb but LXCXE motif is absent in this protein (Acotto et al., 1993; Ach et al., 1997). The functions of the DNA-U3 gene product and that encoded by the small internal ORF of DNA-R are unidentified (Burns et al., 1995; Beetham et al., 1997).

2.1.6 Satellite molecules associated with BBTV

Additional circular ssDNA components have been found associated to South East Asian isolates that encodes Rep-like proteins in addition to the master Rep protein. These proteins are involved in self replication. The satellite is similar to that of the other BBTV genomic

components not only in size but also in the presence of CR-SL (Horser et al., 2001b; Amin et al., 2008). However, CR-SL is not conserved as in other genomic components. Satellite also differs as it lacks an internal ORF and CR-M. Moreover, TATA box is present at the 5' end of the stem loop. The sequence of satellite is similar to that of geminivirus-associated Rep- like alpha satellites however, its phylogeny is more related to that of Rep proteins of *Nanovirus* than *Babuvirus* (Horser et al., 2001b).

2.1.7 Structure of virion

The CP of ~20 kDa size assembles into 18-20 nm icosahedral particles composed of 20 faces and 30 edges with triangulation number one (T=1) (Tsao, 2008). Each genome component is packed in a different capsid. Harding and co-workers (1991) purified BBTV virions using caesium sulphate density gradient centrifugation and observed a major protein with molecular weight 20 kDa in the SDS gel profile. The assembled particles of 18-20 nm banded as a single component with a density of 1.28 to 1.30 g/ml in caesium sulphate.

2.1.8 Phylogenetics and molecular diversity of BBTV

Banana bunchy top virus is believed to have originated from South and Southeast Asian-Australasian region *ie.,* from the centre of origin of banana (Perrier et al., 2011; Stainton et al., 2012). The early reports of the virus infecting banana starting from Fiji in 1889 followed by Egypt in 1901 (source unknown), Australia and Sri Lanka in 1913 reinstates such theories.

Previously phylogenetic studies conducted were based on DNA-R gene sequences yielding two groups of BBTV isolates in the world namely the South Pacific group and the Asian group (Karan et al., 1994; Hu et al., 2007; Stainton et al., 2012; Wickramaarachchi et al., 2016). Yu et al. in 2012 revised the grouping to the Pacific India Oceans (PIO) group (encompasses the isolates from Australia, Egypt, Hawaii, India, Myanmar, Pakistan, Sri Lanka and Tonga) and South East Asia (SEA) group (encompasses the isolates from China, Indonesia, Japan, Philippines, Taiwan and Vietnam). Due to germplasm exchange of banana planting material, the PIO group is seen to have a broader distribution worldwide than the SEA group (Qazi, 2016).

Based on analysis of all the available sequences of BBTV genome, two theories of evolution have been proposed. According to the latest theory the two main BBTV ancestries could have diverged only hundreds of years ago as result of mutation and

frequent homologous inter genomic (Hu et al., 2007; Hyder et al., 2011) and intra genomic recombinations (Hu et al., 2007). The old theory however, states that the divergence of the two groups date back to the evolution of banana cultivation itself (Perrier et al., 2011).

On analysing sequences of all the six genomic components of the multipartite virus in the database, DNA-R was found to be the most conserved with more than 88 per cent nucleotide sequence identity among the isolates, while DNA-U3 was least conserved. DNA-S, DNA-C, DNA-N and DNA-M, recorded sequence similarity of 80 per cent among all isolates (Stainton et al., 2012). The genetic diversity of BBTV isolates within Pakistan, Indonesia, Africa and Oceania was reported to be low and overall genetic diversity of BBTV isolates was calculated to be approximately 15 % (Amin et al., 2008; Selvarajan et al., 2010b; Stainton et al., 2012; Adegbola et al., 2013; Chiaki et al., 2015). However, in the North Eastern region of India, relatively greater diversity for BBTV was detected (Banerjee et al., 2014). Identification of two more viruses in genus *Babuvirus*, *Abaca bunchy top virus* infecting abaca (*M. textilis*) and banana (*Musa sp*.) and *Cardamom bushy dwarf babuvirus* (CBDV) infecting cardamom are indicative of the diversity in the genus (Sharman et al., 2008; Mandal et al., 2013).

2.1.9 Recombination in BBTV

The evolutionary and fitness advantage of segmented genome of plant viruses is testified in various studies (Szathmary, 1992; Chao, 1988; Vetten et al., 2005; Ojosnegros et al., 2011). The replication advantage of shorter genome over the replication of a full length genome is undebated (Nee, 1987; Holmes, 2009). Although, there might be many other undiscussed reasons for evolution of segmented genomes, the consequence of which is increased genetic diversity in the virus population. Genetic reassortment mechanisms among different isolates of virus have been reported to cause variation and lately, that between the segments of the same virus also has been resulting in increased genetic diversity.

Despite being a DNA virus, BBTV genome segments have exhibited high genetic variation attributed to intergenomic recombination and seldom to intragenomic recombination. However, intra and intergenomic recombination events have been detected in DNA-U3 segments of isolates from Pakistan (Hyder et al., 2011) and fewer incidences of recombination are detected in DNA-S and DNA-R segments (Lefeuvre et al., 2009). In DNA-U3 segment recombination hot spots were detected in and around stem loop region

(Hyder et al., 2011) which are also well defined recombination hot spots in *Geminiviridae* (Lefeuvre et al., 2007, 2009; van der Walt et al., 2009). It is testified by Hyder and co-workers (2011) that in this region a nick is created by viral Rep protein (Arguello-Astorga et al., 1994; Hafner et al., 1997b; Wanitchakorn et al., 2000a) while rolling circle replication, thus subjected to high levels of recombination. This mechanism is explained to speed up the process of adaptation and resist the fixation of lethal mutations in virus population (Keightley and Otto, 2006).

2.2 BANANA BRACT MOSAIC VIRUS

2.2.1 History and epidemiology of banana bract mosaic disease

Banana bract mosaic disease (BBrMD) also known as "Kokkan" was first reported from Davao island of Mindanao in Philippines in the year 1979 (Magnaye and Espino, 1990; Thomas and Magnaye, 1996; Rodoni et al., 1997). The disease subsequently spread all over in Philippines and also to major banana growing parts of India, Sri Lanka, Vietnam and Western Samoa. The disease was initially reported from the 'Old World', but recently it has been reported in Columbia, Ecuador and Costa Rica (Quito-Avila et al., 2013) indicative of the spread of BBrMV which is suspected predominantly through export of planting materials (Rodoni et al., 1999).

In India, the disease was first spotted in Kerala, however, the aetiology was unidentified then (Balakrishnan et al., 1996). BBrMD is cosmopolitan in Kerala. The disease causes a heavy economic loss of 40-60 per cent yearly to banana industry in Kerala alone (Cherian et al., 2002). In India, BBrMD is reported from all the major banana growing states like Tamil Nadu, Andhra Pradesh, Karnataka and Assam (Selvarajan and Jeyabaskaran, 2006) (Fig. 1).

2.2.2 Symptoms

The distinguishing symptoms of BBrMD include dark reddish-brown mosaic pattern on the bracts of the inflorescence from where the name of the virus originates. This symptom discerns it from all the other viral diseases of banana. Initial symptoms include reddish brown spindle shaped lesions or streaks running parallel to the veins on the pseudostem, petioles and midrib (Rodoni et al., 1997; Thomas et al., 1997; Selvarajan and Jeyabaskaran, 2006; Kumar et al., 2015). However, symptoms on leaf lamina may or may not be conspicuous (Thomas and Magnaye, 1996). Distinctive dark purple to red coloured mosaic

patterns, stripes or spindle shaped streaks are observed when dead sheaths are pulled away from the pseudostem. At severe conditions the infected leaf sheaths are seen separated from the pseudostem by itself. Seldom observed symptoms include 'Traveller's palm' like congested and parallel arrangement of leaves and black streaks on various plant parts (Balakrishnan et al., 1996). Mosaic patterns are visible on bunch stalks and reduction in the bunch weight and size of fingers are observed with considerable decline in market value of the bunch resulting in huge economic loss (Magnaye and Espino, 1990; Thangavelu et al., 2000; Cherian et al., 2002).

2.2.3 Disease transmission

Banana bract mosaic virus, as BBTV is primarily transmitted through infected planting material. But unlike BBTV, it is transmitted non persistently by several aphid species, apart from *P. nigronervosa, viz., Rhopalosiphum maidis, Aphis gossypii* (Magnaye and Espino, 1990; Munez, 1992) and *A. craccivora* (Selvarajan et al., 2006). The attempts of Dhanya (2004) to mechanically transmit BBrMV through graft and soil were unsuccessful. Vector transmission of BBrMV, however, with *Pentalonia nigronervosa* Coq. resulted in 40 per cent transmission with pre-acquisition fasting of 1 h and acquisition threshold of 30 min. Recently, BBrMV was proved to transmit through seeds in synthetic diploid, H-201 that were raised from embryos of infected seeds of naturally infected plants in the field. Both *in vitro* and potted plants showcased typical BBrMV symptoms (Selvarajan et al., 2020a). BBrMV transmission through seeds is extremely alarming especially in synthetic diploid H-201 which is exclusively developed for panama wilt, nematodes and sigatoka disease resistance breeding programme (Krishnamoorthy, 2002).

2.2.4 Host range

Banana bract mosaic virus belongs to *Potyvirus* genus of *Potyviridae* family, one of the largest genus of plant viruses representing approximately 30 per cent of all the plant viruses identified up to now. Its host range as a whole is broad although that of individual species is restricted (Desbiez et al., 2014; Sharma et al., 2014). *Potyvirus* infects few members of family *Caricaceae, Rosaceae, Amaranthaceae,* and in the order *Zingiberales*, apart from crops belonging to *Solanaceae, Cucurbitaceae*, *Leguminosae* and some ornamental plants causing considerable yield loss in tropical and subtropical countries (Hollings and Brunt, 1981; Wang et al., 2010; Kumar et al., 2015; Sharma et al., 2014; Siljo et al., 2012).

Studies on host range of BBrMV showed that it causes disease in all banana genotypes especially Cardaba (ABB/BBB), Saba (BBB) and Abuhon (BB) (Thomas and Magnaye, 1996). Conversely, on screening the germplasm collection at Banana Research station, Kannara for BBrMD revealed A genotype succumb to the disease than B genotype. The study also highlighted that all the commonly cultivated varieties in Kerala were susceptible to BBrMV (Dhanya, 2004).

Banana bract mosaic virus was believed to be highly specialised, as any other virus in the genus *Potyvirus,* infecting *Musa* sp. including abaca (*M. textilis*), until infection on cardamom (*Elettaria cardamomum* Maton) in India and flowering ginger (*Alpinia purpurata* (Vieill.) K. Schum) in Hawaii was reported (Sharman et al., 2000; Wang et al., 2010; Siljo et al., 2012).

2.2.5 Genome organisation

Banana bract mosaic virus is a positive sense single stranded (ss) RNA virus of 9,711 bp with only one functional ORF coding for a polyprotein (Fig 2) (Allison et al., 1986; Dougherty and Carrington, 1988; Balasubramanian et al., 2014). This is the characteristic of *Potyviridae* family with only genus *Bymovirus* as an exception (Berger et al., 2005). The polyprotein is post-translationally modified by its own proteases to 10 functional polypeptides (Urcuqui-Inchima et al., 2001) and PIPO (Pretty Interesting *Potyviridae* ORF) expressed by an internal frame shift (Chung et al., 2008) as in all other potyvirids. The 5' terminal protein (VPg) is covalently linked to this genomic RNA and it has a poly A tail at the 3' terminal (Riechmann et al., 1992). The 5' untranslated region (UTR) mediates cap-independent translation of the polyprotein (Kneller et al., 2006).

The polyprotein precursor is cleaved at susceptible Q-S (glutamic acid- serine), Q-G (glutamic acid- glycine), and Q-A (glutamic acid- alanine) residues by three virus proteinases (P1, HC-Pro and NIa-Pro) into P3 protein, 6K1 protein, cytoplasmic inclusion (CI) protein, 6K2 protein, nuclear inclusion a protein (NIa), genome linked viral protein (VPg), nuclear inclusion b protein (NIb) and coat protein (CP) (Allison et al., 1986; Shukla and Ward, 1988). Most of the cleavage reactions are catalysed by trans-proteolytic or autoproteolytic mechanisms of small nuclear inclusion protein (NIa-Pro) whereas, P1 and the helper component proteinase (HC-Pro), catalyse autoproteolysis at their respective termini (Dougherty et al., 1988).

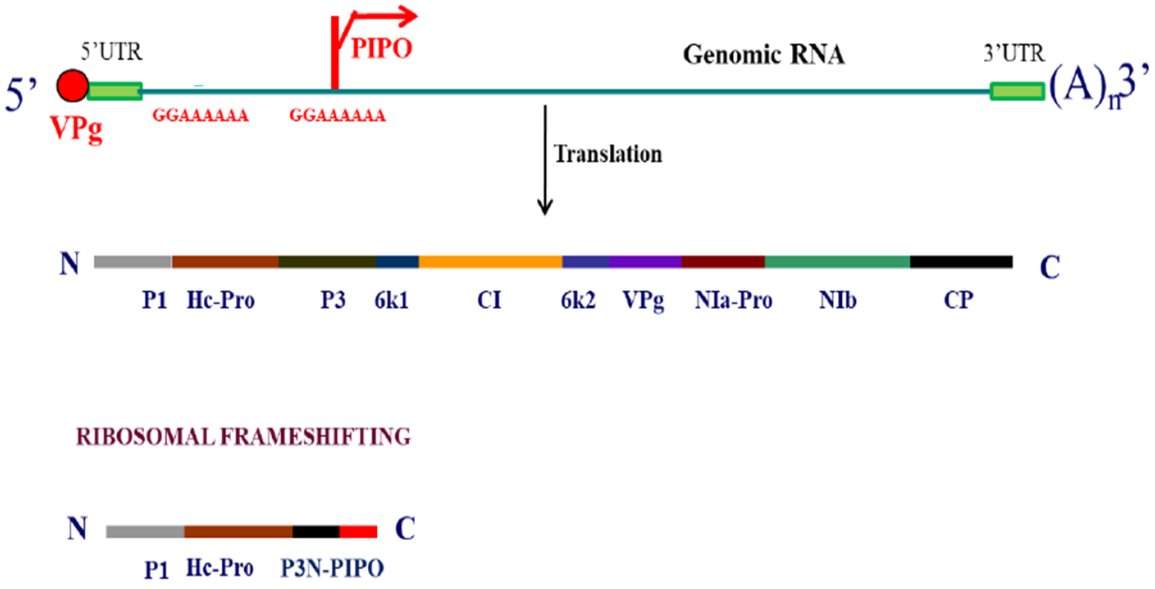


Fig 2. Genome structure and organisation of BBrMV (Adopted and modified from Sabharwal, 2017)

Other than protease activity, HC-Pro facilitates aphid transmission of the virus by interacting with DAG (aspartic acid- alanine- glycine) motif of the capsid protein with its PKT (proline-lysine- threonine) motif (Hull, 2009). The protein is also found to be important in systemic movement of the virus together with 6K2. Development of symptoms are facilitated by HC-Pro, P1 and CI. NIb is the viral RNA dependent RNA polymerase (RdRp) whereas RNA helicase for RNA replication is CI. 6K1 not only enables cell to cell movement but also plays a major role in virulence (Sochor et al., 2012; Balasubramanian et al., 2014).

Coat protein ORF is situated at the 3' end of the ssRNA. The coat protein of *Potyvirus* is proven to be multifunctional as it has significant roles in virion assembly, aphid transmission and cell to cell and systemic movement (Sochor et al., 2012; Balasubramanian et al., 2014).

2.2.6 Structure of virion

The approximately 10 kb genome of BBrMV is encapsidated in long flexuous rod particle of average dimension 725 x 12 nm composed of ~2000 units of capsomeres with molecular weight of 34 kDa arranged in a helical symmetry as in other flexuous rod shaped potyviruses (Anindya and Savithri, 2003; Balasubramanian and Selvarajan, 2014). The

potyvirus nucleocapsid helix consists of minimum seven and not more than nine coat protein (CP) subunits per ring with basic pitch of is 3.3 nm (Goodman et al., 1976; Kendall et al., 2008; Zamora et al., 2017). The coat protein structures of *Watermelon mosaic virus* (WMV) and *Turnip mosaic virus* (TuMV) has been deciphered using cryo-electron microscopy (Zamora et al., 2017; Cuesta et al., 2019). The CP consists of mostly α helices. The secondary structure of N and C terminal is large coiled-coil domain which is flexible. These surface exposed N- and C-terminal arms of CP mediate the polymerisation of CP subunits in the virion (Anindya and Savithri, 2003). The flexuous nature of viral particles can be attributed to the relative movements between CPs facilitated by the N and C terminals (Zamora et al., 2017).

2.2.7 Phylogenetics and molecular diversity of BBrMV

The RdRp of RNA viruses lack proofreading activity, causing errors due to mis insertion in the range of 10^{-3} to 10^{-5} substitutions per nucleotide and per round of replication leading to genetic variability (Steinhauer et al., 1992; Domingo and Holland, 1997). This high rate of mutation and short period for replication favour formation of viral quasi species which can easily adapt to the changes in the environment (Eigen et al., 1988). The coat protein of BBrMV also showcases high variability due to the above reason. The mutations and substitutions however, are fixed in the population, if they have any fitness advantage (Domingo and Holland, 1997).

As mentioned earlier, the CP of potyvirids is multifunctional, facilitating host adaptability, systemic movement and infection in the host and plant- to- plant transmission (Sochor et al., 2012). However, this coat protein is also one of the most variable proteins encoded by the potyvirid genome. It differs across the strains of same virus and different viruses of the same genus and species (Shukla et al., 1988). Particularly, the N and C terminal is hypervariable according to Nigam et al. (2019) and these are the regions under positive selection.

The BBrMV population around the globe was evaluated for diversity and it was evident that isolates displayed 0.3- 5.6 per cent and 0.3- 4.3 per cent variability in the nucleotide and amino acid sequence level respectively (Rodoni et al., 1999). On evaluating the diversity of Indian isolates of BBrMV, based on their CP gene, Balasubramanian and Selvarajan (2014), establishes that the nucleotide and amino acid sequence diversity of the Indian isolates are more than other regions from where the disease has been reported. The

reason stated was the suspected origin of the disease from Southern part of India and its parallel evolution with the host. On assessing the evolutionary lineages of BBrMV isolates infecting *Musa* sp, cardamom and flowering ginger all seem to have diverged from a common ancestor in the course of evolution. Besides, the studies carried out based on VPg and HC-Pro also reinstated that a relationship with geography cannot be established. Furthermore, unregulated movement of planting materials across the states results in frequent gene flow among and between South Indian isolates (Balasubramanian et al., 2014; Balasubramanian and Selvarajan, 2014; Anuradha and Selvarajan, 2018).

2.2.8 Recombination in BBrMV

Recombination occurring in RNA viruses have a major evolutionary and biological significance (Simon-Loriere and Holmes, 2011). Recombination is one of major reasons for genetic diversity in RNA viruses apart from mutations which occur in relatively high frequency in *Potyviridae* (Tomimura et al., 2004; Gibbs and Ohshima, 2010).

Ecological implications of recombination can be, a method for the viral quasi species to adapt to a new host by performing genetic jumps gradually through numerous rounds of replications (Domingo and Holland, 1997). The possible explanation for host range expansion of BBrMV can be corroborated with the frequent recombination events in the genome. An intriguing result obtained from the study conducted by Balasubramanian and Selvarajan (2014) was that the minor parents of two possible recombinant isolates from Tamil Nadu were BBrMV isolates infecting cardamom both found in Southern part of India.

Recombination in BBrMV isolates based on coat protein was analysed by Balasubramanian and Selvarajan (2014) and eight isolates were detected as recombinants including one isolate from Philippines (AF071590.1) and Thailand (AF071589.1). Interestingly, the major and minor parents of Philippines and Thailand isolates were Indian isolates. Among other six Indian isolates, one was collected from Kerala. However, study based on VPg reported only one possible recombinant, KT852552.1, collected from Tamil Nadu and that based on HC-Pro gene reported four isolates one of which was collected form Philippines and the rest from India (Balasubramanian et al., 2014; Anuradha and Selvarajan, 2018). Based on these results conclusion was drawn that BBrMV subsequently moved from India to Philippines and later to other parts of the world (Balasubramanian et al., 2014).

2.3 CODON USAGE BIAS OF VIRUSES

In most of the organisms evaluated so far, it have been revealed that the alternative synonymous codons for one amino acid are not used randomly and that this non-random usage is species or taxon specific (Grantham et al., 1981; Ikemura 1985). It was fascinating to the scientific community that within the species itself there was a considerable heterogenity between the genes. Soon it was establishes that the positive correlation between degree of codon bias and level of gene expression cannot be ignored especially in the well characterised *E. coli* and yeast genome (Gouy and Gautier, 1982; Bennetzen and Hall, 1982).

Codon usage bias (CUB) can be quantified by codon bias index which measures the codon bias based on the effeciency of translation (Bennetzen and Hall, 1982). It is the degree of deviation from a postulated impartial pattern of usage also termed as codon preference bias (McLachlan et al., 1984), chi square value for the deviation from random codon usage (Sharp et al., 1986). However, codon preference statistic is based on the ratio of the likelihood of discovering a particular codon in a highly expressed gene to the likelihood of finding that codon in a random sequence with the same base composition (Gribskov et al., 1984). Apart from these, Codon Adaptation Index (CAI) is a statistical index inclusive of the role of natiral selection on codon bias proposed by Sharp and Li in 1987.

Viral genes are classic examples of foreign genes being expressed in host organims. This implies that the codon usage in the many viruses that do not have their own translation machineries and do not encode their own tRNA molecules should be perceptibily adapted to the translational machinery of the host. Thus, understanding the codon usage pattern of the host will give an insight of the expression levels of certain genes in the virus with understanding the the arms race between the host and virus leading to evolution. It was also established by Adams and Antoniw (2004) that mutational bias is major determinant of codon usage variation midst plant viruses.

Many scientists have endeavoured to define reasons for codon bias in various organisms over time. Evidently, random genetic drift (Sharp and Li, 1986; Bulmer 1991), tRNA abundance, mRNA, protein structure (Knight et al., 2001), apart from replicational, transcriptional, translational bias (Hershberg and Petrov 2008) and other environmental factors (Behura and Severson, 2013) influence codon usage pattern in different organisms.

2.4 DIAGNOSIS OF BBTD AND BBrMD

Viral diseases are effectively controlled by using resistant cultivars in the field. However, even after multiple investigations no variety was identified as a source for resistance breeding. Although, some tolerant cultivars were identified with low rate of infection to BBTV (Jose, 1981; Hooks et al., 2008, 2009; Niyongere et al., 2011) and none were found tolerant to BBrMV (Kumar et al., 2015). Moreover, resistance breeding in *Musa* sp. has not been proved to be successful due to its complex polyploid genome. In the absence of effective management strategy or resistant cultivar, use of superior quality uninfected planting material, early diagnosis and rogueing of infected plants are the best methods to control the spread of the viral diseases and reduce heavy crop loss.

Virus indexing plays a major role in providing high quality planting materials to the stakeholders, especially TC plants which are produced and cultivated with a lot of added investment. Apart from these complications, BBTV and BBrMV stay latent in the host creating an impression of infected plant being healthy and by the time conclusive symptoms are displayed, the management becomes impossible. Even otherwise symptomatology-based diagnosis is not decisive as it overlaps with nutrient deficiencies (Kumar et al., 2015).

Detection of BBTV and BBrMV are primarily based on serological techniques using polyclonal and monoclonal antibodies. The most used serological diagnostic method is Enzyme linked Immunosorbent Assay (ELISA) (Espino et al., 1989, 1990; Wu and Su, 1990b; Thomas and Dietzgen, 1991; Thomas et al., 1997) and its variants like double and triple antibody sandwich ELISA and plate-trapped antigen (PTA) ELISA. Detection of BBTV in TC plants, aphids and field samples have been developed using these methods (Wu and Su, 1990b; Thomas and Dietzgen, 1991; Geering and Thomas, 1996; Selvarajan et al., 2010). Most recently, Lateral Flow ImmunoAssay (LFIA) was developed by Selvarajan et al., (2020b) which can detect up to 10 ng/ ml of BBrMV.

Nucleic acid-based methods have been used for quick and specific detection of both BBTV and BBrMV. Polymerase Chain Reaction (PCR) for detection of BBTV (Xie and Hu, 1995; Hu et al., 1996; Thiribhuvanamala et al., 2005; Galal, 2007) and Reverse-Transcriptase PCR (RT-PCR) for detection of BBrMV (Espino et al., 1990; Bateson and Dale, 1995; Thomas et al., 1997; Rodoni et al., 1999; Kiranmai et al., 2005; Ha et al., 2008; Selvarajan et al., 2010a) has been developed. Besides primers for amplification of all the six components of BBTV, differential primers to differentiate PIO and SEA isolates have also

been designed (Burns et al., 1995; Sharman et al., 2000; Mansoor et al., 2005; Stainton et al., 2012). Additionally, Loop mediated Isothermal Amplification (LAMP) (Peng et al., 2012), Rolling Circle Amplification (RCA) (Stainton et al., 2012) and Nucleic acid Spot Hybridization (NASH) (Harding et al., 1991; Xie and Hu, 1995; Hafner et al., 1997; Selvarajan and Balasubramanian, 2008) have been applied for the sensitive detection of BBTV.

RealTime PCR (qPCR) is used to detect and quantify BBTV (Bressan and Watanabe, 2011; Chen and Hu, 2013) as well as BBrMV (Siljo et al., 2014). Multiplex PCR to detect infection of more than one virus in banana in a single reaction has also been standardized in various laboratories (Sharman et al., 2000; Selvarajan et al., 2004; Liu et al., 2012). Furthermore, Immuno Capture- RT-PCR (IC-RT-PCR), combination of serological and nucleic acid methods is used to detect BBrMV infecting banana (Iskra-Caruana et al., 2008).

The World Health Organisation (WHO) has designated a perfect diagnostic technique to be "ASSURED" *ie.*, Affordable, Sensitive, Specific, User friendly, Rapid and robust, Equipment free, and Deliverable (Kosack et al., 2017). Although nucleic acid based assays qualify as rapid, specific and sensitive than serological assays, the requirement of skilled personnel and sophisticated infrastructure makes it unaffordable and complicated. However, immunodiagnosis can be perfected to fit in all the parameters of good diagnostic assay, provided, it is more specific and sensitive.

2.5 SERODIAGNOSTICS AND RECOMBINANT COAT PROTEIN

Sensitivity, specificity and reliability of any serodiagnostic assay depends on the quality of antiserum used for the same. In serological assays employed for detection of banana viruses, antisera were developed against the surface epitopes present on coat protein of the virus/ virion. Iskra- Caruana *et al.,* and Wu and Su developed and standardised method for purification of BBTV virions from BBTD-affected banana plants in 1989 and 1990, respectively. Wu and Su (1990b) purified BBTV from infected banana sample collected from Taiwan by differential centrifugation (Wu ad Su, 1990a) and used it for immunising BALB/c mice to develop monoclonal antibodies (MAbs) by hybridoma technology. Antibody trapped antigen (ATA) ELISA, PTA- ELISA and immunofluorescence assays were developed for the first time, with BBTV specific MAbs (Wu and Su, 1990b).

Later, more methods for BBTV purification were reported by Thomas and Dietzgen (1991) by modifying the protocol of Wu and Su (1990a). They used two step purification protocol in which virions were partially purified by sucrose density ultracentrifugation and later subjected to equilibrium centrifugation in caesium sulphate (Cs_2SO_4). These were analysed in Sodium Dodecyl Sulphate- PolyAcrylamide Gel Electrophoresis (SDS-PAGE) and prominent band at 21 kDa was observed corresponding to the molecular weight of BBTV CP. This process yielded purified preparations of the virus compared to other methods, nevertheless, it was five days long laborious process. Partially purified virions were injected twice in a New Zealand white rabbit intramuscularly, and antiserum was raised and used to develop ELISA (Thomas and Dietzgen, 1991).

In the case of banana viruses, major drawbacks observed while raising antiserum against partially purified virus is the interference of phenols and secondary metabolites present in the host and low concentration of final preparation. This notably compromises the specificity of the antisera as well as considerably reduces sensitivity of the assay (Hema et al., 2003). Development of recombinant coat protein (rCP) and using highly pure protein as antigen for immunising the animal resolves the above challenges.

Overexpression of recombinant proteins in bacterial systems and its purification in large quantity have revolutionised biochemistry. It has created a surge in number of scientific studies on biochemical and physical characterisation of the proteins which were otherwise difficult to characterise by X-ray diffraction or Nuclear Magnetic Resonance (NMR) spectroscopy (Rosano and Ceccarelli, 2014). Development of many serodiagnostic assays against plant viruses have been developed in the last few years using recombinant technology.

Theoretically, development of recombinant protein can be carried out by the following steps. The gene of interest is cloned in to a suitable expression vector which is later transformed to appropriate bacterial host. The protein is induced and overexpressed in this host and then purified and characterised (Sørensen and Mortensen, 2005a). However, in practice, poor growth of the host, inclusion body (IB) formation and even not obtaining any protein at all are some of the problems often encountered and thus standardisation of each step is inevitable. Nevertheless, deciding expression vector and host combination is the most significant step (Sørensen and Mortensen, 2005b; Rosano and Ceccarelli, 2014). Currently, strains of *Escherichia coli* (*E. coli*) are most preferred as expression hosts owing

to its fast growth in cheap and readily available media and ease in transforming exogenous DNA (Rosano and Ceccarelli, 2014). Rosetta and BL21 (DE3) strains of *E. coli* with pLysS plasmid are the commonly used expression hosts. DE3 is an arrangement of T7 RNA Polymerase gene, under the control of LacUV 5 promotor on a phage genome whereas, pLysS is a plasmid that encodes T7 lysozyme that minimize leaky expression of the protein (Studier and Moffatt, 1986; Dubendorff and Studier, 1991).

Wanitchakorn *et al.,* (1997) cloned BBTV CP to pMAL-c2 vector and expressed in BL21 (DE3) strain of *E. coli*. The fusion protein with maltose binding domain (MBD) was used to raise antiserum in New Zealand white rabbit. However, apart from a specific band corresponding to BBTV CP, non-specific bands were observed in both healthy and infected samples in Western immunoblot with the antiserum produced against the fusion protein. The observation that maltose binding protein (MBP) alone reacted with the antiserum is indicative of the presence of MBP-related proteins in banana tissue with similar epitopes that contributed to formation of polyclonal antibodies. The solution proposed was to use cleaved BBTV CP and not the fusion protein for antiserum development.

Later, overexpression of BBTV rCP was attempted in other expression vectors like pQE30 and pET28a (+) (Abdelkader *et al.,* 2004; Shilpa *et al.,* 2016; Arumugan *et al.,* 2017). In pQE30 expression vector, overexpressed BBTV rCP was soluble. Nevertheless, high molecular weight non-specific proteins were present along with purified rCP. The SDS-PAGE profile displayed prominent band at 21 kDa corresponding to coat protein of BBTV (Abdelkader et al., 2004; Shilpa *et al.,* 2016). In pET28a (+), however, the overexpression of BBTV rCP was less and its solubility was not tested in the study (Arumugam et al., 2017).

Likewise, rCP protein production of BBrMV was attempted by Rodoni *et al.* (1999). The entire coding region of BBrMV CP was cloned into pProEX-1 expression vector and transformed to *E. coli* M15 cells containing the pREP4 repressor plasmid. The 39 kDa purified fusion protein was characterised by Western blot and used to raise antiserum in New Zealand white rabbit. Conversely, BBrMV CP gene cloned to pET28a (+) and overexpressed in *E. coli* BL21 (DE3) was insoluble and present in pellet fraction after sonication. The pellet was subjected to urea denaturation and later purified using Ni- NTA column chromatography (Selvarajan et al., 2020b).

Correspondingly, rCP CMV and *Banana streak mysore virus* (BSMyV) causing infectious chlorosis and streak disease respectively in banana have been developed for detection of the viruses. The rCP has been overexpressed in a series of pET vectors like pET21a, pET21d, pET28a and pMAL-c2X (Rostami *et al.,* 2014; Sharma et al., 2014; Kim *et al.,* 2016; Koolivand *et al.,* 2017). Antiserum raised against the rCP was evaluated for its efficacy in detecting the virus in field samples. Recombinant Viral-associated protein (VAP) of BSV was expressed in *E. coli* and polyclonal antibodies were raised against it in rabbits. The antibody was found to be effective in detecting BSV in assays like Western blot, Immunosorbent electron microscopy (ISEM) and ELISA (Selvarajan *et al.,* 2016).

Monoclonal antibodies and polyclonal antibodies for *Grapevine leafroll associated closterovirus-3* (GLRaV-3) and *Grapevine fanleaf virus* (GFLV) respectively were developed against purified viral rCPs. The CP genes of GLRaV-3 and GFLV were cloned in to pRSET-C and pET28a (+) respectively and was overexpressed in *E. coli* BL21 (DE3) and effectively used in serological assays to detect the viruses in grapevine (Ling et al., 2000; Shibaei *et al.,* 2018).

Furthermore, rCPs were developed for rapid, economical, easy-to-use, serological assays for the detection of *Citrus tristeza virus* (CTV) (Sekiya et al., 1991), *Sugarcane mosaic virus* (SCMV) (Smith et al., 1995), *Tomato spotted wilt virus* (TSWV) (Vaira et al., 1996), *Potato virus A* (PVA) (Čeřovská et al., 2002), *Potato mop top virus* (PMTV) (Čeřovská et al 2003), *Sugarcane streak mosaic virus* (SCSMV) (Hema et al., 2003; Hamdayanty et al., 2016), *Prune dwarf virus* (PDV) (Jawdah *et al.,* 2004); *Potato virus Y* (PVY) (Folwarczna et al., 2008), *Potato virus X* (PVX) (Cerovska et al., 2010), *Papaya ring spot virus* (PRSV) (Valekunja *et al.,* 2016), *Citrus psorosis virus* (CPV) (Salem et al., 2018), and *Citrus yellow mosaic virus* (CYMV) (Kumar et al., 2018) in field samples. The rCPs were cloned and overexpressed in various vector/host combinations.

2.6 PURIFICATION OF rCP

Recombinant proteins are often expressed as fusion protein with a tag on the N and or C terminal of the target protein. These tags are often linked to the target protein by a short linker sequence containing a specific chemical (*viz.,* Met or Asp-Pro) or protease cleavage sites (*viz.*, thrombin, enterokinase). Expression of fusion proteins facilitate the recovery of high quality pure recombinant protein. Hexa-histidine tag (His-tag), glutathione-S-transferase (GST-tag), and thioredoxin (Trx-tag) are few of the popular fusion partners

(Lichty et al., 2005; Waugh, 2005; Esposito and Chatterjee, 2006). Tag occasionally aids in increasing solubility of the protein which is otherwise insoluble (LaVallie et al., 1993; Zhang et al., 1998). Alternatively, the tags may also promote insolubility, especially while expressing short, partially structured polypeptides which need to be protected from host proteases and for expressing proteins that are normally toxic to the host in which it is expressed (Costa et al., 2014).

Purification of the proteins depends on various physico-chemical properties of the proteins most important being the isoelectric point. The expression of fusion proteins with affinity tags, has become extremely popular due to the ease of protein purification under both nondenaturing and denaturing solvent conditions using affinity column chromatography (Wingfield, 2016). Affinity chromatography, selects target proteins from other host proteins by binding with immobilised lectins, dyes, nucleic acids or heparin through its tags (Maity et al., 2013).

Most of the recombinant coat proteins expressed in *E. coli* vectors, reviewed in the previous section are purified using affinity column chromatography. The immobilised matrix is selected according to the tag linked to the target protein (Rodoni et al., 1999; Ling et al., 2000; Hema et al., 2003; Abdelkader et al., 2004; Khan et al., 2011; Rostami et al., 2014; Sharma et al., 2014; Wang et al., 2014; Hamdayanty et al., 2016; Arumugam et al., 2017; Sabharwal et al., 2020).

In case of insoluble protein, the pellets obtained after slow centrifugation containing the protein are treated with denaturants such as 6 to 8 M guanidine hydrochloride or urea. The protein is then purified in the presence of the denaturing agent by affinity column chromatography and later renatured by removing the denaturant through repeated dialysis (Bennion and Daggett, 2003; Camilloni et al., 2008; Almarza et al., 2009; Koolivand et al., 2016; Salem et al., 2018; Selvarajan et al., 2020b).

Apart from affinity column chromatography, differential centrifugation and density gradient ultracentrifugation are also used for rCP purification if the coat protein can self-assemble into virus like particles (VLPs) (Yusibov et al., 1996; Savithri and Murthy, 2010; Gulati et al., 2016a; Sabharwal et al., 2019, 2020). A few instances where rCPs of plant viruses purified by conventional protein purification methods such as ion-exchange and gel filtration chromatography are also reported (Walls and Loughran, 2011).

2.7 BIOPHYSICAL CHARACTERISATIONS OF VIRAL PROTEINS

Proteins that can be crystallised are characterised by X-ray diffraction or NMR spectroscopy. However, many proteins, its truncated versions, domains are unstable and insoluble or partially soluble due to many reasons. Thus, biophysical and biochemical characterisation of proteins opens new avenues of understanding the secondary structure of the protein and the implications of structure on its multifunctionality through dynamic protein-protein or protein-nucleotide interactions. These data are correlated to detailed molecular protein structures from crystallography and high resolution NMR (Neet and Lee, 2002; Vedadi et al., 2010). Biophysical methods include hydrodynamic methods like analytical ultracentrifugation and viscometry, thermodynamic methods like light scattering, microcalorimetry and surface plasma resonance and spectrometry like fluorescence spectroscopy, circular dichroism (CD) and NMR for the characterization of the proteins (Neet and Lee, 2002).

Hitherto, very few plant viral CPs have been characterised biophysically. Fluorescence spectroscopy of Alfalfa mosaic virus (AMV) CP isolated from tobacco was carried out by Kan et al. (1986). Emission spectra was calculated and maximum emission exhibited at 330 nm, a typical fluorescence of tryptophan in a hydrophobic environment and indicative of a well folded protein.

The structure of PVX CP was interpreted to be composed of α-helices by CD spectroscopy. It was evident from the study that the tertiary structure of PVX CP subunits in solution and that of the intact virus particles are similar (Homer and Goodman, 1975).

The secondary structure of PVA was determined by characterisation of virion purified from *Nicotiana benthamiana*. Detailed study of physico-chemical characters through UV absorption, far and near UV CD spectra, intrinsic fluorescence spectra and differential scanning calorimetry (DSC) melting curves, transmission electron microscopy, mass spectroscopy and synchrotron small angle X-ray scattering (SAXS) were conducted (Ksenofontov et al., 2013, 2018). The fluorescence spectra and Far-UV CD spectra showed highly disordered regions in the protein probably due to the surface exposed N and C terminal. The α-helical content in the CP was calculated to be in the range of 25 to 45 per cent and that of β structures were predicted to be 16 %.

Characterisation of VPg, NIa-Pro encoded by *Pepper vein banding potyvirus* (PVBV) was carried out by Far-UV spectroscopy and analytical ultracentrifugation. The protein coding

genes were cloned and expressed in appropriate vectors and strains of *E. coli* respectively to yield high concentration of good quality proteins for characterisation (Sabharwal, 2017).

The intrinsic and extrinsic fluorescence of rCP of *Bean common mosaic virus* (BCMV) studied by Kumar et al. (2019) revealed that N and C terminal of the protein is disordered which corroborated previous research on CP of various potyviruses. The Far UV CD spectra of the protein discovered presence of large unstructured region in the polypeptide which refers to N and C terminal of the CP.

2.8 APPLICATIONS OF rCP

Recombinant technology for investigations of structural and non-structural proteins of plant viruses was indeed a breakthrough. Possibility of obtaining soluble and crystallizable proteins have increased with the popularisation of recombinant protein development techniques using various combinations of expression vectors and hosts. It is also possible to manipulate the protein by expression of certain domains and truncated proteins which are crystallised using this method (Lesley and Wilson, 2005; Vedadi et al., 2007).

The bioactivity of AMV CP was studied by protoplast infectivity assay using the rCP and was proved to be biologically active in early event of infection. The AMV CP gene was cloned into pTrcHisB and expressed in *E. coli*. Inoculation of genomic RNAs alone did not cause infection instead inoculation of rCP with RNA resulted in infection (Yusibov et al., 1996). It implies that the RNA binds to the rCP to form infective virion. The rCP subunits also self assembled *in vitro* even in the absence of RNA to form more or less equal sized isometric particles, however, bacilliform particles like that of AMV virions were formed only in the presence of RNA. It was then subjected to X-ray diffraction after crystallisation.

Jacob and Usha (2002) overexpressed Cardamom mosaic virus (CdMV) CP in *E. coli* to analyse self assembly of CP subunits and observed filamentous aggregates without any association with genomic RNA. After crystallisation of *Tobacco mosaic virus* (TMV) using ammonium sulphate by Stanley in 1935, Li et al. in 2012, attempted crystallisation of TMV rCP expressed in *E. coli*. The protein, was slightly modified at the N and C terminal to obtain more stability in crystallisation buffers and was characterised by size exclusion chromatography as well as gel electrophoresis. The high-resolution crystals were subjected to X-ray diffraction which gave a similar biophysical profile of that of TMV virion. The

authors reported crystallisation of modified rCP to be unsophisticated compared to the process of obtaining crystals from TMV particles.

Gulati et al., (2016a) could solve the structure of Tobacco streak virus (TSV) CP by crystallisation of VLPs formed *in vitro* from rCP. The VLPs were allowed to form *in vitro* and were purified by sucrose density gradient ultracentrifugation which were later crystallised. The architecture of Ilarvirus particles, reasons for their capsid polymorphism and multitasking ability have been decoded with the availability of structural data. TSV CP gene was cloned in to pRSET- C and transformed and expressed in *E. coli* (DE3) BL21.

Apart from using recombinant proteins as an antigen for developing antibodies and aiding in solving their structure and it's functional implications, it has many more advantages of which most are unexplored. The VLPs, coat proteins that self-assemble are nano-sized, molecules that bear a resemblance to viruses but non-infectious due to absence of viral genome (Roldão et al., 2010). The immunogenicity of VLPs enable it's use as vaccines, as they elicit an immune response from the B cells (Chen and Lai, 2013). Their organised repetitive structure is highly preferred to be used as scaffolds for bioimaging, as biosensors, in light-harvesting, synthesis of bionanomaterials, and as nanocarriers in drug and gene therapy as a cheaper alternative to gold nanoparticles which are extensively used at present (Domingo-Espín et al., 2011).

Efforts were made to form chimeric VLPs of *Sesbania mosaic virus* (SeMV) and to crystallise it (Gulati et al., 2016b). The disordered N terminal was replaced by B domain of *Staphylococcus aureus* protein A, which binds to the Fc region of the antibodies. The assembly of T =1 chimeric VLPs instead of T =3 was observed. However, the bioactivity of B domain to was intact and suggested possible biotechnological applications.

Likewise, chimeric PVBV VLPs were produced and it's application as immunodiagnostic and immunotherapeutic agents were explored by Sabharwal et al. (2020). The chimeric CP was expressed in *E. coli* as recombinant protein and purified using nickel-nitrilotriacetic acid (Ni^{2+}-NTA) affinity column chromatography and sucrose density gradient. The chimeric VLPs could internalise into mammalian cells (Sabharwal et al., 2019) and bind to antibodies through the B domain and deliver them was proved in the study (Sabharwal et al., 2020). The inability of antibodies to cross the cell membrane and thus limiting its use in immunotherapy can be resolved this way (Shirbaghaee and Bolhassani, 2016).

Materials and Methods

3. MATERIAL AND METHODS

3.1 CHEMICALS

General biochemicals and accessories such as nickel-nitrilotriacetic acid (Ni^{2+}-NTA) agarose beads, GSH (glutathione) Sepharose beads, isopropyl-β-D-thiogalactopyranoside (IPTG), X-gal, Nitrocellulose membrane (NCM), electrochemiluminescence (ECL) solution, antibiotics, gel elution kits used in this study were obtained from Sigma-Aldrich USA; Merck-Millipore, Germany; Calbiochem, USA; Novagen, USA; ThermoFisher Scientific, USA and Bio-Rad, USA. Ready to use Luria Bertani (LB) broth and other chemicals for preparing media like yeast extract, tryptone, Agar-Agar, were procured from Hi-Media, Mumbai, India. Molecular grade chemicals like agarose, acrylamide, bis acrylamide, tetra methyl ethylene diamine (TEMED), sodium dodecyl sulphate (SDS) and ammonium per sulphate (APS) were procured from Promega, USA and Sigma-Aldrich, USA. Restriction endonucleases, T4 DNA ligase, *Taq* polymerase and Phusion (*Pfu*) polymerase were obtained from ThermoFisher Scientific, USA; New England Biolabs Inc., USA; Genei, India, Meridian Bioscience, USA. Protease inhibitor cocktail were obtained from Sigma. Few other chemicals were of analytical grade, purchased from Hi-Media and some were purchased from local chemical companies.

Protein molecular mass markers besides 1 kb and 100 bp DNA ladder were procured from ThermoFisher Scientific, USA; Bio-Rad, USA. Goat anti-rabbit polyclonal IgG enzyme bioconjugate and Goat anti-mouse monoclonal IgG enzyme bioconjugate were procured from ThermoFisher Scientific, USA and Calbiochem, USA. The BBTV and BBrMV primary antibodies (commercially available) were obtained from National Research Centre for Banana (NRCB), Trichy and Pepper vein banding virus coat protein (PVBV CP) specific antiserum was procured from Prof. Savithri Lab, Indian Institute of Science (IISc), Bengaluru.

The ELISA plate, plasticware and glassware were bought from Tarsons Pvt. Ltd., India; Borosil Pvt. Ltd., India; Axygen Inc., USA and Cole-Parmer, USA. DNA sequencing was carried out by AgriGenome Labs, Kochi and Eurofins, Bengaluru. Formvar coated copper grids for electron microscopy were procured from SPI supplies, USA and electron micrographs were taken from Electron Microscopy facility, Division of Biological

Sciences, IISc, Bangalore. The sources of some infrequently used chemicals are mentioned in the appropriate sections of the thesis.

3.2. BACTERIAL HOSTS AND VECTORS

Escherichia coli DH5α (New England Biolabs, USA) was the recipient strain for the plasmids used in cloning. pGEM-T easy TA cloning vector (Promega) and pUC19 (Invitrogen) were used for molecular cloning of CP genes of the viruses. *E. coli* strains BL21 (DE3) and Rosetta (DE3) containing pLysS (Novagen, Germany), BL21 (DE3), C41 (kind gift from Dr. D. N. Rao Lab and Dr. Tanveer Hussain Lab, IISc) were used for overexpression of coat proteins in bacteria. pRSET-C, pGEX-4T-2 and pET32a (+) vectors (kind gift from Prof. Savithri Lab, IISc, Bengaluru) were used for cloning and overexpression of CPs. Tobacco etch virus (TEV) protease gene containing recombinant plasmid was a kind gift from Dr. Gayathri Pananghat, IISER, Pune.

3.3 SURVEY AND COLLECTION OF VIRUS ISOLATES

A roving survey in 10 districts, divided into North, Central and South zones of Kerala was conducted from 2017 October to 2018 February. The population subset in Northern zone included districts of Kozhikode, Wayanad, Kannur and Kasaragod. From the Central zone, Ernakulam, Thrissur and Palakkad districts were surveyed and Thiruvananthapuram, Kollam and Kottayam were surveyed from Southern part of Kerala. Incidence of BBTD and BBrMD were recorded during the survey. Disease incidence was calculated as follows

$$\text{Per cent disease incidence (PDI)} = \frac{\text{Number of plants infected}}{\text{Total number of plants observed}} \times 100$$

A questionnaire was prepared and details like variety cultivated, the stage of emergence of the disease, association with aphid vector besides personal details were collected (Appendix I). The survey was conducted with an objective to collect virus isolates from various parts of the state. Young leaves of infected samples were collected and stored after flash freezing in liquid nitrogen and stored at -80 ºC until further use. Suckers from BBrMV infected plants were collected and planted in the insect free net house at Banana Research Station, Kannara for further investigations. As necrosis occurs at a later stage of infection, suckers from BBTV infected plants were not maintained in the insect free net house.

3.4 PRELIMINARY ASSAY BY ENZYME LINKED IMMUNOSORBENT ASSAY

A total of 50 samples suspected to be infected with BBTV and BBrMV were collected and subjected to preliminary assay by Direct Antigen Coating- Enzyme Linked Immunosorbent Assay (DAC-ELISA) with slight variation in the standard protocol (Clark and Adams, 1977).

All buffers used for the experiment were prepared in autoclaved Milli-Q water. Coating buffer (carbonate-bicarbonate buffer, 10X) was prepared by dissolving 15.9 g sodium carbonate (Na_2CO_3), 29.3 g sodium bicarbonate ($NaHCO_3$), and 2.8 g sodium nitrate ($NaNO_3$) in 500 ml water which was made up to 1 L after adjusting pH to 9.2. Similarly, 1 L of 10X Phosphate buffered saline (PBS) of pH 7.4 was prepared by dissolving 80 g sodium chloride (NaCl), 2 g potassium dihydrogen phosphate (KH_2PO_4), 11.6 g disodium hydrogen phosphate (Na_2HPO_4) and 2g potassium chloride (KCl). PBS-T (1X PBS of pH 7.4 containing 0.1 per cent v/v Tween-20) was used as washing buffer. Antibody titre of commercially available BBTV and BBrMV specific primary antisera was determined through DAC-ELISA against same concentration of antigen. Primary and secondary antisera were diluted in antibody diluent buffer, PBS-TPO (PBS-T with 2 % Poly vinyl pyrrolidone; PVP-K30 and 0.2 % bovine serum albumin; BSA). Substrate buffer (100 ml) was prepared by dissolving 9.7 ml diethanolamine in 50 ml water, made up to final volume after adjusting pH to 9.8 using concentrated hydrochloric acid (HCl) and stored in an amber coloured bottle.

Antigen was isolated from 1 g of infected leaf sample in 5 ml carbonate buffer (1X) and 200 µl was coated on to the wells of 96 well polystyrene ELISA plate in duplicates (sometimes triplicates) for 1 h at 37 ºC. After washing off the excess antigen using PBS-T, 200 µl blocking buffer (5 % skim milk powder in PBS-T) was added and retained at 37 ºC for 45 min with the intention of reducing non-specific binding of the primary antibody. Appropriately diluted commercially available virus specific antiserum (200 µl) was added to the wells and incubated overnight at 4 ºC. The excess primary antibody was washed off with PBS-T thrice, and 200 µl of 1:10000 v/v anti-rabbit IgG conjugated with alkaline phosphatase enzyme (secondary antibody) was added to the wells and incubated for 2 h at 37 ºC. 200 µl of substrate solution (1000 ppm of para-nitro phenyl phosphate disodium salt; PNPP dissolved in substrate buffer) was added to the wells after washing off excess secondary antibody and incubated for 30 min in dark at room temperature. The absorbance

was recorded at 405 nm in ELISA plate reader (BioRad). Healthy samples as negative control and infected samples as positive control in duplicates were added in the assay. The test samples that showed higher than double the A_{405} value of negative controls were considered as positive.

3.5 TOTAL DNA ISOLATION

Total DNA was isolated from seventeen representative samples selected from those that gave positive results in the preliminary assay using modified CTAB method standardised in Virology lab, BRS, Kannara. In the presence of liquid nitrogen, 1g leaf sample with midrib was pulverised to fine powder and homogenised with 10 ml pre heated 2X extraction buffer (2 % CTAB; 100 mM Tris, pH 8.0; 1.4 M NaCl; 20 mM EDTA; 1 % PVP and 0.1 % β-mercaptoethanol, BME) in autoclaved and prechilled pestle and mortar. The mixture was transferred to an autoclaved Oakridge tube and incubated at 65 ºC in water bath providing intermitted mixing. Ten ml of 24:1 v/v chloroform and isoamyl alcohol was added to the Oakridge tube and placed in an end-to-end rotor for 10 min. Post centrifugation at 10,000 rpm for 15 min at 4 ºC, the aqueous phase was collected into another Oakridge tube. The process was repeated. To the aqueous phase collected post second round of centrifugation, equal volume of ice cold isopropanol was added. Subsequently, incubated it at -20 ºC for 1 h after thorough mixing. The precipitated DNA was collected by centrifugation at 10,000 rpm for 20 min at 4 ºC. The pellet was washed with 70 per cent ice cold ethanol and was blot dried. Pellet was air dried until the alcohol completely evaporated and dissolved in autoclaved Milli-Q water. DNA was quantified in Nanodrop (ThermoFischer Scientific). Quality was checked in 0.8 per cent agarose gel.

3.6 POLYMERASE CHAIN REACTION

Total DNA isolated from the infected tissue was used as template for Polymerase Chain Reaction (PCR) to amplify CP gene of BBTV using reported forward and reverse primers (Wanitchakorn et al., 2000a). The PCR reaction mixture contained 1X Taq buffer with 15 mM $MgCl_2$, 10 mM dNTPs, 40 pmol each of BBTV CP specific forward (5’ GGTTTTC GGATTGAGCCTAC 3’) and reverse (5’ TTGACGGTGTTTTCAGGAACC 3’) primers, 3 units (U) of Taq polymerase (Genei, Bangalore) and 25-30 ng of total DNA isolated was used as template. The PCR was performed in a Master Cycler PCR machine (Eppendorf, Germany) with initial denaturation at 94 ºC for 5 min followed by 35 cycles of denaturation

at 94 ºC for 1 min, annealing at 60.8 ºC for 1 min and extension at 72 ºC for 1 min 30 s with a final extension of 72 ºC for 10 min. The PCR product was analysed on 1.2 per cent agarose and purified by PCR purification kit according to manufacturer's instructions (ThermoFisher Scientific). The purified PCR products were sequenced by Sanger dideoxy sequencing method using virus CP specific forward and reverse primers. Partial BBTV DNA-S with the complete CP coding sequence of 17 BBTV isolates generated in this study were deposited in NCBI GenBank and accession numbers MT174314- MT174330 have been assigned.

3.7 TOTAL RNA ISOLATION

All the glassware, micropipette tips, mortar, pestle, microcentrifuge tubes, PCR tubes and water for preparing buffers were treated with 0.1 per cent diethyl pyrocarbonate (DEPC) overnight and double sterilised. Total RNA was isolated from twelve representative samples that gave positive results in DAC-ELISA using RNeasy Plant MiniKit (Qiagen, USA) according to manufacturer's instructions. Hundred mg of leaf samples were disrupted in presence of liquid nitrogen in mortar and pestle. RLT buffer (450 µl) was added to the pulverised tissue after the liquid nitrogen evaporated. The homogenised lysate was transferred to QIAShredder spin column placed in a collection tube. Supernatant was collected after centrifugation at 10,000 rpm for 2 min at 4 ºC. Ethanol (100 %) equal to half the volume of supernatant was added to clear the lysate and nucleic acids were allowed to precipitate. The lysate along with precipitate was transferred to RNeasy MiniSpin column placed in a 2 ml collection tube. RNA was allowed to bind to the column and the rest were separated by centrifugation for 15 s at 10,000 rpm. Flow through was discarded and the column was washed with 700 µl RW1 buffer. Subsequently, the column was washed with 500 µl RPE buffer twice. Excess wash buffer was discarded by centrifugation at 10,000 rpm for 20 s and alcohol was allowed to evaporate at room temperature. RNA was eluted with RNase free water added directly to the membrane. Centrifugation at 10,000 rpm for 1 min was repeated for maximum recovery of RNA. Quality of RNA was assessed by agarose gel electrophoresis (AGE) and quantified in Nanodrop.

3.8 REVERSE TRANSCRIPTASE- POLYMERASE CHAIN REACTION

First strand cDNA was synthesised from total RNA isolated from the BBrMV infected plants using RevertAid First strand cDNA synthesis kit (ThermoFisher Scientific). Twenty

µl of reaction mixture was prepared by adding 4 µl of 5X reaction buffer, 10 mM dNTP mix, 1 µl Ribolock RNAse inhibitor, 200 units of RevertAid M-MuLV RT, 5 µM Oligo $(dT)_{18}$ as the primer and 50-100 ng of template. The mix was incubated at 42 ºC for 1 h for cDNA synthesis to take place. The reaction was terminated by incubating the mix at 70 ºC for 5 min. The PCR reaction mixture contained 1X Taq buffer with 15 mM $MgCl_2$, 10 mM dNTPs, 40 pmol each of BBrMV CP specific forward (5' TCTGGAACGGAGTCAACC AA 3') and reverse (5' CTCGATCAATACCTCACAGG 3') primers (Sankaralingam et al., 2006), 3 U of *Taq* polymerase (Genei, Bangalore). One µL of the first strand cDNA synthesis reaction mixture was used as template. PCR was performed with initial denaturation at 94 ºC for 5 min followed by 35 cycles of denaturation at 94 ºC for 1 min, annealing at 42 ºC for 1 min and extension at 72 ºC for 2 min with a final extension of 72 ºC for 10 min. The PCR product was analysed on 1.2 per cent agarose and purified. The purified PCR product was sequenced by Sanger dideoxy sequencing method in both directions using BBrMV CP specific forward and reverse primers to obtain complete sequence. The complete CP gene sequences with 3' UTR of 12 BBrMV isolates generated in this study were deposited in NCBI GenBank and were assigned accession numbers MT818176-MT818187.

3.9 AGAROSE GEL ELECTROPHORESIS

Molecular grade agarose was weighed according to the concentration of the gel to be prepared and dissolved in 1X TAE buffer (prepared from 10X TAE buffer containing 2 M Tris, pH 8.0; 0.5 M ethylene diamine tetra acetic acid (EDTA), pH 8.0; 1 M glacial acetic acid) by boiling. When the solution cooled down to ~ 40 ºC, 2 µl of ethidium bromide (EtBr) was added and casted. 1X TAE buffer was poured to electrophoresis tank and the gel was placed in it. After loading samples mixed with 6X DNA loading dye (0.25 % bromophenol blue, 0.25 % xylene cyanol, 30 % glycerol), electrophoresis was carried out at 80 V. In order to run RNA samples, TAE buffer was prepared in DEPC treated double autoclaved water. The casting tray, comb and electrophoresis tank were also washed thoroughly with DEPC treated water prior to casting the gel and electrophoresis.

3.10 IN SILICO ANALYSIS OF COAT PROTEIN GENE

The PCR products (BBTV CP and BBrMV CP) after purification was sequenced by Sanger dideoxy sequencing method in both directions using forward and reverse primers. All the

raw sequences obtained after sequencing, were analysed for its quality score and trimmed accordingly. Forward and reverse sequences were merged to form contig in CAP3-Prabi online tool.

In order to assess genetic variability and phylogeny of the CP genes of BBTV and BBrMV, apart from the sequences generated in this study, complete coding sequences of CP genes of isolates from different parts of the world were retrieved from GenBank database (http://www.ncbi.nlm.nih.gov/) (Appendix IV). In order to make meaningful comparisons, only full length gene sequences were considered and partial sequences were discarded. Coat protein gene sequences of BBTV isolates belonging to both SEA and PIO groups were included in BBTV CP dataset. BBrMV infecting ornamental ginger and cardamom were included in the dataset of BBrMV CP gene. The amino acid sequences were derived from the nucleotide sequences using ExPASy Translate tool (Gasteiger et al., 2003). Rare codons in the protein were determined by Rare codon search tool (https://www.bioline.com/media/calculator/01_11.html).

3.10.1 Phylogenetic analysis

Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018). The coding region of BBTV CP were selected and aligned by MUSCLE (MUltiple Sequence Comparison by Log-Expectation). Phylogenetic tree was constructed by Neighbour Joining (NJ) method (Saitou and Nei, 1987) and evolutionary distances were computed using the Maximum Composite Likelihood (MCL) method (Tamura et al., 2004) and were in the units of the number of base substitutions per site. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) were shown next to the branches (Felsenstein, 1985). The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were total of 528 positions in the final dataset. Separate tree to measure phylogeny of Kerala isolates was constructed by NJ method and evolutionary distances were computed using Kimura 2-parameter method (Kimura, 1980).

Similarly, coding region of BBrMV CP sequences were aligned by MUSCLE and phylogeny were deciphered by Maximum Likelihood method (Felsenstein, 1981). Evolutionary distances were computed using the Kimura 2-parameter method and

represented in the units of the number of base substitutions per site. Bootstrap test with 1000 replications for signifying the confidence of the branches was conducted. Neighbour-Join and BioNJ algorithms were applied to a matrix of pairwise distances estimated using the MCL approach to obtain initial tree(s) for the exploratory analysis (Saitou and Nei, 1987). Later, topology was selected with superior log likelihood value. Phylogenetic tree was constructed with coat protein gene sequences of all the isolates and that of Kerala isolates separately.

3.10.2 Principle co-ordinate analysis

Principle coordinate analysis (PCoA) was carried out to represent the genetic relatedness among different BBTV and BBrMV isolates based on their CP genes. The 2-dimensional (2-D) PCoA based on a pairwise dissimilarity matrix for both datasets were performed in DARwin v6.0.021 separately (Perrier and Jacquemoud-Collet, 2006). Dissimilarity index was calculated by Tajima-Nei method with bootstrap value 500 (Perrier et al., 2003). The aim of factorial methods was to give an overall representation of diversity, dependence on geography and also host in case of BBrMV.

3.10.3 Assessment of genetic variability of BBTV based on coat protein gene

3.10.3.1 Assessment of sequence homology

Total number of 65 BBTV and 56 BBrMV CP sequences were aligned by MUSCLE (Edgar, 2004) and the pairwise nucleotide sequence identity scores were represented as sequence identity matrix and color-coded blocks using SDT v 1.2 software (Muhire et al., 2014). Parameters like theoretical isoelectric point (pI), theoretical molecular weight, extinction coefficient and stability in bacterial system of translated protein were determined by ProtParam online tool. These amino acid sequences were also aligned by MUSCLE and homology matrix was constructed.

3.10.3.2 Genetic variability analysis

Genetic variability analysis of BBTV and BBrMV isolates based on their coat protein gene was carried out. The sequences aligned by MUSCLE were used to detect DNA polymorphism in DnaSP v.6.12.01. Nucleotide diversity, Pi (π) and Watterson estimator or mutation rate (θ) were estimated based on the number of polymorphic sites. Pairwise rate of non-synonymous (Ka) and synonymous (Ks) substitutions per site of isolates were

calculated using method suggested by Nei and Gojobori (1986). The values with $P < 0.05$ were considered significant at 5 per cent level. Ratio of synonymous and non-synonymous substitutions of sequence pairs in the dataset (Ka/Ks also denoted as d_N/d_S or ω) were also calculated using the same tool.

Statistical analysis *viz.*, Tajima D test, Fu and Li D and F tests to evaluate significance of molecular diversity among population subsets were carried out based on the total number of mutation events in the population (Tajima, 1989; Fu and Li, 1993). In case of BBrMV, apart from geography, host range was considered as a different subset.

3.10.3.3 Gene flow between population subsets

During the roving survey, the isolates were collected from different parts of Kerala (North, South and Central zones). DNA divergence and gene flow between these population subsets were analysed using DnaSP software, version 6.12.01. DNA divergence between Kerala isolates which were retrieved from NCBI GenBank database and generated during this study was also analysed in detail to understand the genetic variability occurring with time. Apart from geography, hosts of BBrMV were also considered different population subsets.

3.10.4 Codon usage bias

The Codon Usage Bias (CUB) aids in determining the balance of selection and mutational pressure acting on a gene (Sharp and Li, 1986; Bulmer, 1988). Effective number of codons (Nc) (Wright, 1990) and Relative Synonymous Codon Usage (RSCU) of both BBTV and BBrMV CP genes were calculated using E-CAI web server as indicators of CUB (Puigbò et al., 2008), with the intention of comprehending selective forces acting upon the genes. The RSCU threshold was fixed as >1.30 and synonymous codons with values above threshold were referred to as a high frequency codon (Hfc) in the present investigation. To understand the host adaptability of BBrMV isolates infecting banana, flowering ginger and cardamom, Codon Adaptation Index (CAI) (Sharp and Li, 1987) was calculated.

The Nc and neutrality plots illustrated the role of selection pressure and mutation in CUB of CP gene. The effective number of codons (Nc) was plotted in the Y axis and G+C value at the wobble position (GC3) was plotted in the X axis to obtain Nc plot. The average of GC value at 1st and 2nd position of the codon (GC12) was plotted against GC3 in the X axis to obtain neutrality plot (Biswas et al., 2019).

3.10.5 Recombination analysis

Number of potential recombination events within the dataset was investigated by examining the aligned sequences using the MAXIMUM CHI (Maynard-Smith, 1992), LARD (Holmes et al., 1999), GENECONV (Padidam et al., 1999), Recombination Detection Program (RDP) (Martin and Rybicki, 2000), SISCAN (Gibbs et al., 2000), CHIMAERA (Posada and Crandall, 2001) and BOOTSCAN (Martin et al., 2005) recombination detection methods in RDP v 5.5. Potential recombinants and the recombinant breakpoint positions of these isolates were inferred. Matrices were constructed in a window size of 100 setting step size as 50 and analysed.

3.11 PRIMER DESIGNING AND STANDARDISATION OF ANNEALING TEMPERATURE

The aim of this experiment was to design primers for nucleic acid based detection of BBTV and BBrMV Kerala isolates. Primers specific targeting CP genes were designed, as it is the most abundant gene in all the stages of infection apart from replicase gene. Sequences generated from this study were aligned in CLUSTAL-omega. The most conserved region from the DNA-S of the BBTV was selected as forward primer. In case of BBrMV, forward primer was selected from conserved region of NIb region upstream to CP region in the ssRNA genome. The conserved sequence from the 3' UTR of ssRNA genome was selected and reverse compliment was taken as reverse primer (Table 1). The theoretical annealing temperature was calculated from the theoretical melting temperature (Tm) by substituting number of adenine (A), thymine (T), guanine (G) and cytosine (C) in the formula 2 (A+T) + 4(G+C). GC content of the primers were calculated individually.

Table 1. Details of designed primers to detect BBTV and BBrMV infection in field samples

Name of the primer	Sequence	Amplicon size
BBTV DNA-S FP	5'GATTGCTTGCCCTGCAAGCC 3'	890 bp
BBTV DNA-S RP	5' TCCATTAGCCCATGTCCCGTCC 3'	
BBrMV NIb-CP FP	5' ATATGTGGCTTAGTGTACAGG 3'	1306 bp
BBrMV NIb-CP RP	5' AAGGTGTGCACGCCAGCTAGA 3'	

The annealing temperature was determined by gradient PCR. The PCR mix was prepared as described in 3.6. and 3.8. For amplifying BBTV CP gene using designed primers, denaturation at 94 ºC for 1 min followed by gradient of annealing temperatures between 56-71.4 ºC for 1 min 15 s and extension at 72 ºC for 55 s was programmed. To amplify NIb-BBrMV CP, PCR programme with initial denaturation at 94 ºC for 5 min followed by 35 cycles of denaturation at 94 ºC for 1 min, annealing temperature was set between 51.4-64.7 ºC for 1 min and extension at 72 ºC for 1 min 20 s with a final extension of 72 ºC for 10 min.

3.12 NUCLEIC-ACID BASED DETECTION OF FIELD SAMPLES WITH DESIGNED PRIMERS SPECIFIC TO CP GENE

Total DNA and total RNA from BBTV and BBrMV infected samples respectively collected from germplasm field bank, experimental plots and net house of BRS, Kannara and various fields during survey, were isolated as mentioned in the section 3.5 and 3.7. Presence of BBTV and BBrMV in the infected samples were detected by PCR and RT-PCR respectively using the designed primers. The PCR cocktail was prepared and PCR amplifications were carried out according to the conditions standardised.

3.13 MOLECULAR CLONING OF BBTV AND BBrMV CP

3.13.1 Primer designing for cloning of CP gene to expression vectors

Primers were designed manually for cloning of BBTV CP and BBrMV CP gene to various expression vectors. The complete coding region of the CP gene were identified by ExpASy Translate tool using reference isolates MT174322 for BBTV and MT818184 for BBrMV. Forward primers were designed from the 5’ end of the ORF. Reverse compliment of the sequences from the 3’ end of the gene was selected as the reverse primer. Appropriate restriction enzyme site in the vector sequence were selected so as to avoid many extra amino acids in the N terminal of the recombinant protein which might vary its properties. It was also evaluated for any frame shift while translation of the protein due to cloning in the sites selected. These restriction enzyme recognition site(s) were inserted at the 5’ end of the forward primer and reverse primer to aid in cloning to multiple vectors (underlined) (Table 2).

Table 2. Details of designed primers for cloning BBTV and BBrMV CP gene

Name of the primer	Sequence	Amplicon length
BBTV CP sen	5' GGGGAATTCGCTAGCATGGCTAGGTATCCG 3'	530 bp
BBTV CP anti	5' CCCGGATCCTCAAACATGATATG 3'	
BBrMV CP sen	5' GGGGCTAGCTCTGGAACGGAGTCAACC 3'	1080 bp
BBrMV CP anti	5' CCCGGATCCTTACTCGATCAATACCTCAC 3'	

3.13.2 Competent cell preparation

The *E. coli* strains used in this study were made competent for the uptake of DNA (vectors and recombinant plasmids) using $MnCl_2$ and $CaCl_2$ as described by Alexander (1987). Acid salt buffer (ASB) containing 40 mM sodium acetate, 100 mM $CaCl_2$, 70 mM $MnCl_2$ was made in autoclaved Milli-Q and pH was adjusted to 5.5 using concentrated acetic acid. The buffer was sterilised using 0.22 μm membrane filter (Merck, USA) under aseptic condition and pre-chilled before use.

A single colony of *E. coli* strain was inoculated in 50 ml of sterile 2X LB medium (4 g casein and 2 g yeast extract, supplemented with 34 μg/ml chloramphenicol for maintaining plysS plasmid in Rosetta and BL21(DE3) cells) in 250 ml Erlenmeyer flask and incubated in a shaker at 37 °C overnight. Two ml of this primary inoculum was added to sterile 200 ml 2X LB medium (in 1 L Erlenmeyer flask) supplemented with appropriate antibiotic and grown in an incubator shaker at 30 °C till the A_{600} reached 0.4. The growth of the culture was arrested by chilling for 1 h at 4 °C and the cells were harvested by centrifugation at 3,000 rpm for 15 min at 4 °C. The supernatant was discarded and the cell pellet was resuspended in 30 ml of ASB and then kept in ice for 45 min. The ASB treated cells were centrifuged again at 3,000 rpm for 15 min at 4 °C and pellet was resuspended in 5 ml of ASB containing 20 per cent glycerol and stored as aliquots of 100 μl at -80 °C after flash freezing in liquid nitrogen for future use.

3.13.3 Restriction digestion and ligation

3.13.3.1 BBTV CP

The CP gene of BBTV was amplified using designed primers with EcoR1 and Nhe1 restriction enzyme recognition sites in the forward primer and BamH1 site in the reverse

primer (BBTV CP sen and anti; Table 2) by High Fidelity Pfu polymerase for cloning. The PCR reaction mixture contained 1X HF buffer, 10 mM dNTPs, 10 pmol each of BBTV CP specific forward and reverse primers, 6 per cent dimethyl sulfoxide (DMSO), 2 U of Pfu polymerase and 50 ng of template. PCR to amplify BBTV CP was performed in Thermal cycler (Bio-Rad, USA). Annealing temperature of 38.8 ºC was used for 10 cycles of amplification and in subsequent 20 cycles annealing temperature used was 55 ºC so as to circumvent non-specific binding at low annealing temperatures (Table 3). The PCR products were analysed on 1.2 % agarose gel and purified by PCR purification kit according to the manufacturer's instructions. Purified PCR product was double digested with 0.5 U of EcoR1 and 0.5 U BamH1 site in presence of 1X red buffer for 4 h and ligated with double digested pUC19 cloning vector.

Table 3. PCR conditions to amplify BBTV CP gene using designed primers for cloning

PCR conditions	**BBTV CP specific primers**
Initial denaturation	95 ºC for 5 min
10 cycles of	
Denaturation	95 ºC for 1 min
Annealing	38.8 ºC 1 min
Extension	72 ºC for 45 s
Followed by 20 cycles of	
Denaturation	95 ºC for 5 min
Annealing	55 ºC 1 min
Extension	72 ºC for 45 s
Final extension	72 ºC for 10 min

Cloning of BBTV CP and BBrMV CP genes to various expression vectors (pRSET-C, pGEX-4T-2 and pET32a (+)) were carried out to determine the appropriate vector/ host combination to obtain maximum yield of the recombinant proteins. The BBTV CP gene was cloned to Nhe1 and BamH1 sites of pRSET-C, Sma1 site of pGEX-4T-2 and EcoRV and BamH1 sites of pET32a (+). BBTV CP gene to be cloned into double digested pET32a (+) was digested with 5 U BamH1 alone and was not digested at all before ligating to Sma1 digested pGEX-4T-2. For double digestion appropriate buffers were selected in which both enzymes showed maximum activity.

A molar ratio of 1:10 vector to insert was followed for ligation with 40 U of T4 DNA ligase in the presence if 1X ligation buffer. Molar ratio (1:10) was calculated using the following formula:

$$\frac{\text{ng vector x size of insert} \times\ 10}{\text{size of the vector} \times\ 1} = ng\ of\ insert\ required$$

The ligation mix was incubated at 16 ºC overnight for maximum efficiency of ligase, for both blunt end and staggered end cloning.

3.13.3.2 Restriction free cloning to pGEX-4T-2

Primers were designed for restriction free (RF) cloning (of BBTV CP to pGEX-4T-2 to eliminate thrombin and trypsin cleavage sites and introduce TEV protease cleavage site between the target protein and glutathione-S-transferase (GST) tag. Forward primer, F-RF1 (5'**CCATCCTCCAAAATCGGAT**GAGAACCTGTACTTCCAAGGTCATATGGCTA GTATCCGAAG 3') was designed with C terminal sequence of GST protein (bold) followed by TEV protease cleavage site (underlined) and few N terminal sequences of BBTV CP gene. Reverse primer, R-RF1 (5' TCAAACATGATATGTAATTCTGTTCT GG 3') consisted of reverse compliment of 3' terminus sequences of BBTV CP alone. Annealing temperature was standardised by gradient PCR.

The primer pair F-RF1 and R-RF1 were used to amplify CP gene from total genomic DNA of BBTD infected banana. Reaction mix was prepared as mentioned earlier (section 3.12.2.1.) and PCR was performed with initial denaturation at 98 ºC for 1 min followed by 30 cycles of denaturation at 98 ºC for 30 s, annealing at 61.3 ºC for 30 s and extension at 72 ºC for 25 s with a final extension of 72 ºC for 10 min. The PCR product was separated on 0.8 per cent agarose gel and eluted using GenJet gel extraction kit (ThermoFischer Scientific) based on manufacturer's protocol.

The purified PCR product (PCR product 1) was used megaprimer for RF amplification using BBTV CP/ pGEX-4T-2 clone (mentioned in the previous section) as template. The PCR mix of volume 50 µl was prepared with 1X HF buffer, 10 mM dNTPs, 1 µg megaprimer, 6 % DMSO, 2.5 U of Pfu polymerase and 500 ng of template. Sterile Milli-Q water was added instead of megaprimers in a negative control. The RF reaction was as follows: a single denaturation step (95 ºC, 30 s) was performed followed by 35 cycles of:

denaturation (95 ºC, 30 s), annealing (55 ºC, 1 min), and elongation (68 ºC, 5 min) and a final elongation step of 7 min at 72 ºC.

3.13.3.3 BBrMV CP

For cloning BBrMV CP to pGEM-T easy vector, PCR was performed as described in section 3.8. PCR product amplified by Taq polymerase with an A overhang was purified by PCR purification kit according to manufacturer's instructions (Promega, USA). Four µl of purified PCR product was mixed with 1 µl of pGEM-T easy vector and 1 µl of T4 DNA ligase in 1X ligation buffer. Ligation mix was kept at room temperature for 1 h and later incubated at 4 ºC overnight. BBrMV CP gene was amplified by Pfu polymerase from pGEM-T clone, using BBrMV gene specific primers (BBrMV CP sen and anti; Table 2). BBrMV CP gene was amplified in thermal cycler (Bio-Rad) under the following conditions; initial denaturation at 98 ºC for 1 min followed by 30 cycles of denaturation at 98 ºC for 30 s, annealing at 55 ºC for 30 s, extension at 72 ºC for 45 s and a final extension at 72 ºC for 10 min.

Subsequently, double digested PCR product was cloned at the Nhe1 and BamH1 site of pRSET-C vector and undigested PCR product was cloned into Sma1 site of pGEX-4T-2. Molar ratio of 1:10 vector to insert was followed for ligation with T4 DNA ligase at 16 ºC overnight for both staggered and blunt end cloning.

3.13.4 Transformation of clones and selection

Transformation of recombinant plasmid was performed as per Studier and Moffatt (1986). Ten µl of ligation mix was added to competent DH5α and incubated on ice for 45 min. The cells were given heat shock for 5 min at 37 °C and transferred to ice for 3 min. One ml LB broth was added to the cells and incubated at 37 °C for 1 h with constant shaking, for the cells to recover. The pellet was collected by centrifugation at 10,000 rpm for 5 min, was resuspended in 100 µl of LB broth and spread onto LB agar plates containing 50 µg/ml of ampicillin for selection of transformed cells. For blue white screening of recombinants in pGEM-T and pUC19 vectors, LB agar was supplemented with 80 µg/ ml X-gal, 1 mM IPTG apart from 50 µg/ ml ampicillin.

Transformation of BBTV CP/ pGEX-4T-2 RF clone was carried out as follows. Following the RF reaction ie., second PCR, the parental DNA strand was eliminated by Dpn1

digestion. Nine µl from both RF and control reactions were aliquoted to 0.6 µl microcentrifuge tubes and 1 µl of Dpn1 was added and incubated at 37 ºC for 2 h. The enzyme was heat inactivated at 80 ºC for 20 min on heat block after which it was transformed into DH5α strains of *E. coli*.

In order to amplify plasmids and clones, approximately 20 ng of vectors or recombinant were transformed to 100 µl of competent DH5α as mentioned above and plated on to LB agar supplemented with ampicillin after recovery. Similarly, transformation of the clones to various strains of *E. coli* for expression were carried out following the same procedure. Chloramphenicol (34 µg/ml) was supplemented in the media for selecting transformants in *E. coli* strains with pLysS plasmid.

3.13.5 Clone confirmation

Clones were confirmed by colony PCR, restriction digestion and PCR of isolated recombinant plasmid to amplify the gene of interest and sanger sequencing with primers specific to gene or promoter present in the vector. All the four or atleast any three methods were performed for confirming each clone constructed in this study.

3.13.5.1 Colony PCR

Randomly selected white colonies were streaked on to LB agar plate containing ampicillin, divided into grids and allowed to grow at 37 ºC overnight. Single colony each from each grid was picked up and dissolved in 20 µl autoclaved Milli-Q water taken in a PCR tube and mixed well with micropipette tip. The PCR tube was incubated at 98 ºC for 2 min for denaturation. The cell debris was removed following 2 min of centrifugation at 10,000 rpm. The supernatant (2 µl) was used as the template for conventional PCR reaction with Taq polymerase using appropriate primers as described earlier. The PCR products were visualised on 1.2 per cent agarose gel.

3.13.5.2 Recombinant plasmid isolation

Plasmid was isolated from randomly selected transformed white colonies by alkaline lysis method (Green and Sambrook, 2012). Single *E. coli* DH5α colony was inoculated to 5 ml LB containing 50 µg/ml ampicillin and incubated at 37 ºC for 12 h or until A_{600} reached 0.6. The cells were harvested by centrifugation (10,000 rpm for 10 min) and the supernatant was discarded. The pellet was resuspended in 150 µl of P1 solution (50 mM Tris, 10 mM

EDTA, 100 µg/ml RNase A, pH 8.0) by vortexing. In the next step, equal volume of freshly prepared P2 solution (1 % SDS, 0.2 N NaOH) was added to the resuspended pellet and allowed to stand till the solution became clear. Subsequently, 200 µl of P3 solution (3.0 M potassium acetate, pH 5.5) was added and mixed gently by inverting the tube. Insoluble cell debris was pelleted by centrifugation at 10,000 rpm for 20 min. The supernatant was transferred into a new tube without disturbing the pellet and equal volume of isopropanol (500 µl) was added, mixed gently by inverting 5-6 times and incubated at 4 ºC for 10 min for precipitation. The plasmid DNA was harvested by centrifugation at 10,000 rpm for 15 min and the supernatant was discarded. After washing the pellet with 70 per cent ethanol, the recombinant plasmid was air dried and dissolved in 20 µl sterilised water.

Plasmid was used as template for amplification of the CP gene using gene specific primers to confirm the clones. Restriction digestion of the recombinant plasmid (5 µl of concentrated plasmid) with appropriate endonucleases (0.5 U) was performed to confirm the clones. In case of blunt end cloning with Sma1 recognition sites present in the vector, upstream and downstream of the gene of interest was selected to release the insert. Confirmed clone was transformed to various *E. coli* expression hosts for expression as mentioned in the previous section.

The confirmed clones were designated as follows, pREST/BBTV CP, pGEX/BBTV CP, pET/BBTV CP, pRSET/BBrMV CP and pGEX/BBrMV CP. The BBTV CP cloned in to pGEX vector by RF cloning method was designated as ΔpGEX/BBTV CP.

3.14 CHECKING FOR EXPRESSION OF PROTEIN FROM RECOMBINANT CLONES

The *E. coli* strains BL21(DE3) and Rosetta with pLysS and C41 were transformed with recombinant plasmids as described earlier and induced for protein expression as follows. pRSET clones yielded proteins with hexa histidine tag, pGEX vectors overexpressed GST fusion protein and pET32a (+) gave rise to fusion protein with a thioredoxin (Trx) tag along with an N terminal hexa histidine tag. Single colony was inoculated in 5 ml of LB broth containing required antibiotic(s) and grown at 37 °C in a screw-cap inoculation vial for 4-6 h. The culture was chilled at 4 ºC for 1 h. From the culture, 2 ml was set aside as uninduced *E. coli* extract and the remaining 3 ml culture was induced with 1 mM IPTG. Following induction, the culture was further grown for 12 h at 16 °C. The induced and uninduced cells were harvested by centrifugation and the pellet was resuspended in 100 µl sterilised water

and lysed by sonication (1 min cycle with 2 s interval between 3 s pulse of 35 % amplitude). The sonicated samples, (20 µl) were mixed with equal volume of 2X Laemmli sample loading buffer (100 mM Tris, pH 6.8 containing 2 % SDS, 0.02 % bromophenol blue, 10 % BME and 20 % glycerol), boiled for 10 min and analysed on a 12 per cent polyacrylamide gel by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970).

To obtain maximum expression of the protein, culture media (LB, yeast extract tryptone; YT, terrific broth; TB), temperature (37 ºC, 30 ºC, 16 ºC), period of induction (4 h, 6 h, 12 h, 14 h) and concentration of inducer (0.3 mM, 0.5 mM, 1 mM) were standardised (preparation of culture media is explained in Appendix II). To increase solubility of overexpressed BBrMV rCP cloned to pRSET-C, various lysis buffers like Tris buffer (50 mM Tris, 150 mM NaCl), CAPS buffer (20 mM CAPS, 200 mM NaCl, pH 9.2) and HEPES buffer (20 mM HEPES, pH 7.5) were used. pH of Tris buffer (8.0, 8.5) was also changed to obtain maximum protein in solution.

The *E. coli* BL21(DE3) cells were used for overexpressing tobacco etch virus (TEV) protease. In the LB broth used for culturing the transformants, only 50 µg/ml ampicillin was added. Induction was done with 0.5 mM IPTG at 25 ºC overnight at constant shaking. Rest of the protocol was as described earlier.

The recombinant proteins were designated as follows, pREST/rBBTV CP, pGEX/rBBTV CP, pET/rBBTV CP, pRSET/rBBrMV CP and pGEX/rBBrMV CP. The BBTV CP expressed from ΔpGEX/BBTV CP was designated as ΔpGEX/rBBTV CP.

3.15 SODIUM DODECYL SULPHATE POLYACRYLAMIDE GEL ELECTROPHORESIS

Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis (SDS PAGE) was carried out using discontinuous buffer system and slab gel electrophoresis apparatus. In this study, 1.5 mm thick 12 per cent polyacrylamide (30 %; 29:1 w/w acrylamide and bisacrylamide) gel containing 10 per cent SDS, were used for electrophoretic separation of proteins (Appendix II). The protein samples were mixed with an equal volume of 2X Laemmli loading buffer and boiled for 10 min or heated at 95 ºC for 10 min and were loaded into the wells of the precast polyacrylamide gel. Electrophoresis was performed at a constant voltage of 125 V using 25 mM Tris and 200 mM glycine buffer, pH 8.8, containing 0.1 per cent SDS. After

the electrophoresis, the gel was either stained with Coomassie brilliant blue R 250 (CBB; Merck) or processed for western blotting.

3.16 TRICINE-SDS-PAGE

In order to separate low molecular weight proteins (1-30 kDa) of similar size, Tricine-SDS PAGE was performed (Schägger, 2006). The BBTV rCP fusion protein cleaved using TEV protease was separated in 10 per cent Tricine SDS-PAGE prepared using 50 per cent acrylamide solution in 3X gel buffer (Appendix II). Cathode and anode buffer (both 1X) were poured to upper and lower tanks respectively. Volume of 20 µl of TEV protease treated protein was loaded after boiling with 2X Laemmli sample loading buffer for 10 min in order to visualise 20 kDa BBTV rCP and 17 kDa thioredoxin (tag). Vertical electrophoresis was carried out at constant voltage of 30 V initially which was later increased to 100 V.

3.17 WESTERN BLOT

The cell lysates (induced or uninduced culture or soluble and insoluble fractions of induced culture) or purified protein, was loaded on to 12 per cent gel and SDS-PAGE was performed. The proteins bands were electrophoretically blotted on to NCM (Bio-Rad) for 3 h as per the protocol of Towbin et al. (1979). Subsequently, the blot was subjected to Ponceau's reversible staining to mark the protein molecular weight marker for reference. This step was excluded when pre-stained protein molecular weight marker (Bio-Rad) was used instead of unstained marker. Blocking of unoccupied sites of the blot was done by incubating the blot with blocking buffer (5 % skim milk powder in PBS) for 1 h at room temperature. The blot was washed thrice with PBS-T (PBS containing 0.1 % v/v Tween 20) followed by incubation in primary antibody of appropriate dilution (1:5000 v/v of PVBV CP specific antibody, 1:500 v/v BBTV CP specific antibody, 1:5000 v/v of anti-GST polyclonal and 1:5000 v/v of anti-his monoclonal antibody) at 4 ºC overnight. The blot was washed thrice with PBS-T and twice with PBS to remove unbound antibodies followed by incubation with 1:10000 dilution of secondary antibody (goat anti-rabbit IgG for primary polyclonal antibody and goat anti-mouse IgG for primary monoclonal antibody) conjugated with horse raddish peroxide (HRP). Washing with PBS-T was repeated to remove unbound antibody and the blot was developed by electrochemiluminescence (ECL) with Immobilon chemiluminescent HRP substrate

(Millipore) as per manufacturer's protocol and visualised in a gel-doc accessory (ImageQuant LAS 4000, GE Healthcare).

3.18 PURIFICATION OF rCP

For purification of overexpressed protein, expression hosts harbouring recombinant plasmids were inoculated to 50 ml of bacterial growth media (LB or YT). The primary culture was allowed to grow overnight at 37 ºC. From this primary culture, 5 ml was inoculated to 1 L culture broth with appropriate antibiotic(s) for mass multiplication and incubated at 37 ºC until A_{600} reached 0.6. The culture was chilled at 4 ºC for 1 h prior to induction with 0.5 mM IPTG at 16 ºC. The culture was incubated for 12-14 h under constant shaking at 100 rpm.

3.18.1 Purification of BBTV rCP

Expression of BBTV CP in pGEX-4T-2 (conventionally cloned and RF cloned) produced fusion protein with N terminal GST tag aiding its solubility. The clones pGEX/BBTV CP and ΔpGEX/BBTV CP harbouring C41 strain inoculated in 1 L LB was induced with 0.5 mM IPTG for 14 h. The bacterial cells after induction were harvested by centrifugation at 8000 rpm for 20 min at 4 ºC (JA-10 rotor, Avanti J-E centrifuge, Beckman Coulter) and resuspended in 30 ml of lysis buffer (50 mM Tris, pH 8.0; 200 mM NaCl; 1 mM EDTA, pH 8.0; 10 % glycerol; 5 mM BME). The cells were lysed by sonication to release the protein. Two cycles of 5 min each with 6 s interval between 3 s pulse of 30 per cent amplitude was given. Successively, two more cycles of 3 min each were given after an interval of 5 min between each cycle. The cell debris was removed by centrifugation at 8000 rpm at 4 ºC for 40 min. The supernatant (soluble fraction) was allowed to bind to calibrated GSH sepharose beads (1 ml bead volume) for 5 h at 4 ºC by mild rotation in an end-to-end rotor. The protein-resin complex was packed into a column and washed twice with 10X bed volume with wash buffer (50 mM Tris, pH 8.0, 200 mM NaCl) to remove unbound proteins. BBTV CP-GST fusion proteins (pGEX/rBBTV CP and ΔpGEX/rBBTV CP) were eluted with 100 mM Tris; pH 8.0, 200 mM NaCl and 30 mM glutathione. The purity of protein was checked on 12 per cent SDS-PAGE.

Expression of BBTV CP in C41 *E. coli* cells containing recombinant pET32a(+), yielded pET/rBBTV CP with hexa histidine and thioredoxin tags former aiding in purification and latter aiding in increasing solubility of the protein. The induced cell pellet harvested by

centrifugation, was resuspended in 30 ml of lysis buffer (50 mM Tris, pH 8.0 containing 200 mM NaCl; 10 % glycerol and 5 mM BME). After three cycles of sonication with 30 per cent amplitude providing 6 s interval between 3 s pulse in each cycle (two cycles of 5 min each and one cycle of 3 min), the lysed cell debris was removed by centrifugation at 8000 rpm for 30 min at 4 ºC and the soluble fraction was allowed to bind to calibrated Ni^{2+}-NTA agarose beads (1 ml bead volume) for 6 h at 4 ºC using an end-to-end rotor. The column was packed with the protein-resin complex and flow through was discarded. The beads were washed successively using 10X bead volume of wash buffer (50 mM Tris, pH 8.0 containing 200 mM NaCl) with 10, 25 and 50 mM imidazole. The rBBTV CP-Trx fusion protein was eluted in 50 mM Tris, pH 8.0 containing 200 mM NaCl and 250 mM imidazole. Elutions together with flow through, wash, soluble and insoluble fractions were loaded in to SDS-PAGE to check purity.

3.18.2 Purification of rBBrMV CP

The BBrMV CP gene was cloned to pRSET-C and pGEX-4T-2. The pRSET/BBrMV CP and pGEX/BBrMV CP overexpressed in Rosetta pLysS or BL21 pLysS was purified by Ni^{2+}-NTA chromatography and GSH sepharose affinity column chromatography respectively. The procedure employed for purification was similar as followed for purification of rBBTV CP with few modifications.

Lysis buffers used for Ni^{2+}-NTA purification were Tris buffer (20 mM Tris, pH 8.0 containing 200 mM NaCl, 5 % glycerol, 1 % Triton-X-100 and 10 µL protease inhibitor cocktail), CAPS buffer (40 mM CAPS, 200 mM NaCl, 5 % glycerol and 0.5 % Triton X-100, pH was adjusted to 9.2 by NaOH) and HEPES buffer (20 mM HEPES, glycerol and 0.5 % Triton X-100, pH 7.5). Rosetta pLysS cells harbouring pRSET/BBrMV CP recombinant plasmid was cultured in 1 L YT broth and induced by 0.3 mM IPTG for 12 h at 16 ºC. Induced cells were harvested by centrifugation and resuspended in 15 ml lysis buffer and lysed by sonication for 10 min (two cycles of 5 min each). The soluble fraction obtained after centrifugation was allowed to bind to Ni^{2+}-NTA beads (500 µl bead volume) for 4 h and purification was carried out at 4 ºC. Appropriate wash buffer and elution buffer were also selected according to the lysis buffer used. Elution was carried out in appropriate buffer containing 300 mM imidazole.

The BBrMV CP gene cloned to pGEX-4T-2 was overexpressed and purified in both BL21 and Rosetta (DE3) pLysS strain of *E. coli*. The induced cells harvested by centrifugation was lysed by sonication in Tris buffer, pH 8.5 (50 mM Tris, pH 8.5 containing 150 mM NaCl, 1 mM EDTA, 5 % glycerol, 1 % Triton-X 100, 5 mM BME and 10 µl protease inhibitor cocktail). The GSH sepharose beads (500 µl) was added to the soluble fraction and incubated at 4 ºC in an end-to-end rotor for 4 h and subsequently packed to a column. Washing was done with Tris buffer (50 mM Tris, pH 8.5 containing 150 mM NaCl) equal to 10X bed volume followed by elution with 30 mM glutathione. Elutions were loaded on to SDS-PAGE to confirm purity.

3.18.2.1 Purification of rBBrMV CP in denaturing conditions

Purification of rBBrMV CP cloned to pRSET-C and expressed in Rosetta pLysS was also attempted at denaturing conditions in the presence of 8 M urea and 6 M guanidine hydrochloride in order to bring the recombinant protein in to solution. Entire process of purification was carried out at room temperature. For urea denaturation, the harvested pellet from 1 L Rosetta pLysS harbouring pRSET/BBrMV CP cultured in YT medium was treated with lysis buffer (20 mM Tris, pH 8.0; 200 mM NaCl; 5mM BME; 10 µl protease inhibitor cocktail and 10 mM imidazole) supplemented with 8 M urea and incubated for 1 h by constant stirring. Further lysis of the cells was carried out by sonication for 5 min at 30 per cent amplitude (6 s intervals were given between 3 s pulse). Supernatant after centrifugation (8000 rpm, 15 min) was collected and allowed to bind to Ni^{2+}-NTA (500 µl bead volume) for 3 h. After packing the column with protein-resin complex, washing with wash buffer (20 mM Tris, pH 8.0; 200 mM NaCl) containing 20 mM imidazole was repeated twice. The protein was eluted in Tris buffer (20 mM Tris, pH 8.0; 200 mM NaCl) with 300 mM imidazole. Wash buffer and elution buffer with 8 M urea were also used for standardising purification protocol to obtain higher yield of protein. For denaturation with 6 M guanidine hydrochloride, CAPS buffer of pH 9.2 was used for lysis, washing and elution of the protein instead of Tris buffer. Washing and elution were done only in the presence of denaturing agent in this case.

3.18.3 Purification of TEV protease

The *E. coli* BL21(DE3) harbouring recombinant plasmid with TEV protease gene was induced by 0.5 mM IPTG and incubated at 25 ºC overnight. The pellet was harvested by centrifugation at 8000 rpm, 4 ºC for 20 min and resuspended in 30 ml lysis buffer (50 mM

Tris, pH 8.0; 200 mM NaCl; 10 % glycerol) and sonicated with 30 per cent amplitude (3 s on 6 secs off cycle) for 5 min. Two sonication cycles of 3 min each were given after an interval of 5 min. Consequently, the supernatant was collected after centrifugation and was allowed to bind to 1 ml calibrated Ni^{2+}-NTA beads in an end-to-end rotor for 6 h at 4 ºC. The lysate-resin mix was loaded to the column, and after removing the flow through, beads were successively washed with Tris NaCl buffer, pH 8.0 containing 10 mM, 25 mM and 50 mM imidazole to remove the unbound and weakly bound proteins. For elution, 50 mM Tris buffer of pH 8.0, containing 200 mM NaCl and 100 mM imidazole was used. Eluted samples were subjected to SDS-PAGE to confirm the presence of TEV protease.

3.19 CLEAVAGE OF AFFINITY TAG FROM FUSION PROTEIN

The attempt to cleave off GST tag of BBTV CP expressed in C41 harbouring ΔpGEX/BBTV CP was carried out using TEV protease overexpressed and purified in the laboratory. Different concentration of protease (2.0, 1.0, 0.5 and 0.25 µg) was used to cleave 100 µg of ΔpGEX/rBBTV CP fusion protein at 20 °C for 12 h. Attempt to remove Trx tag from pET/rBBTV CP overexpressed in C41 was carried out by thrombin (Millipore) (0.2 µg, 0.1 µg) at 20 ºC for different time intervals (4 h, 6 h, 8 h and 12 h). the GST tag of BBrMV fusion protein was removed using thrombin, 5 U/ mg protein. In order to separate BBrMV rCP from GST tag, the digested protein was allowed to bind to GSH sepharose beads and tag less recombinant protein alone was collected. On-column digestion using 10 U/ml bead volume of thrombin to obtain the tag less protein was attempted at 25 °C for 8, 12 and 14 h to compare the best method for removal of the tag.

3.20 DIALYSIS OF PURE PROTEIN

Recombinant coat protein was dialysed to concentrate the protein and also to remove imidazole, glutathione or proteases used for cleavage of the fusion protein in the solution. A high buffer-to-sample volume-ratio was maintained to aid in maintaining the concentration gradient (50 X volume of all the elutions pooled) (Walker, 2009). Dialysis buffer corresponding to the elution buffer was prepared with 10 per cent glycerol and without eluent. Dialysis membrane of molecular weight cut off (MWCO) of 8 kDa was used for dialysis. The membrane was washed thoroughly with distilled water and with the dialysis buffer. Pooled rCP was loaded into the membrane which was sealed and placed in the buffer. The set up was kept at 4 ºC overnight. Dialysis buffer was changed 2-3 times in order to obtain required concentration of the protein.

3.21 QUANTIFICATION OF RECOMBINANT COAT PROTEIN

The rCP (fusion proteins and untagged proteins) were quantified using Bradford's reagent. Protein working standards of Bovine serum albumin (BSA) of concentrations 0.1 mg/ml to 1.0 mg/ml were prepared. In to 950 µl Bradford's reagent (Sigma) 50 µl of 0.1, 0.2, 0.4, 0.8 and 1.0 mg/ml BSA was added, allowed to react in dark. Two replications each containing 25 µl and 50 µl of purified rCP in 950 µl Bradford's reagent was prepared. The volume of the samples was adjusted to 1 ml using buffer whenever necessary. Absorbance at 595 nm of known concentration of protein was measured and standard graph was plotted. Absorbance value of rCP was recorded and its concentration was calculated from the linear equation in Microsoft Excel.

3.22 IN SILICO ANALYSIS OF CP

3.22.1 Fold analysis

Understanding the folding of protein is an important determinant of its function. Folding of BBTV and BBrMV CPs were analysed using FoldIndex (Prilusky et al., 2005). The ordered and disordered regions in the protein were predicted. The folding of fusion proteins *viz.*, BBTV CP-GST, BBTV CP-Trx and BBrMV CP-GST, were also analysed separately and mapped using the software.

3.22.2 Analysing protease cleavage sites in rCP

The peptide was analysed for presence of protease cleavage sites using PeptideCutter tool. This analysis was done to determine the protease responsible for degradation of the protein in solution. Sites of cleavage of trypsin, chymotrypsin and thrombin were determined and mapped on the translated protein.

3.22.3 Secondary structure prediction

Secondary structure of the protein was predicted by PSIPRED workbench. The helices, β sheets, turns and coils present in the CP of BBTV and BBrMV were determined. The prediction was based on three methods, PSIPRED (Jones, 1999), MEMSAT 2 (Nugent and Jones, 2010) and GenTHREADER (McGuffin and Jones, 2003) from the translated amino acid sequences of the isolates.

3.22.4 Protein structure prediction

Threading of CP of BBTV and BBrMV was carried out in i-TASSER suite and Phyre2 software using the amino acid sequence, by comparing with most similar amino acid sequences of protein deposited in the protein databank with solved structure (Yang et al., 2014; Mezulis et al., 2015). The structure was evaluated by Ramachandran plot in MolProbity web server (Williams et al., 2018). The quality of the model was assessed in ERRAT and ProSA web servers (Colovos and Yeates, 1993; Wiederstein and Sippl, 2007). The SuSPect analysis was carried out to predict effect of mutation in a particular position of amino acid sequence that has an impact on the function of the protein (Yates et al., 2014) and mutational analysis graph was plotted. The sequence profile graph representing residue preferences in the protein at a particular sequence position was also mapped out. Structures of the most homologous and the most diverse sequence were predicted and superimposed using FATCAT (Flexible structure AlignmenT by Chaining Aligned fragment pairs allowing Twists) (Li et al., 2020). The molecular models were superimposed to identify the structural differences between the isolates that is reflected from the variation in amino acid sequences.

3.22.5 Epitope prediction

Analysis of major and minor epitopes on the antigen (CP used for immunising animal for antiserum production) and binding to B cells has enormous potential in diagnostics. It is equally important to assess whether these epitopes are conserved across different variants of the virus especially RNA viruses which are prone to mutation than DNA viruses. The IEDB web server was used to predict continuous B cell binding epitopes on BBrMV CP and BBTV CP based on BepiPred linear epitope prediction 2.0 method (Jespersen et al., 2017). Epitope prediction was also done using DiscoTope 2.0 and mapped on the predicted molecular protein structure (Kringelum et al., 2012).

3.23 BIOPHYSICAL CHARACTERISATION OF rCP

3.23.1 Assessment of self-assembly of CP by ultracentrifugation

The BBTV CP fusion protein with Trx tag was subjected to ultracentrifugation to assess whether CP assembles itself into VLPs *in vitro*. Fusion protein was overexpressed in C41 strain of *E. coli*. The cells harbouring pET/BBTV CP recombinant plasmid was induced with 0.5 mM IPTG for 14 h and pellets harvested by centrifugation at 8000 rpm for 20 min

at 4 ºC. The pellets were resuspended in Tris buffer (50 mM Tris, pH 8.0 containing 200 mM NaCl) and lysed by sonication. The cell debris was removed by centrifugation at 8000 rpm for 30 min at 4 ºC. The supernatant was subjected to high speed ultracentrifugation at 26000 rpm for 3 h at 4 ºC (MLS 50 rotor, Benchtop ultracentrifuge, Beckman coulter). The supernatant and ultra pellet was viewed on 12 per cent SDS-PAGE. The pellet was resuspended in 500 µl – 1 ml of Tris buffer and incubated in an end-to-end rotor overnight at 4 ºC. Linear sucrose gradient (10-40 %) were prepared and poured to ultracentrifuge tubes and equilibrated overnight at 4 ºC. The ultra pellet suspension was layered on the linear sucrose gradient and centrifuged at 26000 rpm for 3 h at 4 °C. Fractions of 250 µl were collected and analysed by SDS-PAGE. The peak fractions were pooled and subjected to ultracentrifugation at 26000 rpm for 3 h at 4 ºC again. The pellet obtained was resuspended in Tris buffer and used for electron microscopy.

The purified tag less rBBrMV CP and rBBrMV CP-GST fusion protein were subjected to ultracentrifugation at 26000 rpm for 3 h at 4 °C to assess self-assembly. The supernatant and pellet were separated using SDS-PAGE. The pellet was dissolved in minimum volume of Tris NaCl buffer (50 mM Tris, pH 8.0 and 150 mM NaCl) and laid over equilibrated linear sucrose gradient poured to ultracentrifuge tubes. The resuspended ultra pellet was also subjected to electron microscopy. Rest of the steps were followed as described earlier.

3.23.2 Electron microscopy

A concentration of 0.1 mg/ml protein (ultracentrifuged BBTV CP-Trx fusion protein and tag less rBBrMV CP) was adsorbed on formvar coated copper grids (SPI Supplies, USA, Code: 3440C-MB) for 2-3 min followed by staining with filter sterilised 1 per cent uranyl acetate for 2 min. Excess stain was absorbed by tissue paper and then air dried for 5 min. The grids were viewed in Tecnai F30 Transmission electron microscope at 120 kV (ThermoFisher, USA) at 1,60,000X magnification at Electron microscopy facility, Biological Sciences, IISc, Bengaluru.

3.23.3 Fluorescence spectroscopy

Changes in intrinsic fluorescence contributed by three amino acid residues in the polypeptide *viz.*, tryptophan, tyrosine and phenylalanine facilitate in monitoring structural changes of the protein. The intrinsic fluorescence was measured using TECAN LS55 Luminescence Spectrometer (Perkin-Elmer). Spectra of rBBTV CP fusion protein with

GST tag and Trx tag were compared. 0.1-0.5 µg protein was dissolved in 50 mM Tris, pH 8.0 and 200 mM NaCl buffer for analysis. Spectra of BBrMV CP with and without tag were also recorded at protein concentrations of 0.1-0.5 µg in buffer containing 50 mM Tris, pH 8.5 and 150 mM NaCl. Blank was set with the same buffer without recombinant protein. The excitation wavelength was 280 nm and the emission was scanned between 300-400 nm.

3.23.4 Mass spectroscopy of rBBTV CP

Self-assembled rBBTV CP separated in 12 per cent SDS was eluted and prepared for mass spectroscopy as per the protocol of Gundry et al., (2010). The band at 37 kDa was cut in to small pieces using a sterilised blade and transferred to sterile microcentrifuge tube. The band was destained by treating the gel pieces with destaining solution (50 % v/v acetonitrile; ACN, 50 % 25 mM ammonium bicarbonate; NH_4HCO_3). After removing destaining solution, it was treated with 100 per cent ACN. After several rounds of repeating steps one and two alternatively, 400 µl autoclaved Milli-Q water was added to the destained gel. Water was removed and the gel was dehydrated with 400 µl 100 per cent ACN for 5 min. The solution was discarded and to the dried gel, 100 µl of 10 mM dithiothreitol (DTT) was added and incubated at 55 ºC for 45 min. The supernatant was removed and 100 µl of freshly prepared 50 mM iodoacetamide was added to the gel and incubated in dark for 40 min at room temperature. After removing iodoacetamide, 400 µl of gel wash solution (50 % v/v acetonitrile, 25 mM NH_4HCO_3) was added and incubated for 15 min. This step was repeated. The gel was again dehydrated by 400 µl of 100 per cent ACN for 10 min at room temperature. The enzyme stock solution was diluted to 1:1000 in 25 mM NH_4HCO_3. From the working stock (~20 ng/µl), 20µl was added to the gel so that the gel immerses in the enzyme solution and incubated on ice for 1 h. Excess solution was removed and sufficient volume of 25 mM NH_4HCO_3 to immerse the gel was added. Incubation was performed at 37 ºC overnight. The supernatant containing peptides were removed to new microcentrifuge tube post incubation. To the gel, 50 µl gel extraction solution (50 % v/v ACN, 1 % v/v trifluoroacetic acid) was added and vortexed at room temperature for 20 min. The solution was removed and added to the solution of peptides previously transferred to a separate microcentrifuge tube. To extract maximum peptide from the gel aforementioned two steps were repeated. The pooled peptide solution was incubated at 40 ºC for 30 min to concentrate to a final volume of 20-25 µl. The peptide solution was desalted before it was subjected to Matrix-Assisted Laser Desorption/Ionization-Time Of Flight mass spectrometry (MALDI-TOF MS).

3.23.4.1 Data analysis

Mass spectrometry data was examined using Mascot search algorithm against UniProt protein database. The search parameters included a maximum of 1 missed cleavage without any fixed or variable modifications. The MS error tolerance was set at 1.2 Da and MS/MS error tolerance at 0.6 Da. The data were searched against a decoy database and the results were used to estimate q values using the Percolator algorithm within the Proteome Discoverer suite. Peptides were considered identified at a q value of < 0.05.

3.24 ANTISERUM PRODUCTION AGAINST pET/rBBTV CP AND TITRE ESTIMATION

The BBTV CP fusion protein with thioredoxin tag was used for immunising New Zealand white rabbit to raise rCP specific antiserum. The experiment was carried out at the animal facility, LivGen Biotechnologies, Bengaluru. Before immunisation, pre-immune bleed was drawn and designated as control serum. Primary injection with 300 µg of antigen in Freund's complete adjuvant was administered to the animal. Three booster doses with 100 µg antigen emulsified with Freund's incomplete adjuvant was given at 10-15 days interval. First and second bleeds were collected at an interval of five days after the third booster. Two more booster doses were given at an interval of 15 days with 100 µg antigen emulsified with Freund's incomplete adjuvant 20 days after the second bleed. Third bleed was collected one week after administering the final booster dose.

Blood samples collected from the jugular vein of immunised rabbit was left undisturbed at room temperature until the blood clotted. The clotted blood was gently cut with the help of sterile toothpick and refrigerated overnight for the serum to ooze out. The clear serum was collected carefully in microcentrifuge tube and centrifuged at 1000 rpm for 15 min at 4 ºC. The fluid at the top containing the immunoglobulins were collected in sterile microcentrifuge tube for further analysis. Bleeds after each booster dose was collected similarly and titre was examined by DAC-ELISA.

3.24.1 Determination of reactivity of recombinant antiserum

Dot blot analysis was conducted to determine the reactivity of serum to specific protein. One µL of purified protein (used for immunisation) was spotted on the nitrocellulose membrane and air dried. Once the spots were dried, the membrane was blocked with a blocking buffer composed of 0.1 per cent BSA dissolved in TBS-T buffer (20 mM Tris; pH

8.0, 500 mM NaCl, 0.05 % Tween 20). Later, 30 µL of serum diluted to 1: 10000 (v/v) in TBS-T was added and allowed to bind at room temperature for 1 h. Unbound serum was washed with wash buffer (TBS-T) allowed to dry. Subsequently, 50 µL of Protein A gold conjugate was used to visualize the antibody antigen interaction.

3.25 STANDARDISATION OF ELISA FOR BBTV DETECTION AND COMPARISON WITH COMMERCIAL ANTISERUM

3.25.1 DAC-ELISA

The protocol followed for performing DAC-ELISA as mentioned previously in section 3.4. Titre value of the antiserum for detecting field and tissue culture samples were determined. Various dilutions *viz.*, 1:100000, 1:50000, 1:25000, 1:10000, 1:5000, 1:2500, 1:1000, 1:500 was prepared in PBS-T supplemented with 2 per cent PVP and 0.2 per cent BSA. The titre value of commercially available antiserum was determined earlier and it was kept as a control in the experiment. The reaction between primary and secondary antibody was studied using various dilutions of primary and secondary antibody (enzyme conjugate). The sensitivity of ELISA was tested by diluting antigen further to 1:10, 1:20, 1:40, 1:80 with coating buffer and detecting it at the tire value obtained from previous experiments.

3.25.2 Immuno Capture-Polymerase Chain Reaction

A concentration of 10 ng/ml of antiserum raised against rBBTV CP and commercial antiserum in coating buffer was coated on to PCR tubes for overnight at 4 ºC. Antigen was isolated from infected sample in various buffers like PBS (supplemented with 0.05 % Tween-20, 2 % PVP), coating buffer, extraction buffer (500 mM Tris, pH 8.0; 0.01 M sodium sulphite; 2 % PVP; 140 mM NaCl; 0.05 % Tewwn-20) and Tris buffered saline (20 mM Tris, pH 7.5; 200 mM NaCl). One g of infected plant tissue was ground in 1 ml of buffer and transferred to 1.5 ml microcentrifuge tubes. The supernatant was collected after centrifugation at 8000 rpm for 10 min. To the precoated PCR tubes 200 µl of antigen was added and incubated for 5 h at 37 ºC. After washing off the excess antigen, the tubes were heated at 80 ºC for 5 min to release the genome. Release of the genome was also tested in the presence of transfer buffer (10 mM Tris, pH 8.0; 0.05 % Tween-20) and GES buffer (0.1 M glycine, pH 9.0; 50 mM NaCl; 1mM EDTA and 1 % BME) at 65 ºC for 10 min. The tube was transferred on to ice after heating for 3 min. The PCR was done directly in these tubes. The PCR mix was prepared as mentioned in the methods section previously

(3.11) and added to the tubes and centrifuged briefly. The designed BBTV DNA-S primers were used for amplifying the gene. PCR conditions standardised earlier for these designed primers were used for this experiment as well.

3.25.3 Dot Immuno Binding Assay

For performing Dot Immuno Binding Assay (DIBA), antigen was isolated in TBS buffer (20 mM Tris, 500 mM NaCl; pH 7.5) in 1: 10 w/v ratio. The homogenised plant tissue with buffer was transferred to Oakridge tubes and centrifuged at 5000 rpm for 5 min to remove the debris. The supernatant was transferred to another tube and equal amount of chloroform was added to it. After brief vortexing, the mix was centrifuged at 8000 rpm for 10 min to remove chlorophyll pigment in the sample. Later, 10 µl of the clarified solution was spotted on to Nitro cellulose membrane (NCM) activated with TBS buffer. Antigen was spotted in 1 cm^2 grids marked on the membrane. Separate strips for comparing the reactivity of antiserum produced in the present study and commercial antiserum were used in the experiment. After the membrane was air dried, blocking with 5 per cent SDM dissolved in TBS was carried out at room temperature for 1 h giving a gentle oscillation. The membrane was washed thrice with TBS for 5 min each to remove excess blocking buffer. Subsequently, 1:10000 v/v of recombinant antiserum or 1:1000 v/v of commercial antiserum diluted in TBS-SDM was added and incubated for 1 h 30 min at room temperature. Post washing with wash buffer, secondary antibody (1:10000) also diluted in TBS-SDM was added and allowed to react for 1 h at room temperature. 3,3'-diaminobenzidine (DAB) was used for developing the membrane. Five µl of 30 per cent H_2O_2 was added to 2 ml DAB (0.5 mg/ml) prior to developing the membrane. This solution was added on to the membrane and incubated in dark for 5 min for colour development. To estimate the sensitivity of the assay, 10 µl of antigen diluted up to 1:80 v/v was spotted on to the membrane. The highly concentrated unclarified antigen was also spotted on to the membrane to compare the results.

3.25.4 Triple antibody sandwich-Enzyme Linked Immunosorbent Assay

Triple antibody sandwich-Enzyme Linked Immunosorbent Assay (TAS-ELISA) was also standardised using the recombinant antiserum. BBTV specific mouse monoclonal antibody (Agdia) was diluted to 1:500 (v/v) in coating buffer and used as capture antibody. The microtitre plate coated with 100 µl of capture antibody was incubated at 37 ºC for 1 h. After

washing off the excess antibody with PBS-T, the unbound regions of the wells were blocked with 5 per cent SDM dissolved in PBS-T for 1 h at 37 ºC. Post washing thrice with PBS-T, 100 µL of antigen (1:10 w/v) extracted 1X PBS with 2 per cent PVP was allowed to bind to the capture antibody for 1 h at 37 ºC. Various dilutions of the antigen (1:10, 1:20, 1:40 and 1:80) was allowed to bind to capture antibody so as to assess the sensitivity of the assay. Primary antibody (1:10000 v/v PBS-TPO) was added after removing the excess unbound antigen and allowed to react with the trapped antigen. Anti-rabbit IgG conjugated with alkaline phosphatase enzyme was diluted to 1: 10000 v/v in PBS-TPO was added for detection. PNPP substrate solution (100 µl) was added and OD value was read in ELISA-Plate reader.

Results

4. RESULTS

4.1 SURVEY AND SAMPLE COLLECTION

Survey was conducted in 10 out of 14 districts in Kerala (Table 4). Locations selected for survey were divided into three population subsets *viz.*, Northern, Central and Southern zones (Plate 2). The main aim of the survey, was to examine if there are any variation in symptoms caused by BBTV and BBrMV in various cultivars of banana and also to ascertain, if there are any correlation in the symptoms with geographical area and cultivars. Therefore, symptomatology was studied in detail during the survey. Majority of the farmers cultivated the most popular variety Nendran followed by Njalipoovan, and Rasthali. A total of 50 BBTD and BBrMD infected plant samples were collected during the survey.

4.2 STUDIES ON BANANA BUNCHY TOP VIRUS

4.2.1 Symptomatology and disease incidence

The BBTV infected plants showed bunchy appearance of the leaves, later the margins of the leaves turned yellow and necrotic (Plate 3). Vein clearing and Morse code symptoms were observed on young leaves. However, rosetting or bunching of the leaves were observed in various stages (3-8 months old) of infected plants. Banana infected at earlier stage of growth showcased severe stunting of growth coupled with yellowing and necrosis from the leaf margin. Banana infected at a later stage, exhibited rosetting of leaves and bunching was absent causing 100 % yield loss. There were no considerable variations in the symptomatology of infected plants across various zones.

The disease incidence of BBTD was calculated as per the observations at the time of the survey (Fig 3). Maximum incidence of BBTD was recorded at Peruvazhikadavu (5.50 %) followed by Mala (4.80 %). Per cent disease incidence of BBTD in various blocks ranged from 0- 4.80 per cent, maximum in Mala. Among the districts, maximum incidence of the disease was recorded at Kozhikode (2.4 %) district followed by Thrissur district (2.36 %). At the time of survey, in three districts *viz.*, Thiruvananthapuram, Ernakulam and Kasaragod, incidence of BBTV was not recorded.

Table 4. Location details of survey conducted for the study

Zone	District	Block	Location	Latitude and longitude
Northern zone	Kasaragod	Neeleswaram	Palathadom	12°16'32.9"N 75°10'46.4"E
	Kannur	Thalassery	Kizhakumbhagam	11°48'01.2"N 75°30'18.3"E
	Kozhikode	Kunnamangalam	Chathamangalam Peruvazhikadavu Cheruppa Urkadavu Ayyappankavu	11°18'34.4"N 75°55'06.3"E 11°17'22"N 75°53'43"E 11°15'41"N 75°54'58"E 11°15'37.2"N 75°55'07.9"E 11°15'41"N 75°54'53"E
		Koduvally	Vavoor	11°14'11.2"N 76°00'11"E
	Wayanad	Kalpetta	Kalpetta Vellarimala	11°37'36.8"N 76°04'39"E 11°30'39"N 76°09'09"E
		Sulthan Bathery	Kuppady Sulthan Bathery	11°40'23.6"N 76°15'50.4"E 11°40'22.0"N 76°15'46.1"E
Central zone	Thrissur	Mala	Mala	10°14'11.4"N 76°16'10.2"E
		Ollukkara	Maraikal Pananchery	10°32'14"N 76°19'09"E 10°32'35.5"N 76°19'18.3"E
		Wadakkanchery	Naduthara Thekkumkara	10°39'49.3"N 76°14'19.8"E 10°38'24.0"N 76°16'21.0"E
		Chalakudy	Kottat	10°18'38.1."N 76°19'44"E

	Palakkad	Alathur	Pullode Kizhakanchery	10°38'32.6"N 76°33'45.1"E 10°32'14.8"N 76°29'28.3"E
		Ottapalam	Ottapalam	10°46'50.5"N 76°22'59.7"E
		Sreekrishnapuram	Karakurissi	10°53'54.1"N 76°24'20.0"E
	Ernakulam	Angamaly	Karukutty Kalady Vengoor	10°14'02.4"N 76°22'55.0"E 10°10'24.0"N 76°25'05.3"E 10°09'03.0"N 76°32'39.0"E
Southern zone	Kottayam	Erattupetta	Keeriyathottam	9°41'16"N 76°47'34"E
	Kollam	Kottarakkara	Vembayam	8°38'14.1"N 76°55'59.3"E
	Thiruvananthapuram	Neyyattinkara	Kannjiramkulam	8°21'05.9"N 77°03'28.1"E
		Athiyannor	Russelpuram	8°26'27.2"N 77°04'05.5"E
		Parassala	Parassala	8°20'26.8"N 77°09'42.9"E
		Nedumangad	Nedumangad	8°37'20"N 76°59'28.5"E

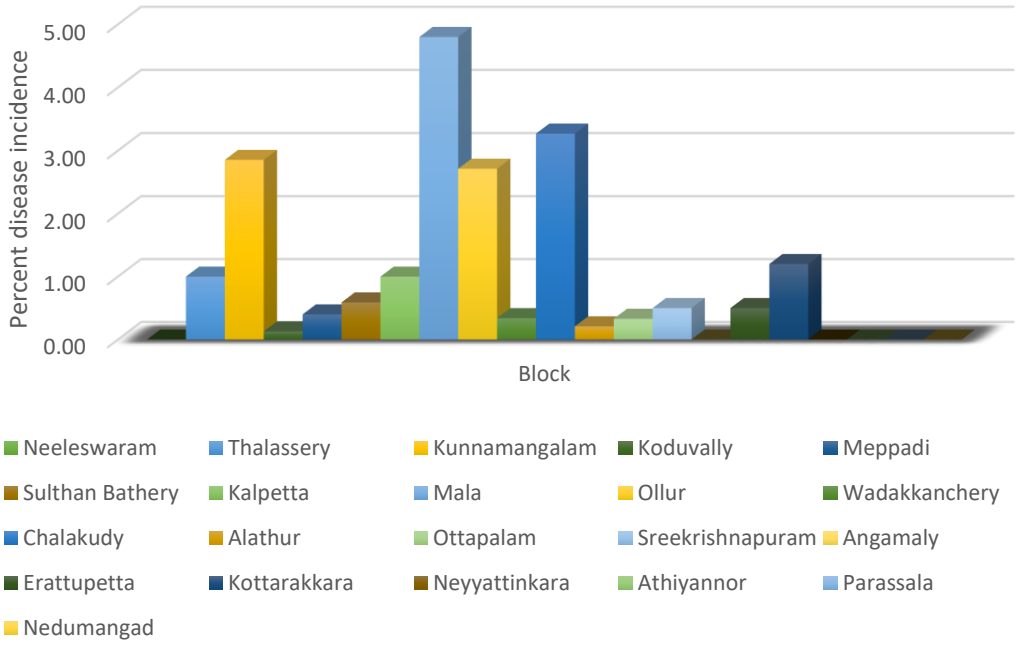


Fig 3. Block wise per cent incidence of banana bunchy top disease

Through detailed data collection, it was understood that in most of the places, disease was present however, as an when the symptoms are noticed, farmers rogued off the virus infected plants. Management practices like vector control by timely spraying of organic or inorganic insecticides in the field was practices by most of the farmers apart from cutting and burning the infected plants. Disease was seen to spread mainly through the suckers which are purchased in bulk without virus indexing from Tamil Nadu. Very meagre association of vector (*Pentalonia nigronervosa*) was observed in the fields.

4.2.2 Preliminary assay by DAC-ELISA

Antiserum titre of BBTV specific commercial antiserum procured from NRCB, Trichy was determined to be 1:1000 v/v through DAC-ELISA (Plate 4). Plant extracts suspected to be infected by virus were coated on to ELISA plate for DAC-ELISA. The extraction was done in coating buffer as per the method mentioned in section 3.4. Absorbance value at 405 nm more than twice that of negative control (NC) was only considered positive (Plate 4).

4.2.3 Total DNA isolation and PCR

Total DNA from 17 representative samples suspected to be infected with BBTV that gave positive results in DAC-ELISA was isolated by modified CTAB method (Plate 5A). High

Plate 2 Glimpses of survey conducted in northern, central and southern zones of Kerala for data and sample collection

Plate 3 **Symptoms of banana bunchy top disease recorded during the survey. A** Yellowing of the parallel veins of infected plant giving Morse code like appearance **B** Reduction in size of leaf lamina **C** and **G** Rosetting of leaves, dwarfing and absence of new flush **D** Necrosis starting from the leaf margin E Absence of bunching **F** Virus transmitted to suckers through infected mother plant

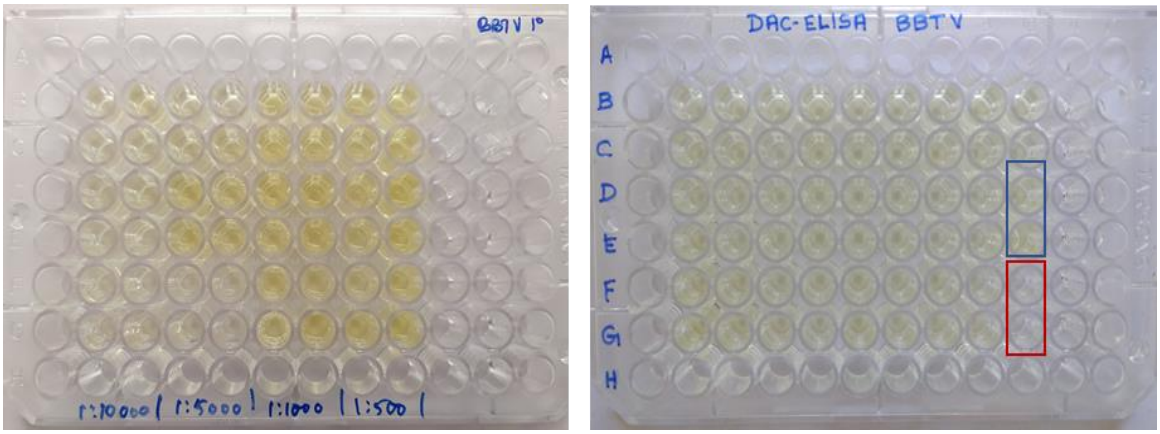


Plate 4 **Determination of antiserum titre and DAC-ELISA of field samples suspected to have BBTV infection.** Red squares denote negative control (NC) and blue squares denote positive control (PC)

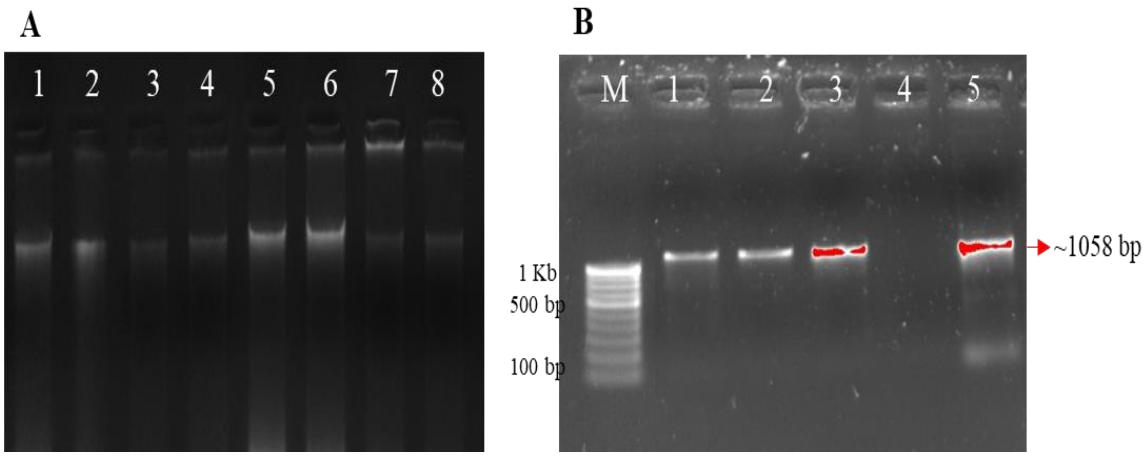
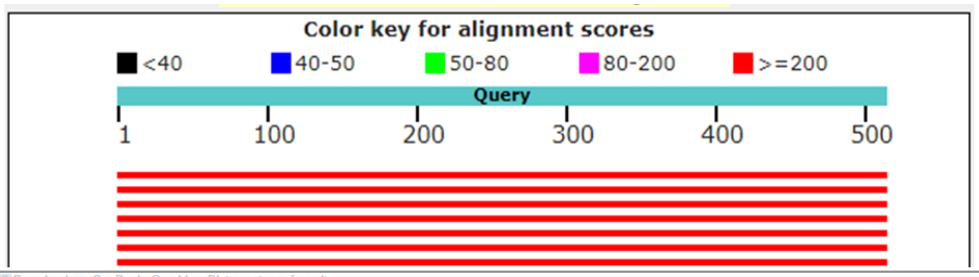


Plate 5 **Total DNA isolation and PCR A** Isolated DNA separated in 0.8 % agarose gel by electrophoresis **B** PCR performed using BBTV coat protein specific primers. An amplicon of 1058 bp obtained in positive samples. M: DNA ladder; Lanes 1-3: infected samples 4: negative control (NC) 5: positive control (PC)

molecular weight band corresponding to the genomic DNA was visible in the gel. Contamination with RNA was also observed. Polymerase chain reaction to amplify CP gene was carried out using reported primers (Plate 5B). An amplicon of 1058 bp was obtained after PCR. No bands were observed in negative control and primer dimers were also absent. Partial sequence of the DNA-S segment of BBTV was amplified using the reported primers.

4.2.4 Sanger dideoxy sequencing of coat protein gene and submission to GenBank

The pre-processed sequences obtained after Sanger dideoxy sequencing, were searched against the NCBI nucleotide repository and identified as CP gene of BBTV (Fig 4) having 99.81 % identity with the sequences in the database. Seventeen BBTV CP gene sequences were submitted in NCBI and accession numbers MT174314- MT174330 was obtained (Appendix III).



Alignments | Download | GenBank | Graphics | Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Banana bunchy top virus isolate Bmul-77 segment DNA-S, complete sequence	942	942	100%	0.0	99.81%	KU759885.1
Banana bunchy top virus isolate Luburn-9 segment DNA-S, complete sequence	942	942	100%	0.0	99.81%	KU759873.1
Banana bunchy top virus clone pTzBBSTri7 segment DNA-S, complete sequence	942	942	100%	0.0	99.81%	KJ513017.1
Banana bunchy top virus isolate Malawi 73 segment DNA-S, complete sequence	942	942	100%	0.0	99.81%	JQ820455.1
Banana bunchy top virus isolate Rwanda 138 segment DNA-S, complete sequence	942	942	100%	0.0	99.81%	JQ820461.1
Banana bunchy top virus isolate Rwanda 142 segment DNA-S, complete sequence	942	942	100%	0.0	99.81%	JQ820467.1
Banana bunchy top virus isolate Lucknow segment DNA 3, complete sequence	942	942	100%	0.0	99.81%	EF687856.1
Banana bunchy top virus coat protein (cp) gene, complete cds	942	942	100%	0.0	99.81%	AY534140.1
Banana bunchy top virus DNA-3 isolate Burundi coat protein gene, complete cds	942	942	100%	0.0	99.81%	AF148943.1
Banana bunchy top virus isolate AS_JOR_B5 segment S, complete sequence	937	937	100%	0.0	99.61%	KX592198.2
Banana bunchy top virus segment DNA-S, complete sequence	937	937	100%	0.0	99.61%	KU559331.1
Banana bunchy top virus isolate BU11_CD_2012 segment S, complete sequence	937	937	100%	0.0	99.61%	KM607492.1
Banana bunchy top virus isolate 2557_AU_2010 segment S, complete sequence	937	937	100%	0.0	99.61%	KM607442.1
Banana bunchy top virus isolate Bangalore (GKVK) coat protein (S) gene, complete cds	937	937	100%	0.0	99.61%	JN250599.1
Banana bunchy top virus isolate Bangalore-GKVK segment DNA3, complete sequence	937	937	100%	0.0	99.61%	JN243753.1

Fig 4. NCBI BLAST to identify the gene. Graphical summary and descriptions of BBTV CP gene blasted against NCBI database

4.2.5 Phylogenetic analysis of BBTV based on coat protein nucleotide and amino acid sequences

Phylogeny was interpreted using Neighbour-Joining method and a tree was constructed based on coat protein gene sequences separately for Kerala isolates (Fig 5A) and all the isolates considered in this study (Fig 5B).

Fig 5A. Phylogenetic tree of BBTV Kerala isolates based on CP gene. The sum of branch length equal to 0.04722093 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches

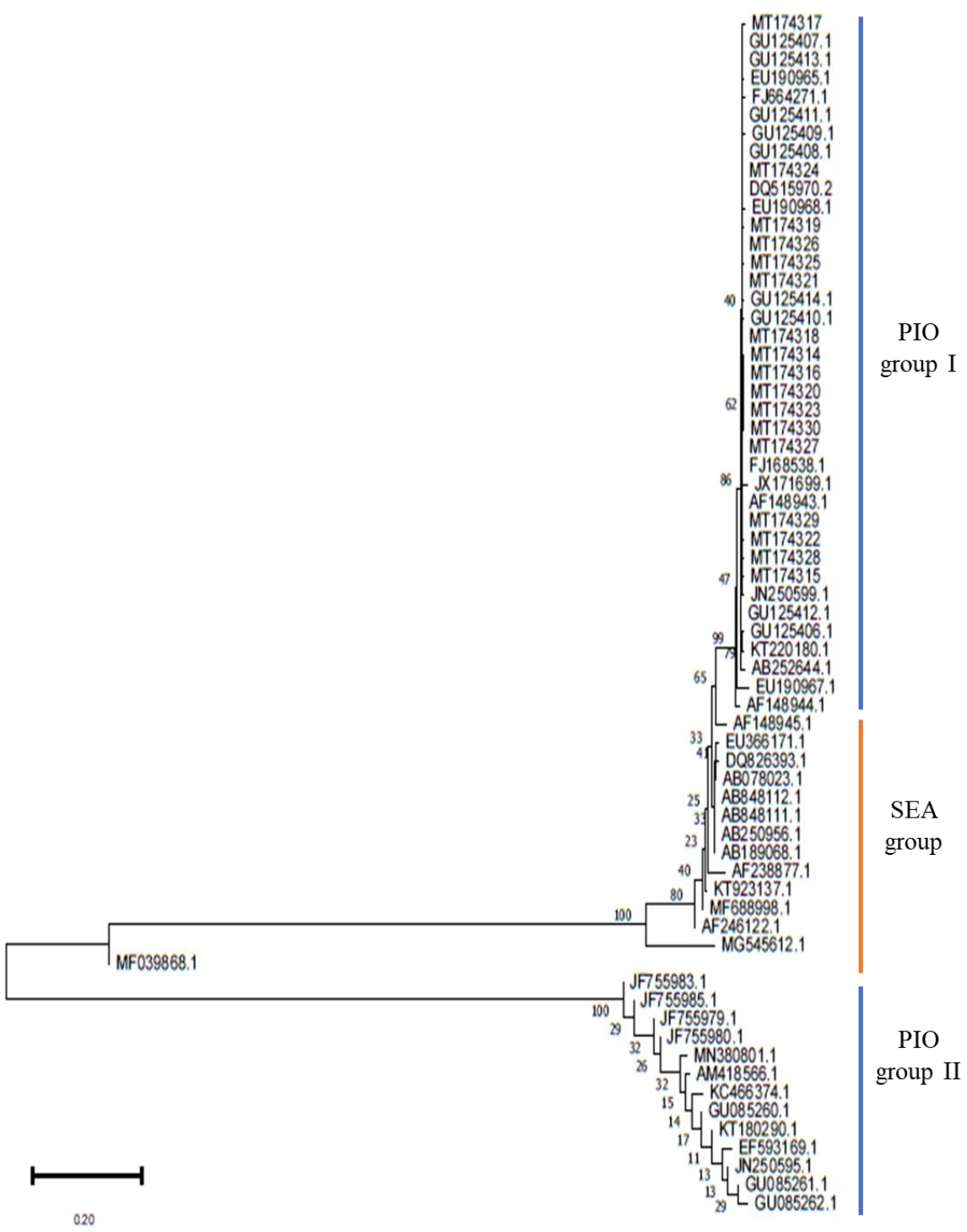


Fig 5B. Phylogenetic tree of all the BBTV Kerala isolates in the dataset. The sum of branch length equal to 3.01480964 is shown. The bootstrap values are shown next to the branches

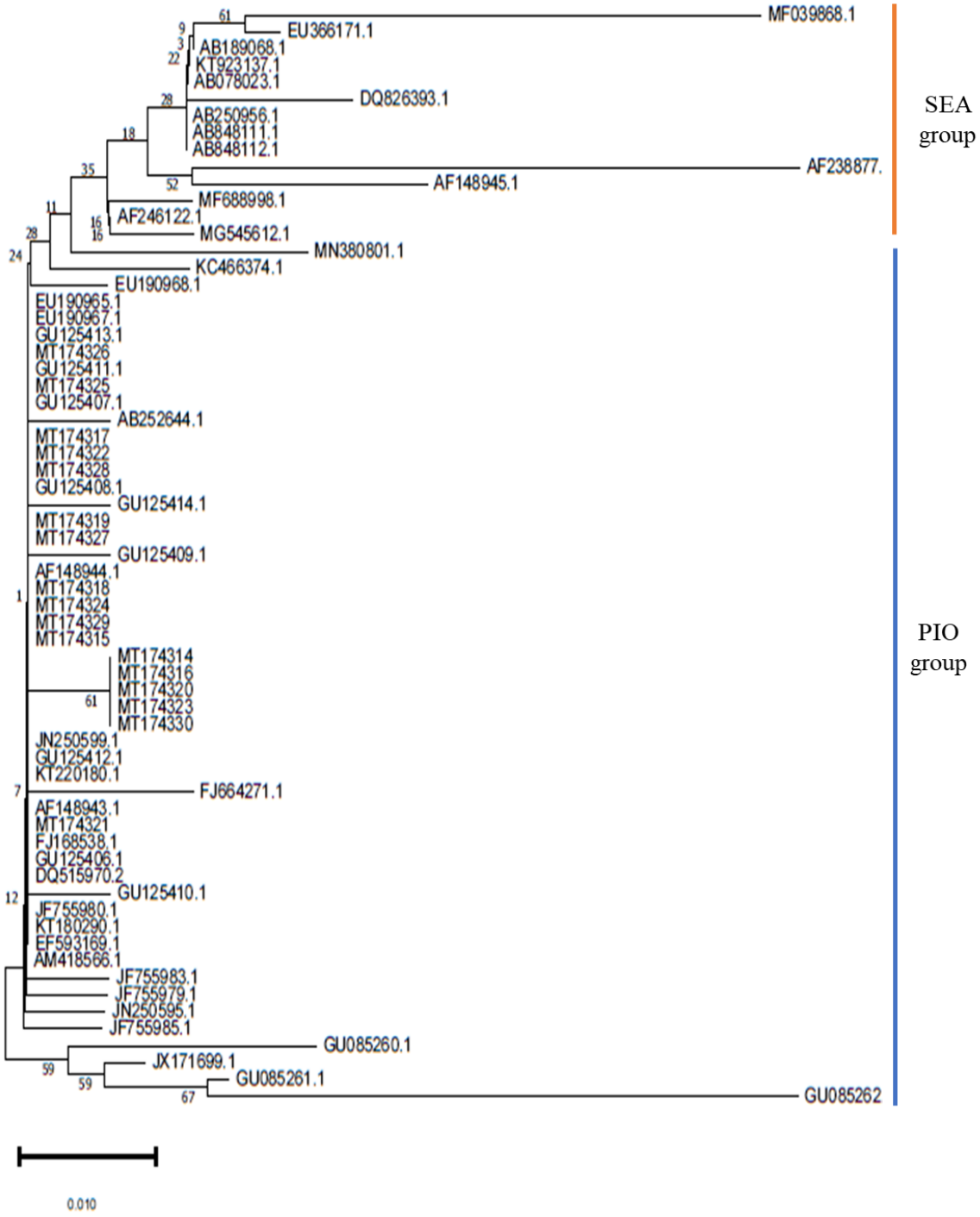


Fig 5C. Phylogenetic tree based on amino acid sequences of all BBTV isolates in the dataset. The sum of branch length equal to 0.325 is shown. The bootstrap values are shown next to the branches. There were total of 196 positions in the final dataset

From the phylogenetic tree of the Kerala isolates, grouping based on geography was observed with few exceptions. There was a very prominent demarcation of Southern and Northern isolates whereas those collected from Central zone paired with the isolates collected from both North and South. The isolates collected from Northern Kerala *viz.*, MT174317, MT174328, MT174315, MT174319, MT174314 and MT174318 clustered together with MT174321 collected from Thrissur, Central Kerala. Alternatively, isolates collected from Central zone (MT174322, MT174325, MT174323, MT174326, MT174327) clustered together. However, MT174329 collected from Kottayam clustered with the isolates from Central zone of Kerala. Three isolates *viz.*, MT174330 (Kollam), MT174316 and MT174320 (Wayanad) formed outgroups and clustered together. From the branch lengths, it was evident that accessions MT174314 and MT174317 from Kannur and Wayanad respectively are the most divergent among the group (Fig 5A).

All the isolates from India, Sri Lanka, Myanmar, Egypt, Tonga and Fiji belonging to Pacific Indian Oceans group clustered together (PIO group). The South East Asian isolates consisting of isolates from China, Taiwan, Indonesia, Philippines, Japan and Vietnam clustered together. The isolate from Thailand also clustered with these isolates however, was separated from other SEA isolates (SEA group). Few isolates from India (Tripura, Meghalaya, Tamil Nadu), Pakistan and Africa (Nigeria, Angola, Cameroon and Malawi) clustered separately into another monophyletic group (PIO group II) (Fig 5B).

All the seventeen Kerala isolates generated in this study grouped in the biggest subcluster (subcluster Ia) in PIO group I with other Indian isolates deposited from states like Tamil Nadu, Andhra Pradesh, Karnataka, Maharashtra, Gujarat, Bihar, Uttar Pradesh, New Delhi, West Bengal, Arunachal Pradesh and Assam apart from Kerala. Besides Indian isolates, an African isolate (AF148943.1), an Egyptian isolate (KT220180.1) and a Myanmarese isolate (AB252644.1) were also grouped in the same subcluster. Accessions, EU190967.1 (Meghalaya) and AF148944.1 (Fiji) also belonged to PIO group I, but in a different subcluster (subcluster Ib). From the branch lengths it can be inferred that the isolates in subcluster Ia of PIO group I, may have evolved simultaneously.

In contrast to the tree constructed using nucleotide sequences of CP gene, the phylogenetic tree based on amino acid sequences (Fig 5C) had only two major clusters. The evolutionary pattern of isolates belonging to PIO group and SEA group was similar to that depicted in

the former phylogenetic tree based. Except three Kerala isolates (MT174321, MT174325, MT174326) all others clustered together in PIO group.

Although, major clusters were as observed in phylogenetic tree based on nucleotide sequences, subclusters were different in the tree based on amino acid sequences. However, BBTV isolate from Thailand (MF039868.1) was seen to be the most divergent amongst the dataset considered for the present study in both the evolutionary trees. Interestingly, there was no noticeable phylogenetic variance of BBTV infecting abaca (AB250956.1) collected from Philippines, with other BBTV isolates infecting banana isolates as it clustered together in SEA group, indicating divergence from a common ancestor.

4.2.6 Principal co-ordinate analysis

The results obtained in Principal co-ordinate analysis (PCoA) was consistent with the phylogeny of the isolates considered in the present study (Fig 6).

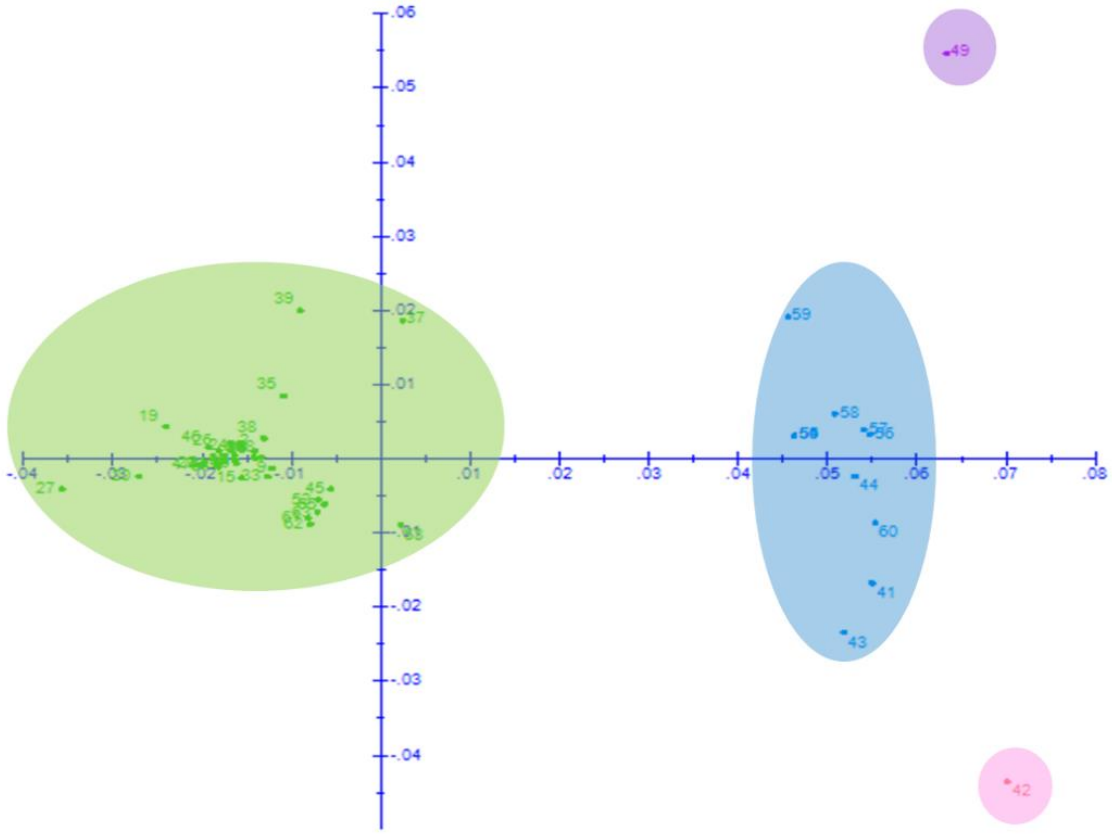


Fig 6. Principal co-ordinate analysis of BBTV isolates based on the nucleotide sequences of CP gene

There were four major clusters comprising of two major groups corresponding to PIO and SEA isolates and two clusters with only one isolate each. MF039868.1 isolate from Thailand (49) and AF238877.1 China (42) formed other two groups. These two isolates belonged to SEA group as shown in the phylogenetic analysis.

From the PCoA analysis, clustering based on geography was unambiguous. Isolate infecting abaca collected from Philippines grouped together with other SEA isolates. However, In PIO group isolates MN380801.1 (37), AM418566.1 (53), EU190967.1 (35), KC466374.1 (39) collected from Tripura, Meghalaya and Pakistan respectively were seen to be distinct from other isolates in the major cluster encompassing of PIO isolates.

4.2.7 Assessment of genetic variability of BBTV based on coat protein gene

4.2.7.1 Nucleotide and amino acid sequence homology of BBTV isolates

The nucleotide and amino acid sequences of the coat protein gene of all the seventeen Kerala isolates were aligned using MUSCLE and sequence identity was calculated in pairs to obtain an identity matrix (Table 5). The nucleotide sequence identity ranged from 99.2-100 %. Majority of sequence pairs were 99.8-99.6 percent identical. Most variable among all the isolates was MT174317 isolated from banana cv. Nendran collected from Wayanad district.

Amino acid sequence identity also displayed similar trend with identity ranging from 98-100 % (Fig 7). However, based on the amino acid sequence the most diverse isolate was MT174314, with 1.2- 1.8 percent amino acid sequence variability. This isolate was collected from an infected Nendran variety in Kannur district.

On comparing the sequence similarity of all the nucleotide sequences in the entire dataset, it was evident that the pairwise nucleotide identity ranged from 90- 100 % (Fig 8). The SEA as well as PIO isolates were included in the data set for comparison. Isolate from Maharashtra with and accession number GU085262.1 belonging to PIO and that from China, AF238877.1 a member of SEA group exhibited maximum diversity. Based on only the coding region of coat protein gene, SEA and PIO isolates were 4-10 per cent diverse. However, diversity among the isolates in these two groups were comparatively less.

Table 5. Homology matrix of BBTV Kerala isolates colour coded to differentiate most similar and dissimilar sequence pairs

	MT174314	MT174315	MT174316	MT174317	MT174318	MT174319	MT174320	MT174321	MT174322	MT174323	MT174324	MT174325	MT174326	MT174327	MT174328	MT174329	MT174330
MT174314	100.00	99.60	100.00	99.22	99.80	99.60	100.00	99.80	99.60	100.00	99.80	99.60	99.80	99.80	99.60	99.80	100.00
MT174315	99.60	100.00	99.60	99.20	99.80	99.61	99.60	99.80	99.61	99.60	99.80	99.60	99.80	99.80	99.61	99.80	99.60
MT174316	100.00	99.60	100.00	99.20	99.80	99.60	100.00	99.80	99.60	100.00	99.80	99.60	99.80	99.80	99.60	99.80	100.00
MT174317	99.22	99.20	99.20	100.00	99.40	99.20	99.20	99.40	99.22	99.22	99.40	99.20	99.40	99.40	99.22	99.40	99.22
MT174318	99.80	99.80	99.80	99.40	100.00	99.80	99.80	100.00	99.80	99.80	100.00	99.80	100.00	100.00	99.80	100.00	99.80
MT174319	99.60	99.61	99.60	99.20	99.80	100.00	99.60	99.80	99.60	99.60	99.80	99.60	99.80	99.80	99.60	99.80	99.60
MT174320	100.00	99.60	100.00	99.20	99.80	99.60	100.00	99.80	99.61	100.00	99.80	99.60	99.80	99.80	99.60	99.80	100.00
MT174321	99.80	99.80	99.80	99.40	100.00	99.80	99.80	100.00	99.80	99.80	100.00	99.80	100.00	100.00	99.80	100.00	99.80
MT174322	99.60	99.61	99.60	99.22	99.80	99.60	99.61	99.80	100.00	99.60	99.80	99.60	99.80	99.80	99.60	99.80	99.60
MT174323	100.00	99.60	100.00	99.22	99.80	99.60	100.00	99.80	99.60	100.00	99.80	99.60	99.80	99.80	99.60	99.80	100.00
MT174324	99.80	99.80	99.80	99.40	100.00	99.80	99.80	100.00	99.80	99.80	100.00	99.80	100.00	100.00	99.80	100.00	99.80
MT174325	99.60	99.60	99.60	99.20	99.80	99.60	99.60	99.80	99.60	99.60	99.80	100.00	99.80	99.80	99.60	99.80	99.60
MT174326	99.80	99.80	99.80	99.40	100.00	99.80	99.80	100.00	99.80	99.80	100.00	99.80	100.00	100.00	99.80	100.00	99.80
MT174327	99.80	99.80	99.80	99.40	100.00	99.80	99.80	100.00	99.80	99.80	100.00	99.80	100.00	100.00	99.80	100.00	99.80
MT174328	99.60	99.61	99.60	99.22	99.80	99.60	99.60	99.80	99.60	99.60	99.80	99.60	99.80	99.80	100.00	99.80	99.60
MT174329	99.80	99.80	99.80	99.40	100.00	99.80	99.80	100.00	99.80	99.80	100.00	99.80	100.00	100.00	99.80	100.00	99.80
MT174330	100.00	99.60	100.00	99.22	99.80	99.60	100.00	99.80	99.60	100.00	99.80	99.60	99.80	99.80	99.60	99.80	100.00

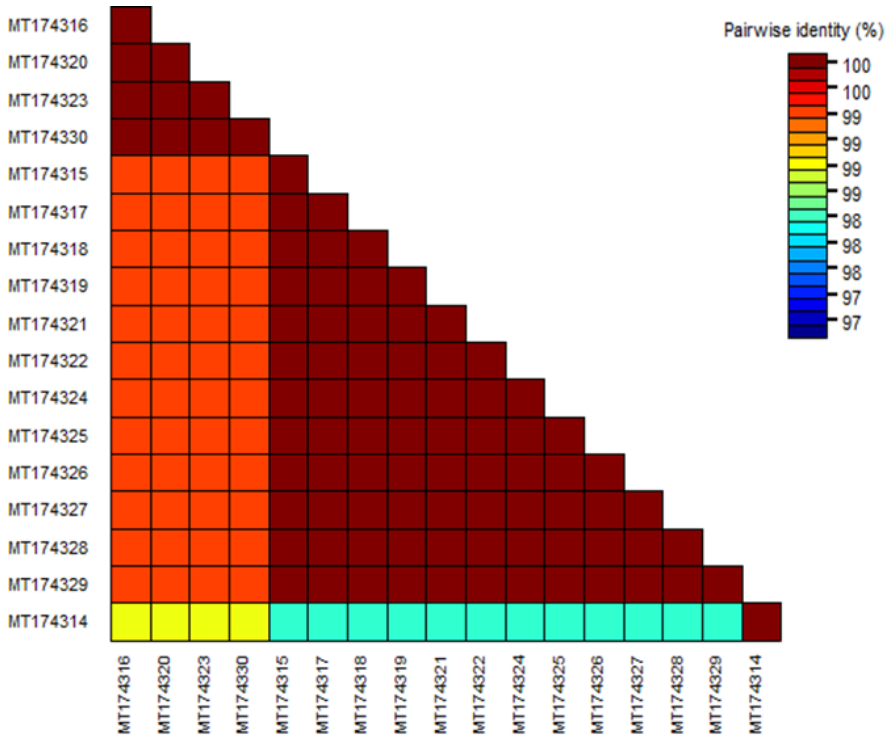


Fig 7. Amino acid sequence identity heat map denoting sequence identity of BBTV isolates

4.2.7.2 Nucleotide diversity and mutation rate

Parameters like nucleotide diversity, Pi (π) and Watterson estimation or mutation rate (θ) parameters are measures of mutation rate in a population and signify the genetic diversity in the sequence. Nucleotide diversity of CP gene of BBTV was estimated to be 0.03246 indicating a very low diversity and higher stability of the gene (Fig 9). Pi calculated for nucleotide positions ranged from 0.04-.078. There was no particular region in the gene that was highly variable compared to other regions. However, maximum variability was observed between nucleotide positions 286-310 and minimum between 166-190. Mutation rate, theta (θ) indicated high mutation rate at nucleotide positions 41-65, 286-310, 499-523 and 574-591. A total of 135 mutations and 111 polymorphic sites on coding region of coat protein were detected from the analysis. Evidently, analysis indicated low variability in the coding region of the coat protein gene of BBTV.

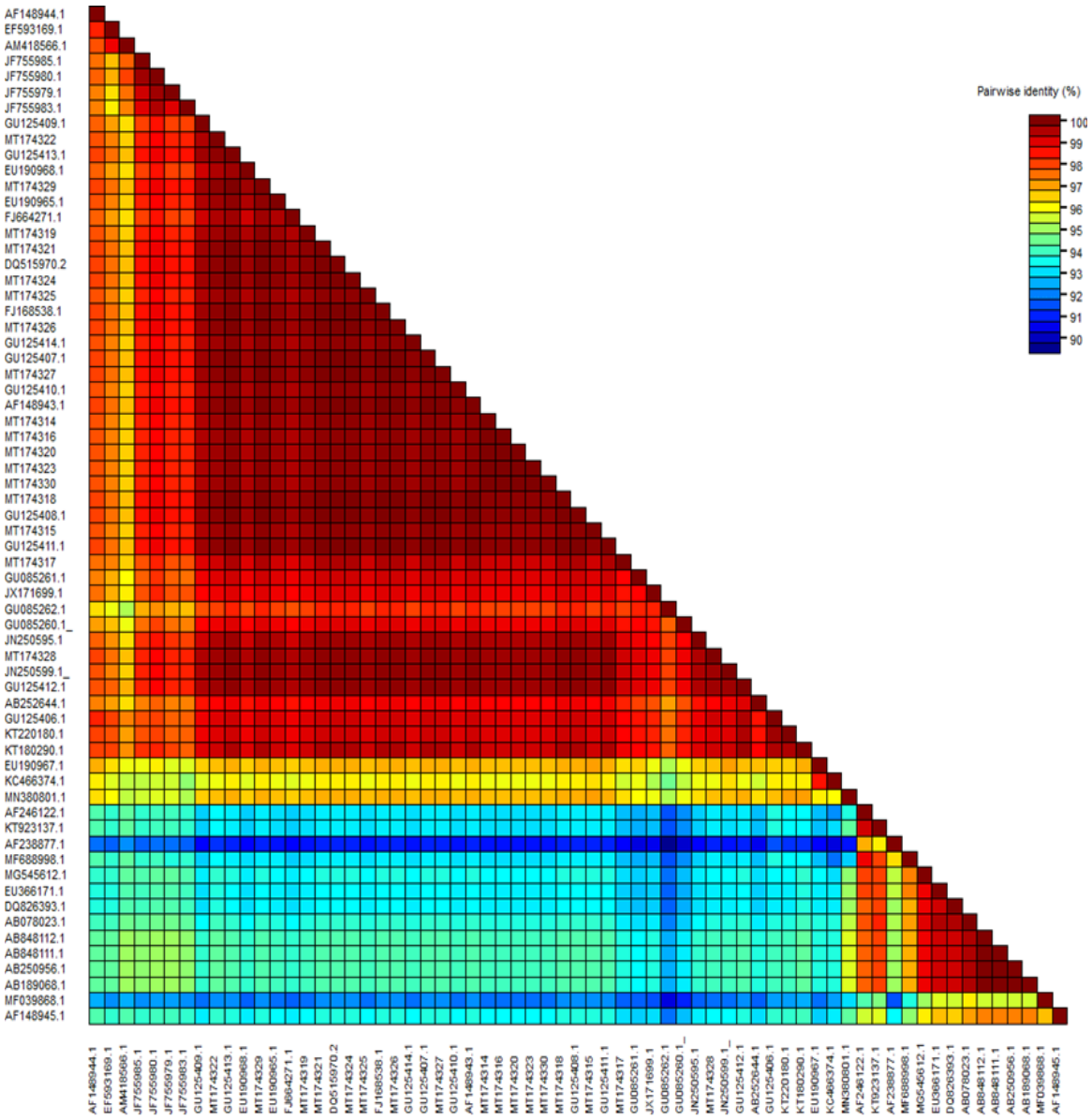


Fig 8. Pairwise sequence identity heat map denoting sequence identity of coat protein gene of all the Kerala BBTV isolates and NCBI accessions

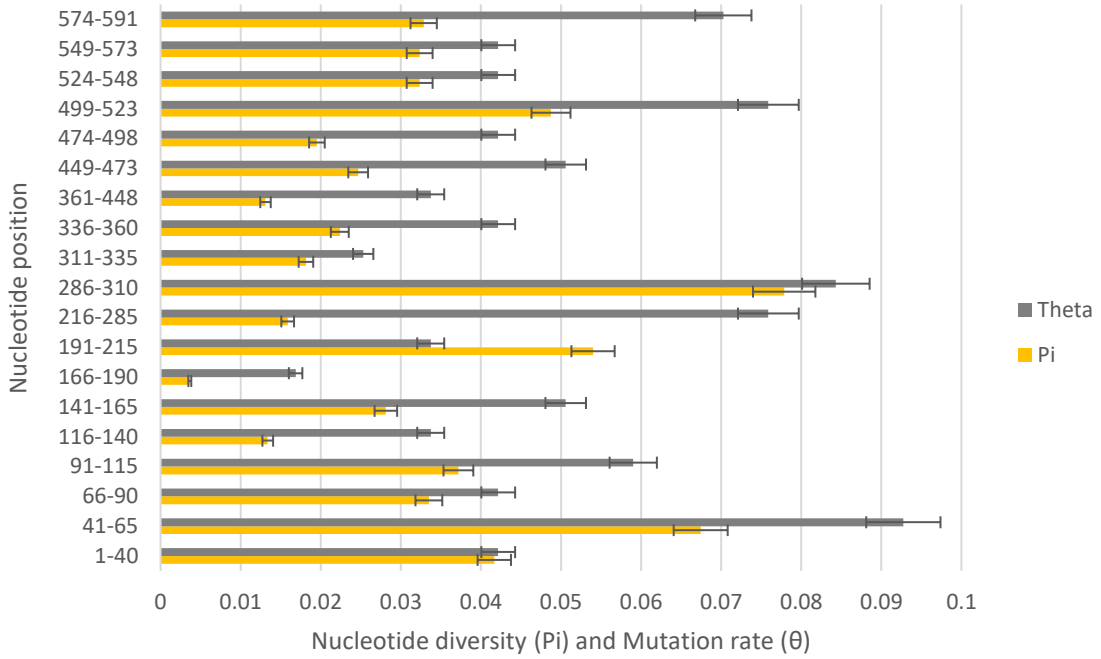


Fig 9. Nucleotide diversity and mutation rate estimated based on total number mutations per site. The result is interpreted in window size 25. Error bars at 5 % are displayed on the graph.

Out of the total number of mutations, 123 were detected as frameshift mutations caused by insertion or deletion (In-Del). A major deletion of 45 nucleotides at the 5' end and an insertion of 60 nucleotides towards the 3' end of CP gene of SEA isolate collected from Thailand (MF039868.1) was observed making it most distinct among the dataset.

Pi value of the Kerala isolates estimated separately was 0.0027 with only nine polymorphic sites and mutations each in the seventeen sequences. The value of θ was calculated to be 0.0052 which was much lesser than the average of the total dataset. Although nucleotide variation was negligible, variations were observed between nucleotide positions 201 and 325. Apart from the core region of the CP, regions between 26-50 and 351-375 was also seen to be variable, considering π and θ values (Table 6).

DNA divergence was calculated for population subsets and it was evident that divergence among and between the isolates collected from North, Central and South zones of Kerala were negligible. However, diversity among the SEA isolates were greater than PIO isolates. Although no particular region in the gene was identified to be hypervariable, net nucleotide substitutions were higher in the 5' and 3' terminus of the gene followed by the core region (Fig 10).

Table 6. Nucleotide diversity and mutation rate of coat protein gene of BBTV Kerala isolates

Nucleotide position	Midpoint	Pi (π)	Theta (θ)	Nucleotide position	Midpoint	Pi (π)	Theta (θ)
1-25	13	0.000	0.000	276-300	288	0.022	0.024
26-50	38	0.009	0.024	301-325	313	0.005	0.012
51-75	63	0.000	0.000	326-350	338	0.000	0.000
76-100	88	0.000	0.000	351-375	363	0.005	0.012
101-125	113	0.000	0.000	376-400	388	0.000	0.000
126-150	138	0.000	0.000	401-425	413	0.000	0.000
151-175	163	0.000	0.000	426-450	438	0.000	0.000
176-200	188	0.000	0.000	451-475	463	0.000	0.000
201-225	213	0.005	0.012	476-500	488	0.000	0.000
226-250	238	0.005	0.012	501-513	507	0.000	0.000
251-275	263	0.005	0.012				

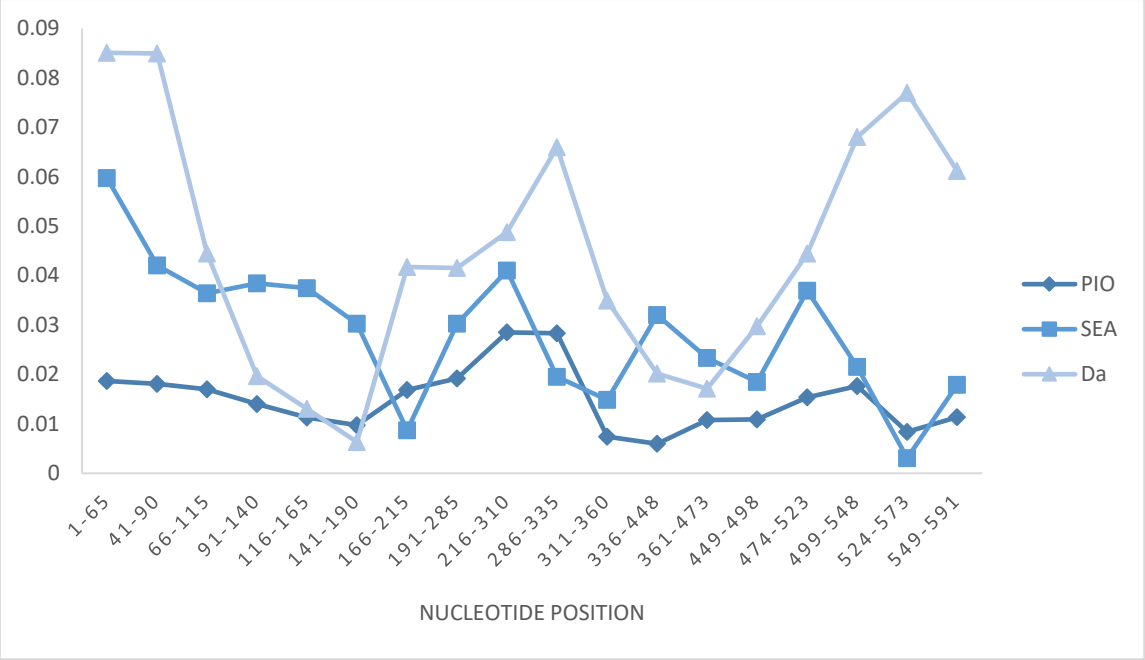


Fig 10. DNA divergence between and among PIO and SEA populations. Number of net nucleotide substitutions per site between populations, Da is also represented

4.2.7.3 Synonymous and non-synonymous substitutions

Pairwise rate of non-synonymous (Ka) and synonymous (Ks) mutations per nucleotide site were calculated and ratio between rate of non-synonymous and synonymous mutations (Ka/Ks) were obtained. Most of the pairwise Ka/Ks ratio (ω) were less than 1, signifying purifying or negative selection. However, ω of few sequence pairs the PIO isolates were more than 1 indicating positive or Darwinian selection (Table 7).

Positive selection (ω >1) or neutral selection (ω ~1) was observed between isolates with accession numbers GU085260.1, GU085261.1, JX171699.1 (Tamil Nadu), GU085262.1 (Maharashtra) and Kerala isolates generated in the present study. Nevertheless, between the Kerala isolates, ω values were less than 1 indicative of stabilising selection.

Table 7. Ratio of non-synonymous and synonymous mutation in CP gene sequences between PIO isolates

Sequence 1	Sequence 2	Ka/Ks	Sequence 1	Sequence 2	Ka/Ks
GU085261.1	MT174314	1.253	GU085262.1	MT174315	1.245
	MT174316	1.253		MT174317	1.245
	MT174318	0.934		MT174318	2.516
	MT174320	1.253		MT174319	1.245
	MT174321	0.934		MT174320	2.835
	MT174322	0.934		MT174321	2.516
	MT174323	1.253		MT174322	2.516
	MT174324	0.934		MT174323	2.835
	MT174326	0.934		MT174324	2.516
	MT174327	0.934		MT174325	1.245
	MT174329	0.934		MT174326	2.516
	MT174330	1.253		MT174327	2.516
GU085260.1	MT174314	1.560		MT174328	1.245
	MT174316	1.560		MT174329	2.516
	MT174318	1.253		MT174330	2.835
	MT174320	1.560	JX171699.1	MT174314	1.253
	MT174321	1.253		MT174316	2.835
	MT174322	1.253		MT174321	0.934
	MT174323	1.560		MT174322	0.934
	MT174324	1.253		MT174323	1.253
	MT174326	1.253		MT174324	0.934
	MT174327	1.253		MT174326	0.934
	MT174329	1.253		MT174327	0.934
	MT174330	1.560		MT174329	0.934
GU085262.1	MT174314	2.835		MT174330	1.253

4.2.7.4 Gene flow between population subsets

Gene flow between the population subsets were concluded based on the fixation index (F_{ST}) *ie.,* population differentiation. As gene flow reduces the genetic drift, population differentiation will be insignificant. This trend was seen between the population subsets of Kerala (Table 8).

There were variations in the G_{ST}, N_{ST} (genetic differentiation indices) and F_{ST} values of the subpopulations. However, Gene flow between PIO and SEA isolates were observed to the least. This implied genetic differentiation between these group. The average (Dxy) and net (Da) nucleotide substitutions per site were also highest (0.067 and 0.045 respectively) between PIO and SEA subpopulation.

Evidently, high gene flow was observed between Northern, Central and Southern isolates of Kerala with very low F_{ST} and N_{ST} values indicative of minimal genetic differentiation between the subpopulations. Considering the phylogeography of the BBTV Kerala isolates, it is evident that, the clustering based on geography has no strict delineation between the subpopulations. Isolates from Central zone of Kerala clustered with BBTV strains collected from both Northern and Southern zones.

Interestingly, there was evident gene flow between all the subpopulations in India, especially between the South Indian states *viz.,* Kerala and Karnataka. The gene flow between Kerala and Tamil Nadu is also evident from the analysis. From the survey it was learnt that most of the planting materials for banana cultivation in Kerala are procured from Tamil Nadu substantiates the above result.

Statistical significance of nucleotide diversity analysis was carried out by three tests. Tajima's D as well as Fu and Li's D and F tests gave negative values implying the presence of rare alleles in the population. Tajima's D test (-1.620) was statistically non-significant in contrast to highly significant Fu and Li's D and F tests (-3.714 and -3.458).

However, Tajima's D test at non-synonymous sites were highly significant negative value (-2.459). Codons with multiple evolutionary paths were detected in the analysis denoting the PIO and SEA isolates present in the dataset.

Table 8. Gene flow and differentiation between the population subsets

Population 1	Population 2	Hs	Ks	G_{ST}	N_{ST}	F_{ST}	Dxy	Da
North Kerala	Central Kerala	0.740	1.333	0.067	0.025	0.025	0.003	0.000
North Kerala	South Kerala	0.857	1.778	0.035	0.235	0.235	0.003	-0.001
Central Kerala	South Kerala	0.643	0.800	0.033	0.166	0.167	0.002	0.000
Kerala	Tamil Nadu	0.779	2.166	0.072	0.127	0.128	0.009	0.001
Kerala	Karnataka	0.752	1.182	0.062	0.056	0.056	0.002	0.000
Tamil Nadu	Karnataka	1.000	4.667	0.020	0.116	0.118	0.009	0.001
South Indian	North India	0.843	2.412	0.020	0.016	0.016	0.006	0.000
South Indian	North East India	0.891	4.441	0.038	0.185	0.184	0.020	0.004
North India	North East India	0.857	8.071	0.077	0.183	0.181	0.021	0.004
PIO isolates	SEA isolates	0.924	8.355	0.034	0.677	0.670	0.067	0.045

Hs denotes the genetic diversity within the population, **Ks** denotes the rate of synonymous substitutions per site, genetic differentiation is denoted by N_{ST} and G_{ST} values and gene flow is denoted by fixation index, F_{ST}. **Dxy** is the average number of nucleotide substitutions per site between populations and **Da** denotes the number of net nucleotide substitutions per site between populations. All the values were estimated using the DnaSP6 program with 1,000 permutation tests

4.2.8 Codon usage bias

Codon usage bias of CP gene of BBTV was calculated in CAI-Cal web server. The overall nucleotide composition (% A, adenine; % T/U, thymine/uracil; % C, cytosine; % G, guanine) of CP gene was analysed (Table 9). The percentage values of A (30.626 ± 0.512) and T (26.621 ± 0.440) were highest followed by G (25.497 ± 0.340) and C (17.256 ± 0.548). Both PIO and SEA isolates were both AT rich. Composition of A (30.428 ± 0.345, 31.347 ± 0.348), T (26.680 ± 0.228; 26.408 ± 0.807), G (25.458 ± 0.270; 25.639 ± 0.493) and C (17.434 ± 0.327; 16.606 ± 0.682) of PIO and SEA separately was similar to the overall trend. Mean value of % AT was 57.247 ± 0.579 whereas that of % GC was 42.753 ± 0.579.

Apparently, the CP gene of BBTV which is AT rich is most likely to have A or T residues in the wobble position (3rd position of the synonymous codon). To the contrary, the results showed that G or C at the wobble position was preferred. Percentage of G at wobble position (% G3) was 29.159 ± 0.893 followed by % C3 (28.442 ± 1.118), % A3 (26.656 ± 1.181) and % T3 (15.743 ± 0.942). Interestingly, % GT3 (57.601 ±0.769) is higher than % GA3 (55.815 ± 1.392), % AT3 (55.098 ± 1.342) and % GC3 (44.902 ± 1.342). This unusual variation from the hypothesis implies that the compositional patterns of BBTV CP gene is complex and is anomalous from what is commonly observed in most of the AU and or GC rich virus genes.

The effective number of codons (Nc) in the dataset was ranging from 48.60-59.50 with an overall mean of 54.70 ± 2.65. The Nc value of PIO isolates ranged from 48.60 to 59.50 and that of SEA isolates were between 48.80 and 53.10. The values indicate high stability of the gene.

The CP is highly expressed gene of BBTV, and significant high codon usage bias (CUB) was observed from the relative synonymous codon usage (RSCU) value calculated. A total of 29 codons were observed to be desirable (UUU, UUA, UUG, CUG, CUU, AUC, GUU, GUA, GUG, AGU, AGC, CCU, CCC, ACA, GCU, GCA, UAU, CAU, CAG, AAC, AAG GAU, GAA, UGU, CGG, AGA, AGG, GGA, GGG) out of which 22 were identified as high frequency codons (Hfc) (Table 10).

Table 9. Nucleotide composition and effective number of codons of BBTV CP gene

Parameter / Sequences	Length	%A	%C	%U	%G	%G+C	%A+U	%GC1	%AU1	%GC2	%AU2	%GC3	%AU3	Nc
PIO isolates														
MT174314	513	30.21	17.74	26.71	25.34	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	54.50
MT174315	513	30.21	17.35	26.90	25.54	42.89	57.12	42.11	57.90	41.52	58.48	45.03	54.97	56.00
MT174316	513	30.21	17.74	26.71	25.34	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	54.50
MT174317	513	30.41	17.15	27.10	25.34	42.50	57.51	42.11	57.90	41.52	58.48	43.86	56.14	54.60
MT174318	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
MT174319	513	30.41	17.54	26.71	25.34	42.89	57.12	42.11	57.90	41.52	58.48	45.03	54.97	56.00
MT174320	513	30.21	17.74	26.71	25.34	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	54.50
MT174321	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
MT174322	513	30.41	17.54	26.71	25.34	42.89	57.12	42.11	57.90	41.52	58.48	45.03	54.97	56.10
MT174323	513	30.21	17.74	26.71	25.34	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	54.50
MT174324	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
MT174325	513	30.21	17.74	26.71	25.34	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.50
MT174326	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
MT174327	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
MT174328	513	30.41	17.54	26.71	25.34	42.89	57.12	42.11	57.90	41.52	58.48	45.03	54.97	56.50
MT174329	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
MT174330	513	30.21	17.74	26.71	25.34	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	54.50
GU125413.1	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
GU085261.1	528	31.06	17.61	26.14	25.19	42.80	57.20	42.05	57.96	40.91	59.09	45.46	54.55	56.40
GU085260.1	528	31.06	17.61	26.33	25.00	42.61	57.39	42.05	57.96	40.91	59.09	44.89	55.11	59.50

FJ664271.1	513	30.02	17.54	26.90	25.54	43.08	56.92	42.69	57.31	40.94	59.06	45.61	54.39	56.90
JX171699.1	513	30.60	17.15	26.90	25.34	42.50	57.51	42.11	57.90	40.94	59.06	44.44	55.56	56.10
GU125411.1	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
JN250599.1	513	30.02	17.54	26.90	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	55.70
GU125412.1	513	30.21	17.35	26.90	25.54	42.89	57.12	42.11	57.90	41.52	58.48	45.03	54.97	56.20
GU125409.1	513	30.02	17.74	26.51	25.73	43.47	56.53	42.11	57.90	42.11	57.90	46.20	53.80	55.90
GU085262.1	528	31.25	17.24	26.89	24.62	41.86	58.14	41.48	58.52	39.77	60.23	44.32	55.68	58.50
GU125408.1	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
DQ515970.2	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
GU125414.1	513	30.21	17.54	26.51	25.73	43.28	56.73	42.11	57.90	42.11	57.90	45.61	54.39	56.00
FJ168538.1	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
GU125407.1	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
GU125406.1	513	30.41	17.15	27.10	25.34	42.50	57.51	42.11	57.90	41.52	58.48	43.86	56.14	54.00
EU190968.1	513	30.60	17.54	26.51	25.34	42.89	57.12	42.11	57.90	41.52	58.48	45.03	54.97	56.20
EU190967.1	513	30.41	17.15	26.32	26.12	43.28	56.73	42.11	57.90	41.52	58.48	46.20	53.80	50.90
EU190965.1	513	30.21	17.74	26.51	25.54	43.28	56.73	42.11	57.90	41.52	58.48	46.20	53.80	55.50
MN380801.1	528	31.06	16.67	26.33	25.95	42.61	57.39	42.05	57.96	41.48	58.52	44.32	55.68	54.00
KT180290.1	528	30.87	17.24	26.89	25.00	42.24	57.77	42.05	57.96	40.91	59.09	43.75	56.25	57.00
KC466374.1	528	30.87	16.86	26.33	25.95	42.80	57.20	41.48	58.52	40.91	59.09	46.02	53.98	53.90
GU125410.1	513	30.21	17.54	26.51	25.73	43.28	56.73	42.11	57.90	42.11	57.90	45.61	54.39	56.60
AF148944.1	513	30.41	17.74	26.71	25.15	42.89	57.12	42.69	57.31	41.52	58.48	44.44	55.56	54.30
AB252644.1	513	30.41	17.35	26.71	25.54	42.89	57.12	42.69	57.31	41.52	58.48	44.44	55.56	54.80
JN250595.1	528	30.68	17.61	26.71	25.00	42.61	57.39	42.05	57.96	40.91	59.09	44.89	55.11	57.90
EF593169.1	528	30.68	17.24	27.27	24.81	42.05	57.96	42.61	57.39	40.91	59.09	42.61	57.39	57.10
AM418566.1	528	30.49	17.24	26.71	25.57	42.80	57.20	42.05	57.96	40.91	59.09	45.46	54.55	57.90

KT923137.1	513	31.77	15.79	27.10	25.34	41.13	58.87	40.35	59.65	41.52	58.48	41.52	58.48	48.60
JF755985.1	528	30.49	17.42	26.52	25.57	42.99	57.01	41.48	58.52	40.91	59.09	46.59	53.41	56.80
JF755983.1	528	30.49	17.05	26.52	25.95	42.99	57.01	42.05	57.96	40.91	59.09	46.02	53.98	57.50
JF755980.1	528	30.49	17.42	26.33	25.76	43.18	56.82	41.48	58.52	40.91	59.09	47.16	52.84	58.00
AF148943.1	513	30.21	17.54	26.71	25.54	43.08	56.92	42.11	57.90	41.52	58.48	45.61	54.39	56.00
JF755979.1	528	30.87	17.42	26.14	25.57	42.99	57.01	40.91	59.09	40.91	59.09	47.16	52.84	57.40
Average		**30.43**	**17.43**	**26.68**	**25.46**	**42.89**	**57.11**	**42.03**	**57.97**	**41.37**	**58.64**	**45.28**	**54.72**	**55.80**
SD		**0.35**	**0.33**	**0.23**	**0.27**	**0.39**	**0.39**	**0.37**	**0.37**	**0.39**	**0.39**	**0.96**	**0.96**	**1.70**
SEA isolates														
AB250956.1	513	31.38	16.57	26.12	25.93	42.50	57.51	41.52	58.48	41.52	58.48	44.44	55.56	50.30
AB189068.1	513	31.38	16.57	26.12	25.93	42.50	57.51	41.52	58.48	41.52	58.48	44.44	55.56	50.30
EU366171.1	513	31.38	16.18	26.71	25.73	41.91	58.09	40.35	59.65	41.52	58.48	43.86	56.14	50.60
DQ826393.1	513	31.19	16.18	26.32	26.32	42.50	57.51	40.94	59.06	42.69	57.31	43.86	56.14	51.60
AB078023.1	513	31.77	16.18	26.32	25.73	41.91	58.09	40.94	59.06	41.52	58.48	43.28	56.73	50.20
AF148945.1	513	30.80	16.57	26.51	26.12	42.69	57.31	42.11	57.90	41.52	58.48	44.44	55.56	52.80
AF246122.1	513	31.77	15.98	27.10	25.15	41.13	58.87	40.35	59.65	41.52	58.48	41.52	58.48	49.00
AF238877.1	513	31.38	16.57	27.68	24.37	40.94	59.06	40.94	59.06	42.11	57.90	39.77	60.23	50.80
MF688998.1	513	31.58	16.57	26.90	24.95	41.52	58.48	40.94	59.06	42.11	57.90	41.52	58.48	48.80
MG545612.1	516	31.59	16.28	26.36	25.78	42.05	57.95	41.86	58.14	41.28	58.72	43.02	56.98	49.20
KT220180.1	513	30.41	16.76	27.29	25.54	42.30	57.70	41.52	58.48	41.52	58.48	43.86	56.14	53.10
MF039868.1	528	31.44	18.94	24.05	25.57	44.51	55.49	43.75	56.25	43.18	56.82	46.59	53.41	52.60
AB848112.1	513	31.38	16.57	26.12	25.93	42.50	57.51	41.52	58.48	41.52	58.48	44.44	55.56	50.30
AB848111.1	513	31.38	16.57	26.12	25.93	42.50	57.51	41.52	58.48	41.52	58.48	44.44	55.56	50.30
Average		**31.35**	**16.61**	**26.41**	**25.64**	**42.25**	**57.76**	**41.41**	**58.59**	**41.79**	**58.21**	**43.54**	**56.46**	**50.71**
SD		**0.35**	**0.68**	**0.81**	**0.49**	**0.82**	**0.82**	**0.82**	**0.82**	**0.53**	**0.53**	**1.61**	**1.61**	**1.32**

The wobble position of 17 of the 22 Hfc were ending in A or U indicating that the overall nucleotide composition of high importance in codon usage bias of a highly expressed gene.

The correlation analysis of various nucleotide compositions in BBTV CP gene and the effective number of codons (Nc), was successful in establishing a strong positive correlation between Nc and % C, % T/U, GC content at first position of the codon (GC1) and GC content at the wobble position of the codon (GC3) (Table 11). A significant negative correlation with % A, % G, % AU, % GA, GA3 and AU3 (AU content at the wobble position) was also observed. The analysis confirms statistically that CUB is greatly dependent on the nucleotide composition of the CP gene in BBTV.

The Nc plot (Fig 11) and neutrality plot (Fig 12) illustrated the effect of selection pressure and mutation on the CUB of BBTV CP gene. The magnitude of impact of mutational pressure and natural selection on CUB of BBTV CP was investigated by neutrality plot. In the Nc plot all the PIO and SEA isolates were below the standard curve indicating the equal contribution of mutation and selection pressure on the codon usage of the virus coat protein and a major role of overall nucleotide composition in CUB. If selection pressure was major driving force for the codon stability, the points would have been on the standard curve. This indicated why certain synonymous codons are selected over the other by process of mutation and selection in order to obtain a high stability of the gene as indicated by the Nc value.

The slope (0.0197) of the regression line was close to zero and the range of GC3 was narrow ranging from 0.398 to 0.472. The neutrality plot emphasised the role of natural selection on codon usage bias.

Table 10. Relative Synonymous Codon Usage (RSCU) of BBTV CP. High frequency codons are indicated by (*) next to RSCU value. Desirable codons are indicated by (#) next to the codon.

Amino Acid	Codon	RSCU	Amino Acid	Codon	RSCU
Phenylalanine	UUU#	1.614*	Asparagine	AAU	0.472
	UUC	0.386		AAC#	1.528*
Leucine	UUA#	1.420*	Lysine	AAA	0.674
	UUG#	1.596*		AAG#	1.326*
	CUU#	1.713*	Aspartic acid	GAU#	1.492*
	CUC	0.022		GAC	0.508
	CUA	0.068	Glutamic acid	GAA#	1.564*
	CUG#	1.181		GAG	0.436
Isoleucine	AUU	0.995	Cysteine	UGU#	1.502*
	AUC#	1.107		UGC	0.498
	AUA	0.898	Arginine	CGU	0.039
Valine	GUU#	1.344*		CGC	0.395
	GUC	0.525		CGA	0.015
	GUA#	1.057		CGG#	1.406*
	GUG#	1.074		AGA#	1.701*
Serine	UCU	0.777		AGG#	2.445*
	UCC	0.337	Glycine	GGU	0.486
	UCA	0.900		GGC	0.420
	UCG	0.957		GGA#	1.514*
	AGU#	1.807*		GGG#	1.579*
	AGC#	1.222	Alanine	GCU#	1.011
Proline	CCU#	1.489*		GCC	0.305
	CCC#	1.023		GCA#	2.104*
	CCA	0.497		GCG	0.580
	CCG	0.990	Tyrosine	UAU#	1.501*
Tryptophan	ACU	0.415		UAC	0.499
	ACC	0.907	Histidine	CAU#	1.469*
	ACA#	2.390*		CAC	0.531
	ACG	0.288	Glutamine	CAA	0.207
				CAG#	1.793*

Table 11. Correlation of nucleotides compositions and effective number of codons of BBTV coat protein

	%A	%C	%U	%G	%GC1	%GC2	%GC12	%GC3	Nc	%AU	%GA	%GU3	%GA3	%AU3
%A	1.000													
%C	-0.702*	1.000												
%U	-0.267*	-0.262	1.000											
%G	-0.031	-0.215	-0.469*	1.000										
%GC1	-0.597	0.798*	-0.282*	-0.021	1.000									
%GC2	0.032	0.061	-0.320*	0.268	0.105*	1.000								
%GC12	-0.415	0.627*	-0.410	0.146	0.801*	0.678	1.000							
%GC3	-0.642	0.704*	-0.450	0.413*	0.504*	-0.105	0.311*	1.000						
Nc	-0.685**	0.704**	0.112**	-0.247*	0.515*	-0.441	0.112	0.607*	1.000					
%AU	0.682*	-0.820**	0.523**	-0.382**	-0.742	-0.215	-0.678*	-0.909	-0.520**	1.000				
%GA	0.828*	-0.714**	-0.489**	0.535**	-0.516	0.177	-0.269*	-0.311	-0.718**	0.362	1.000			
%GU3	-0.569	-0.069	0.708	0.051*	-0.030	-0.172	-0.135	0.026*	0.186*	0.035*	-0.452	1.000		
%GA3	0.685	-0.550*	-0.603	0.634	-0.365	0.074*	-0.219	-0.063	-0.548**	0.148	0.934**	-0.449**	1.000	
%AU3	0.642*	-0.705*	0.450**	-0.413	-0.504*	0.105	-0.311	-1.000*	-0.607**	0.909**	0.311**	-0.026	0.063	1.000

* Correlation is significant at the 0.05 level;

** Correlation is significant at the 0.01 level (two-tailed). Significant values are highlighted with blue colour.

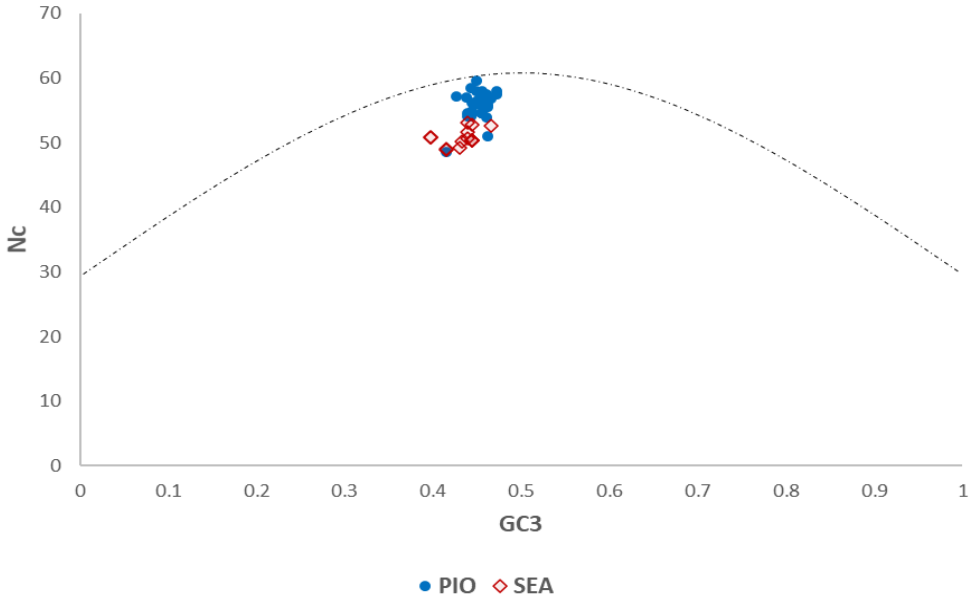


Fig 11. Nc plot. The relationship between number of codons (Nc) and GC content at the wobble position (GC3) of coat protein of BBTV is represented in the graph. The dotted line represent the standard curve of expected GC3 content and Nc value under random codon usage

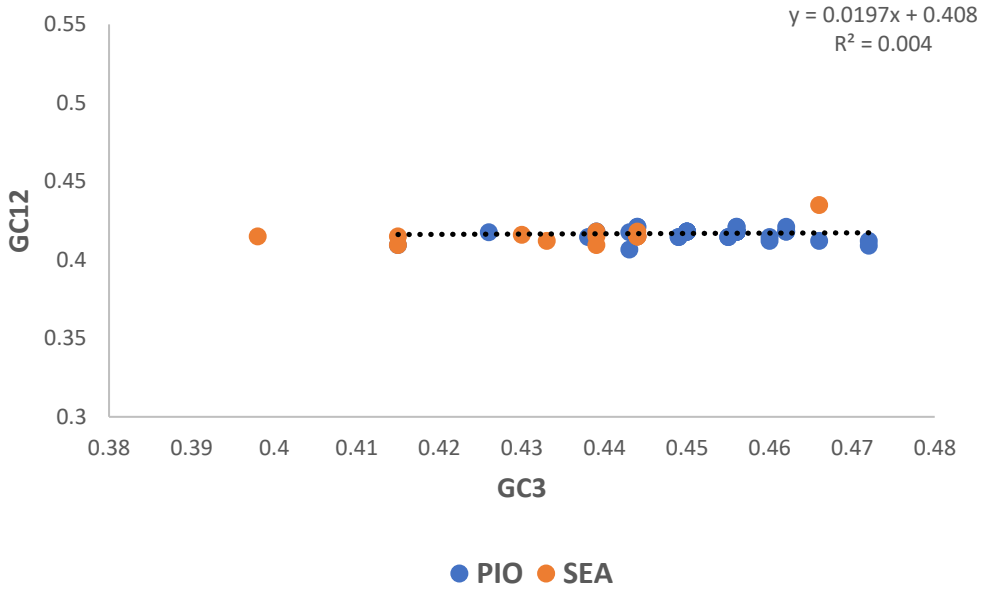


Fig 12. Neutrality plot. Investigation of the magnitude of natural selection and mutational pressure in codon usage bias of BBTV CP. The linear regression of GC12 against GC3 is represented in the graph by the black dotted line

4.2.9 Recombination analysis

From the molecular evolutionary analysis based on the CP gene of all the isolates, it was evident that the mutations in BBTV population is the result of neutral selection or purifying

selection in majority of the isolates whereas, stabilising selection is acting upon the Kerala isolates. The role of recombination in the variability of the isolates was detected using RDP 5.0. Only one recombinant was detected from the entire dataset. The most diverse isolate with accession number, MF039868.1 (Thailand) was detected to be recombinant with a Pakistani isolate as major parent (AM418566.1) and Taiwanese isolate as minor parent (DQ826393.1) (Fig 13). Recombination was detected by MaxChi and SiScan methods.

The Robinson-Fould's matrix and Shimodaira-Hasagawa (SH) matrix (Fig 14A) helped to visualise the effect of recombination on phylogeny. Although, from the white regions in matrices it is evident that CP gene of BBTV mostly have low degrees of phylogenetic compatibility, it was difficult to predict recombination hot spots. However, from the McVean's LD matrix, Low marginal likelihood values (<-2.52) are indicative of recombination cold spots (Fig 14B).

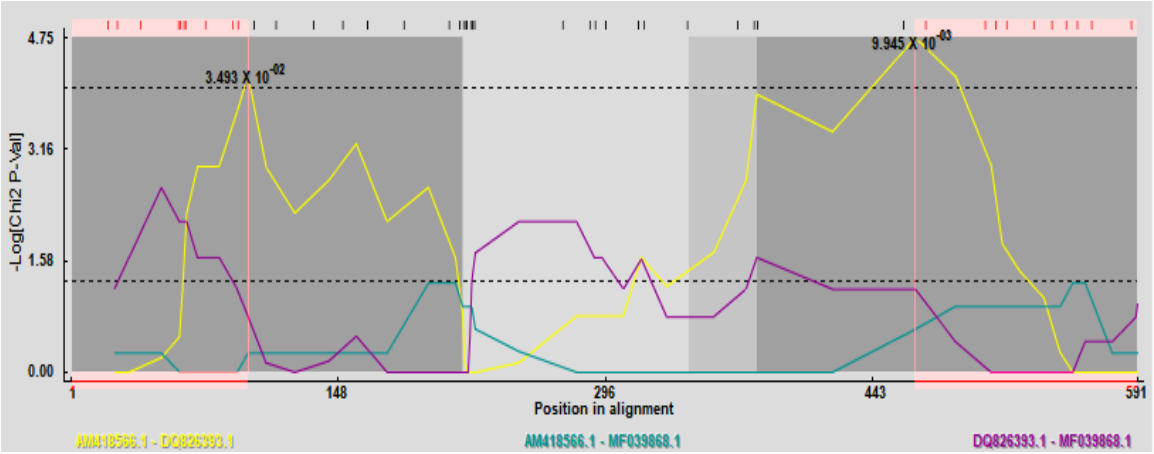


Fig 13. MaxChi plot. Major parent-minor parent plot (yellow), major parent-recombinant plot (green) and minor parent-recombinant plot (purple) are represented. Position of informative sites are displayed as black vertical lines on the top of the identity plot. Appropriate p values of the peaks are displayed inside the plot with p value cut off represented ad black dotted lines. Pink region highlights the approximate breakpoint positions.

The dark red regions, between ~240-254 and ~469-501 in the LARD breakpoint matrix and MaxChi breakpoint matrix indicated the probable positions of breakpoint pairs (Fig 14C). Hudson and Kaplan's (HK) matrix predicted the minimum number of recombination events (RMin) separating every pair of nucleotide position in an alignment. Large changes in RMin occurring over short genetic distances were calculated using RMin/D matrix. Red spots along the diagonal of the HK RMin/distance (Fig 14D) correspond with recombination breakpoint hotspot positions.

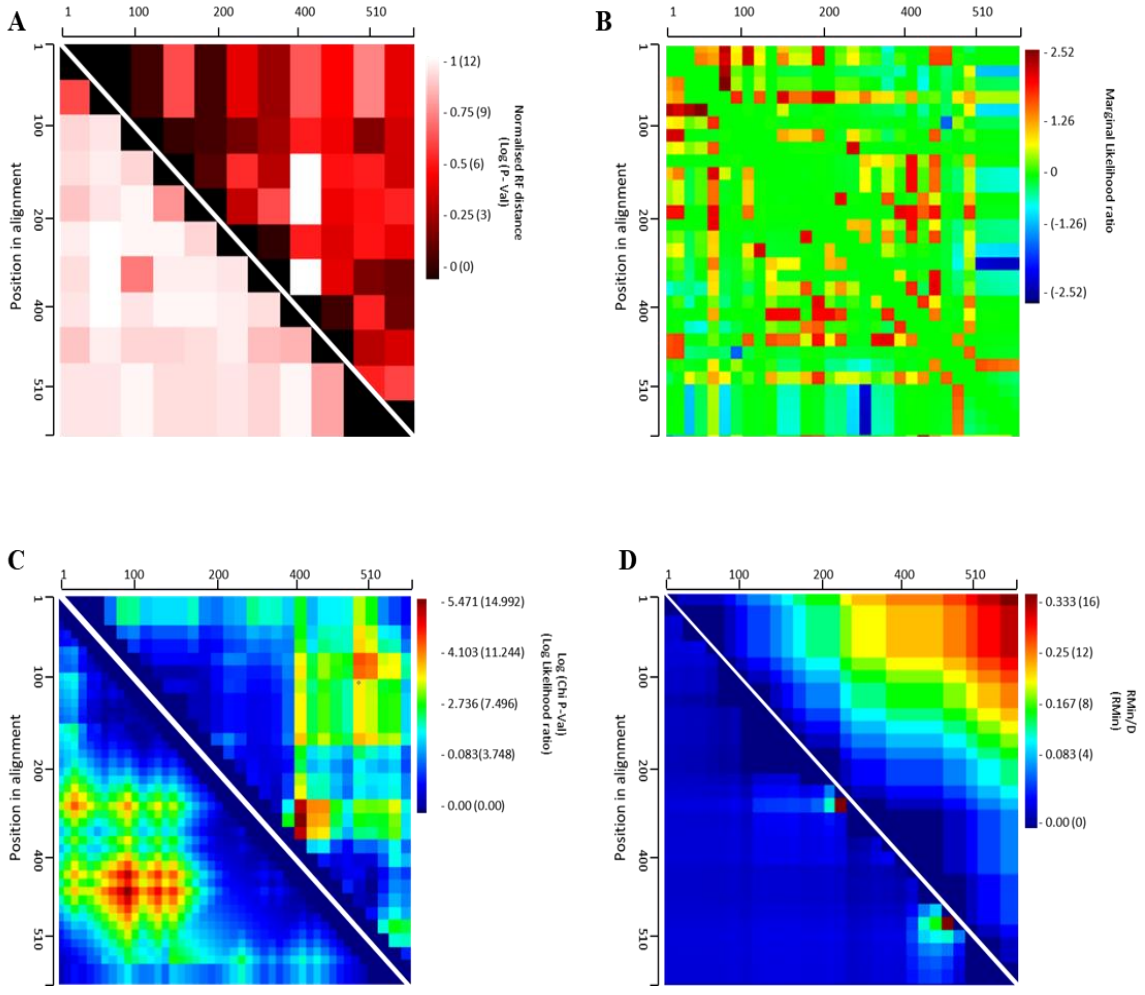


Fig 14. Recombination matrices. A Robinson-Fould's (RF) (lower) and Shimodaira-Hasagawa (SH) (upper) compatibility matrices B McVean's Linkage disequilibirum (LD) matrix. C MaxChi Breakpoint matrix (lower) and LARD breakpoint matrix (upper) D Hudson and Kaplan's (HK) RMin matrix (upper) and HK RMin/ Distance matrix (lower).

4.2.10 Validation of designed primers for detection of BBTV

Primers specific to DNA-S of BBTV was designed so as to amplify approximately 900 bp product comprising of 513 bp of complete BBTV CP gene and downstream untranslated region (UTR) region (DNA-S FP and RP). The annealing temperature was deduced experimentally by gradient PCR as 66.9 ºC (Plate 6). Samples were collected from plants that showcased typical symptoms of BBTD and total DNA was isolated. PCR amplification was done to confirm the presence of the virus and thus validate the primers for nucleic-acid based detection. Samples that gave positive reactions with reported primers also were amplified using the designed primers thus confirming that it can be used for detection of BBTV detection and routine virus indexing of banana (Plate 7).

4.2.11 Molecular cloning

For molecular cloning, competent cells of various strains of *E. coli* were prepared in ASB as mentioned in the methods section (3.13.2). The competent cells were streaked on to LB agar plate with and without antibiotics to ascertain that contamination was absent. The colonies were examined thoroughly to differentiate the strains of *E. coli*.

The CP gene was amplified using designed primers at annealing temperature 38.8 ºC (Plate 8). Restriction digestion sites of EcoR1 and Nhe1 were added on to the 5' end and BamH1 site was added to the 3' end of the BBTV CP gene. The CP gene digested by EcoR1 and BamH1 was inserted to double digested pUC19 vector and transformed to DH5α (Plate 9).

The plasmid was isolated from randomly selected white colonies and loaded on to 0.8 % gel to check difference in mobility compared to the non-recombinant plasmid. A reduced mobility indicated presence of the insert and those plasmids were selected for restriction digestion and PCR (Plate 10A). Double digestion with EcoR1 and BamH1 was performed to observe the release of ~540 bp insert (Plate 10B). Similarly, 540 bp PCR product was also observed after amplifying the CP gene using the designed primers with the recombinant plasmid as the template (Plate 10C). Few white colonies were also screened for positive recombinants by colony PCR using the designed primers (Plate 10D). Later, plasmid was isolated from the positive clones and subjected to restriction digestion for confirmation. Sanger dideoxy sequencing using M13 forward and reverse primers were carried out and ascertained that complete coding sequence of BBTV CP was present in the clone. After confirmation by any of the two methods, clones were selected and recombinant plasmids were amplified by transforming to DH5α.

Similarly, BBTV CP gene was cloned to Nhe1 and BamH1 sites of pRSET-C, Sma1 site of pGEX-4T-2 and EcoRV and BamH1 sites of pET32a (+) (Plate 11). Since lac Z gene was absent in the expression vectors used, transformed white colonies were randomly selected for colony PCR and plasmid isolation. Clones were confirmed as mentioned earlier. Given that, the CP gene cloned to pGEX-4T-2 by blunt end restriction digestion and cloning, checking the orientation of the gene was inevitable. Since the vector has a BamH1 site upstream to the Sma1 site and BamH1 recognition site is incorporated at the 3' end of the gene using the reverse primer, restriction digestion with BamH1 was carried out to confirm the orientation of the gene (Plate 11E). However, cloning to pET32a (+) was done using EcoRV and BamH1 producing blunt end and staggered end respectively after restriction, checking the orientation was not necessary. The clones of BBTV CP in pRSET-C, pGEX-4T-2 and pET32a (+) were designated as pRSET/BBTV CP, pGEX/BBTV CP and pET/BBTV CP respectively.

The BBTV CP gene was also cloned to pGEX-4T-2 by restriction free (RF) cloning to manipulate the recognition site of protein cleavage and augment the overexpression of the protein. Primers were designed (F-RF1 and R-RF1) such that the trypsin and thrombin cleavage sites between the affinity tag and the gene were eliminated and TEV protease recognition site was inserted. The primers were designed and annealing temperature was standardised as 61.3 ºC after gradient PCR (Plate 12). The 550 bp PCR product of the primary PCR reaction was used as megaprimer for the secondary PCR.

The template used was pGEX/BBTV CP clone. After secondary PCR, 5.6 kb of product was obtained (Plate 21C) which was transformed to DH5α after Dpn1 digestion. The control reaction without the megaprimer was also digested with Dpn1 enzyme and transformed. No colonies were grown in control plates whereas plate with transformants with RF plasmid had plenty of colonies (Plate 13).

The randomly selected colonies were subjected to colony PCR and plasmids isolated were subjected to PCR using the newly designed primers (Plate 14A). A PCR was also carried out using M13 reverse and gene reverse primers to obtain an amplicon of length 1.9 kb as expected in case of positive clone (Plate 14B). The elimination of trypsin digestion site and insertion of TEV protease site was ascertained by restriction digestion with BamH1. The vector was seen to be linearised in case of positive clones where the BamH1 site on the

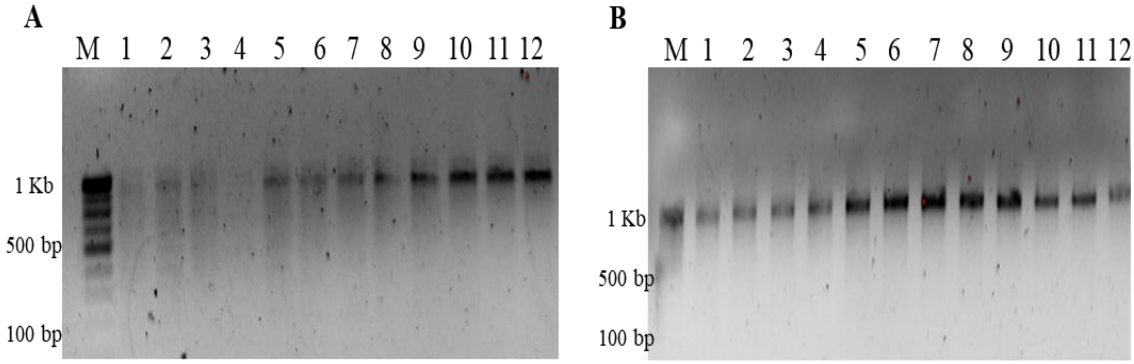


Plate 6 **Gradient PCR for standardisation of annealing temperature of designed primers for BBTV detection. A** M: 100 bp DNA ladder; Lanes 1-12: BBTV infected sample amplified using designed primers at annealing temperatures 56.0 ºC, 56.2 ºC, 56.7 ºC, 57.5 ºC, 58.5 ºC, 59.7 ºC, 60.9 ºC, 62.2 ºC, 63.3 ºC, 64.3 ºC, 65.0 ºC, 65.4 ºC. **B** M: 100 bp DNA ladder; Lanes 1-12: BBTV infected sample amplified using designed primers at annealing temperatures 64.0 ºC, 64.1 ºC, 64.6 ºC, 65.2 ºC, 66.0 ºC, 66.9 ºC, 67.9 ºC, 68.9 ºC, 69.7 ºC, 70.5 ºC, 71.1 ºC, 71.4 ºC.

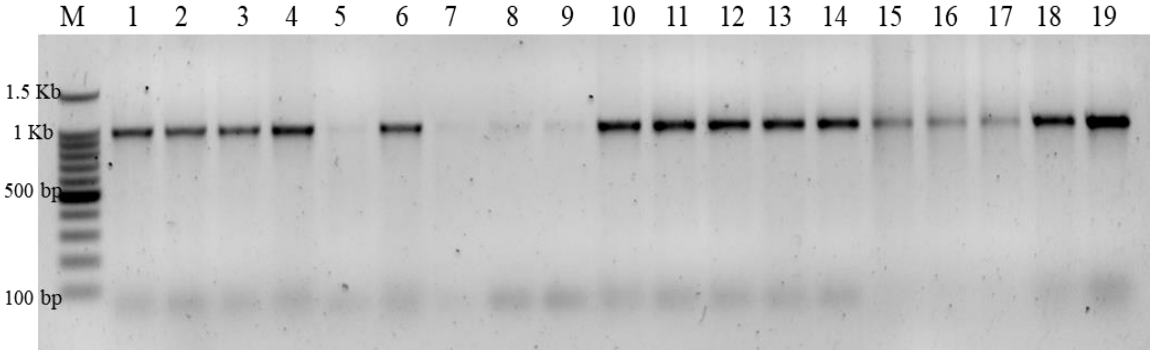


Plate 7 **Amplification of BBTV CP using designed primers from field samples suspected to be infected with BBTV.** M: 1 kb DNA ladder; Lane 1: CT1 (Kunnan), Lane 2: WYT1 (Nendran), Lane 3: BRS4 (Grand Naine), Lane 4: KKT17 (Mysore poovan), Lane 5: CKT2 (Palayankodan), Lane 6: KNT1 (Nendran), Lane 7: PKD1 (Nendran), Lane 8: BRST9 (Nendran), Lane 9: BRST6 (Kasaka), Lane 10: TCR4 (Robusta), Lane 11: BRS10 (Chengalikodan), Lane 12: BRS11 (Njalipoovan), Lane 13: BRS12 (Kunnan), Lane 14: BRS13 (Palayankodan), Lane 15: KOT1 (Nendran), Lane 16: TCR3 (Nendran), Lane 17: PKD 3 (Nendran), Lane 18: WYT3 (Njalipoovan), Lane 19: WYT4 (Nendran)

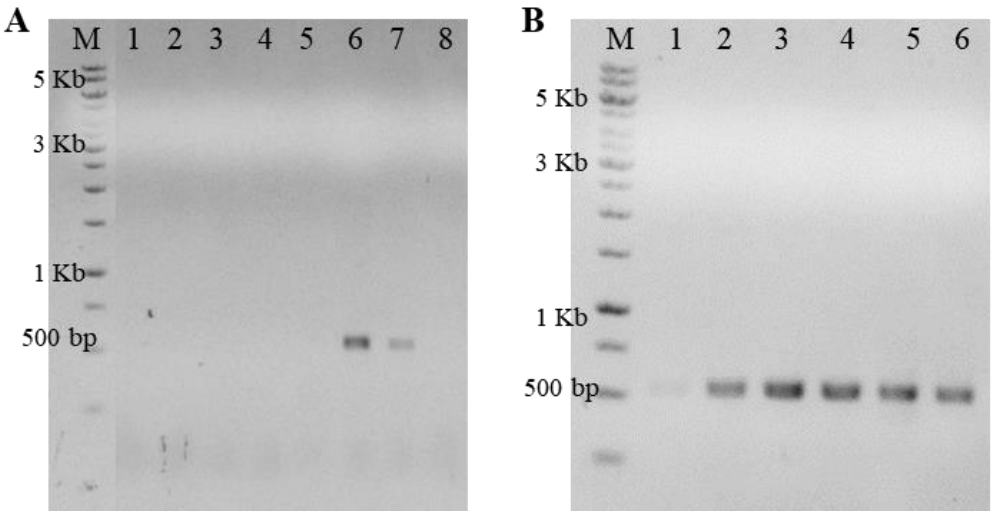


Plate 8 **Standardisation of annealing temperature of the designed primers for cloning of BBTV CP. A** M: 1 kb DNA ladder, Lanes 1-8: PCR performed at annealing temperatures 45.0 ºC, 44.4 ºC, 43.4 ºC, 42.0 ºC, 40.2 ºC, 38.8 ºC, 37.7 ºC, 37.0 ºC **B** PCR using Phusion polymerase with designed primers for cloning M: 1 kb DNA ladder, Lane 1: NC, Lanes 2-6: ~540 bp amplicon

Plate 9 **Blue white screening of transformants.** The recombinants plated on LB supplemented with ampicillin, X-gal and IPTG

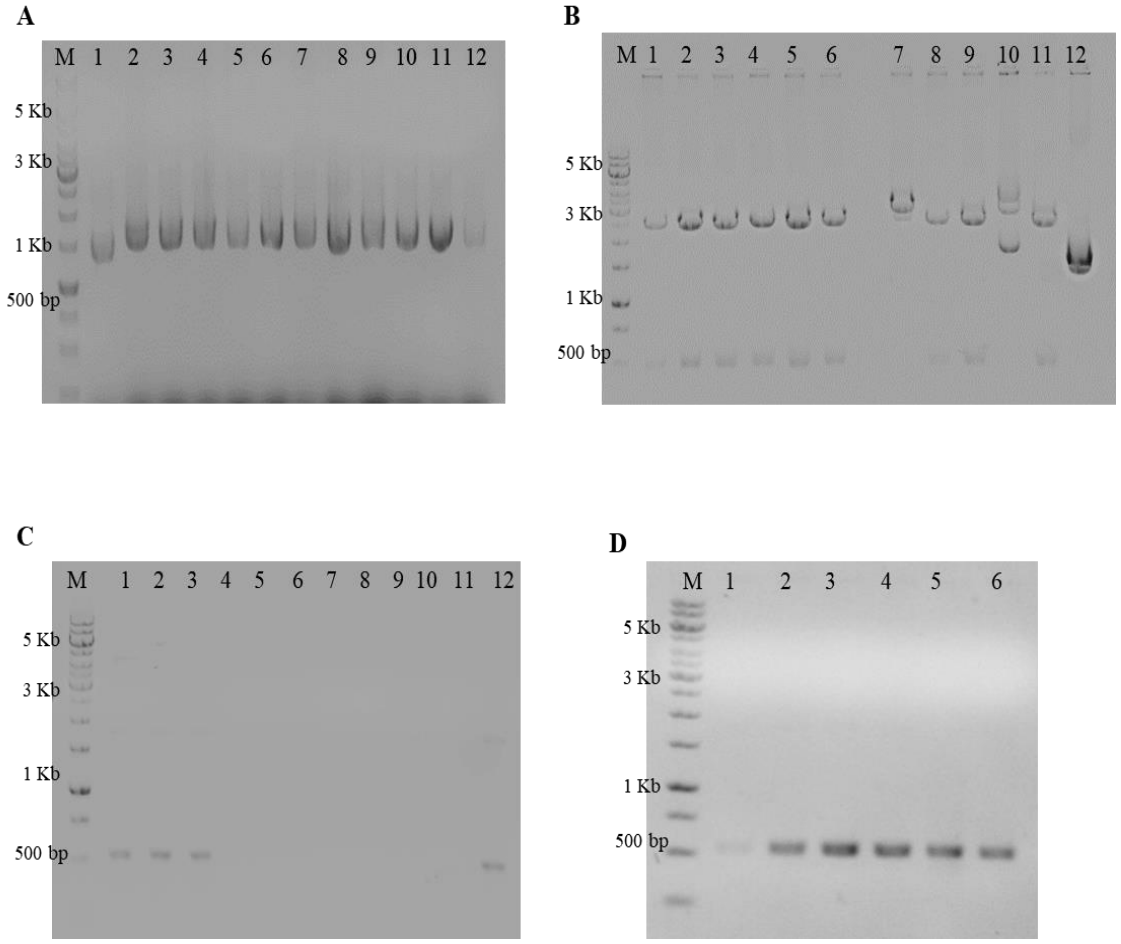


Plate 10 **Selection of positive clones and confirmation. A** Plasmid isolation by alkali lysis method and comparison of mobility with non-recombinant pUC19 vector. M: DNA ladder, Lane 1: empty vector, Lane 2-12: Plasmid isolated from randomly selected white colonies. **B** Restriction digestion of the plasmid to observed insert release at ~540 bp. Lane 1-11: double digested pUC19, Lane 12: undigested pUC19. Insert released from samples 1, 2, 3, 4, 5, 6, 8, 9, 11. **C** Amplification of coat protein gene from the recombinant plasmids using designed primers. Amplification of coat protein gene observed in samples 1, 2, 3 and 12. **D** Colony PCR of Lane 1: randomly selected blue colony, Lane 2-6: randomly selected white colonies

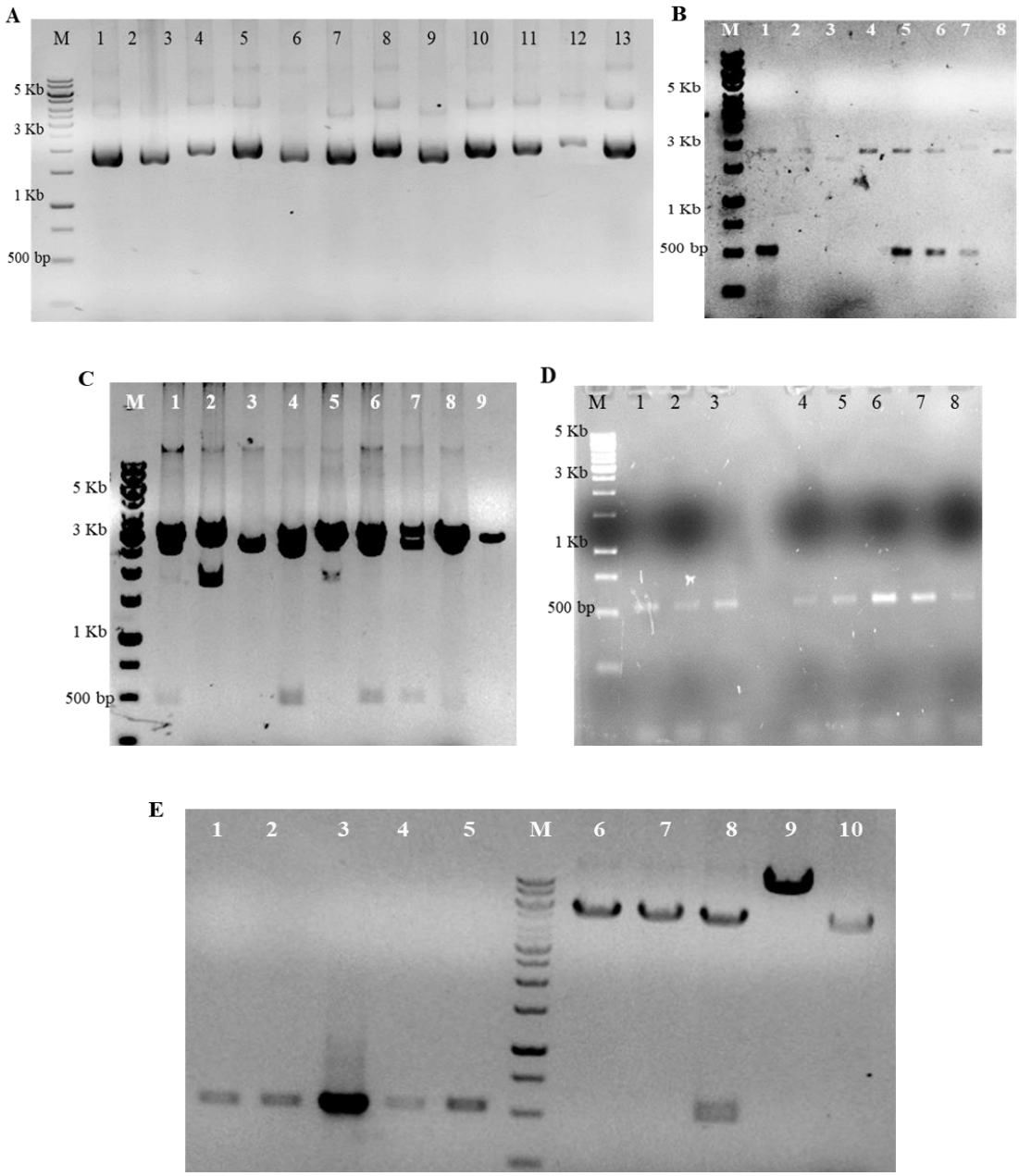


Plate 11 **Cloning of BBTV CP to expression vectors. A** Plasmid isolation by alkali lysis method and comparison of mobility with non-recombinant pRSET-C vector. M: DNA ladder, Lane 1: empty vector, Lanes 2-13: Plasmid isolated from randomly selected white colonies. **B** Clone confirmation by PCR using designed primers. Amplification of coat protein gene observed in samples 1, 5, 6 and 7. **C** Restriction digestion of the plasmid for clone confirmation. Lane 1: double digested pRSET-C, Lane 9: digested empty pRSET-C. Insert released from samples 1, 4, 6, 7 and 8. **D** Confirmation of BBTV CP clone in pET32a (+) by colony PCR of randomly selected white colonies. E Clone confirmation in pGEX-4T-2, Lanes 1-5: by PCR with BBTV CP specific primers, Lanes 6-10: orientation confirmation by restriction digestion with BamH1

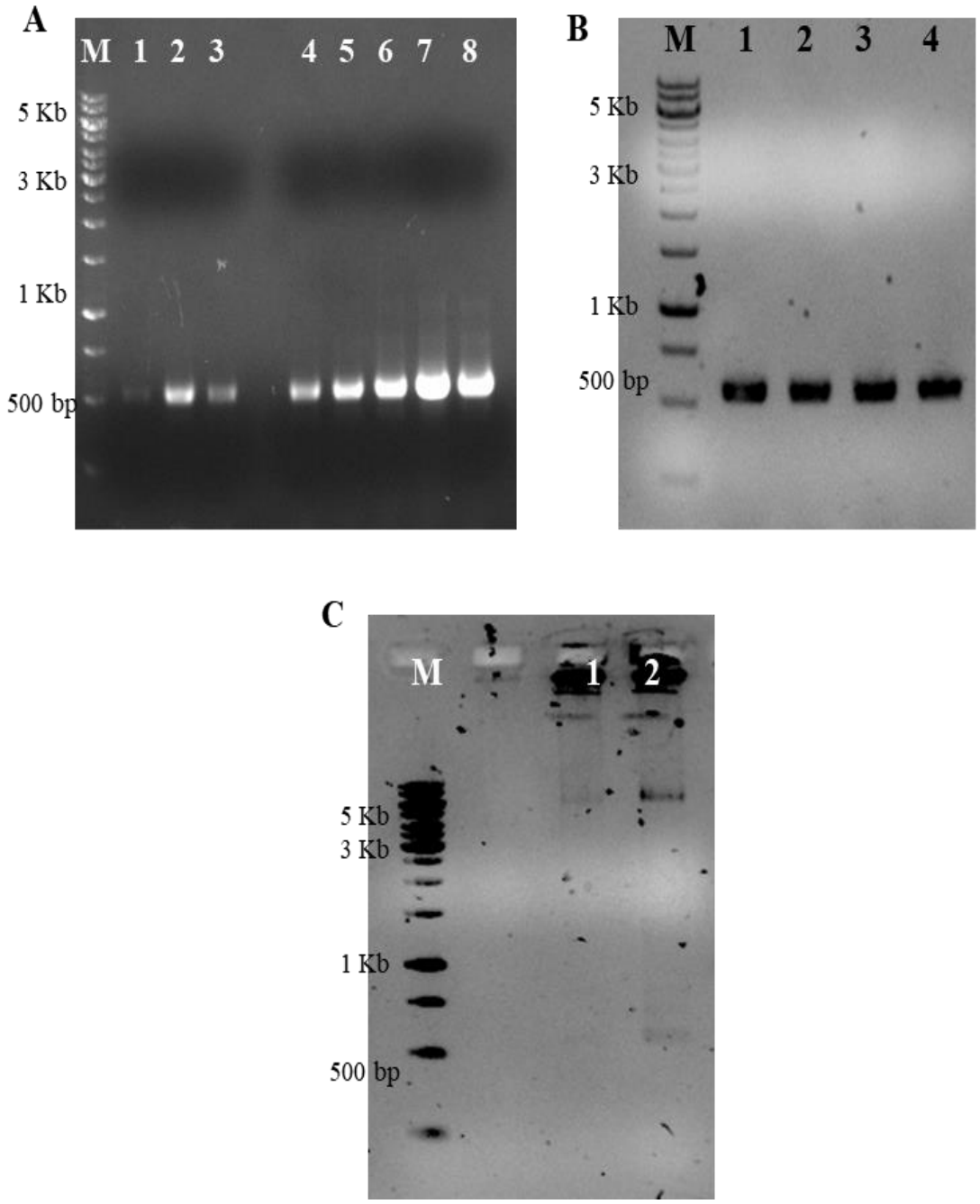


Plate 12 **Restriction free cloning of BBTV CP to pGEX-4T-2. A** Standardisation of annealing temperature for the designed primers M: 1 kb DNA ladder, Lanes 1-8: PCR performed at annealing temperatures 60.0 ºC, 60.2 ºC, 60.5 ºC, 60.8 ºC, 61.3 ºC, 61.6 ºC, 61.9 ºC, 62.0 ºC. **B** PCR amplification of BBTV CP with new primers **C** Secondary PCR using megaprimers using pGEX/BBTV CP clone as template

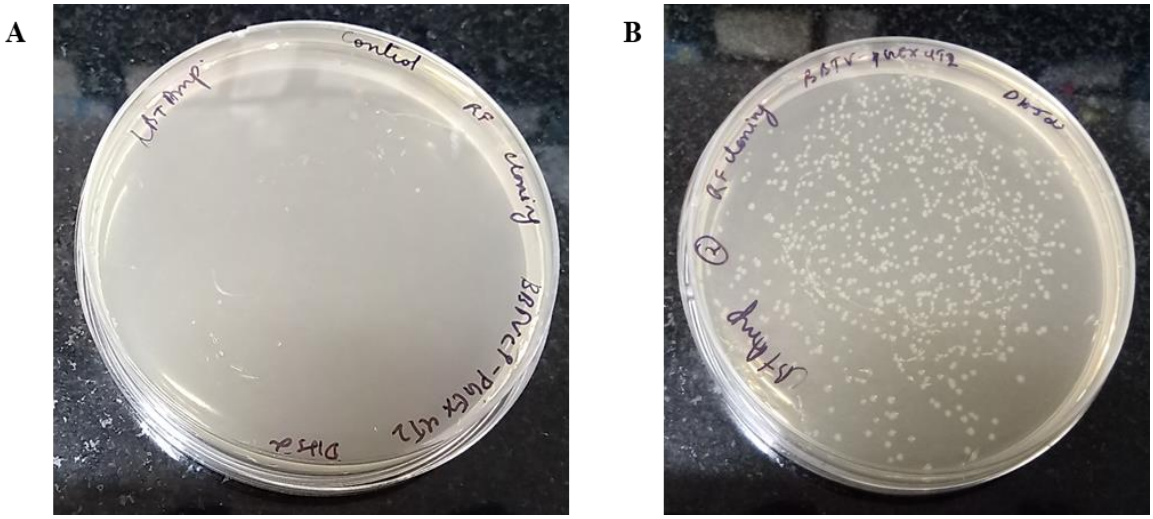


Plate 13 RF clone selection A Control plate with DH5α transformed with PCR product of reaction without the megaprimer. **B** plate with transformed DH5α after restriction free cloning of BBTV CP to pGEX-4T-2

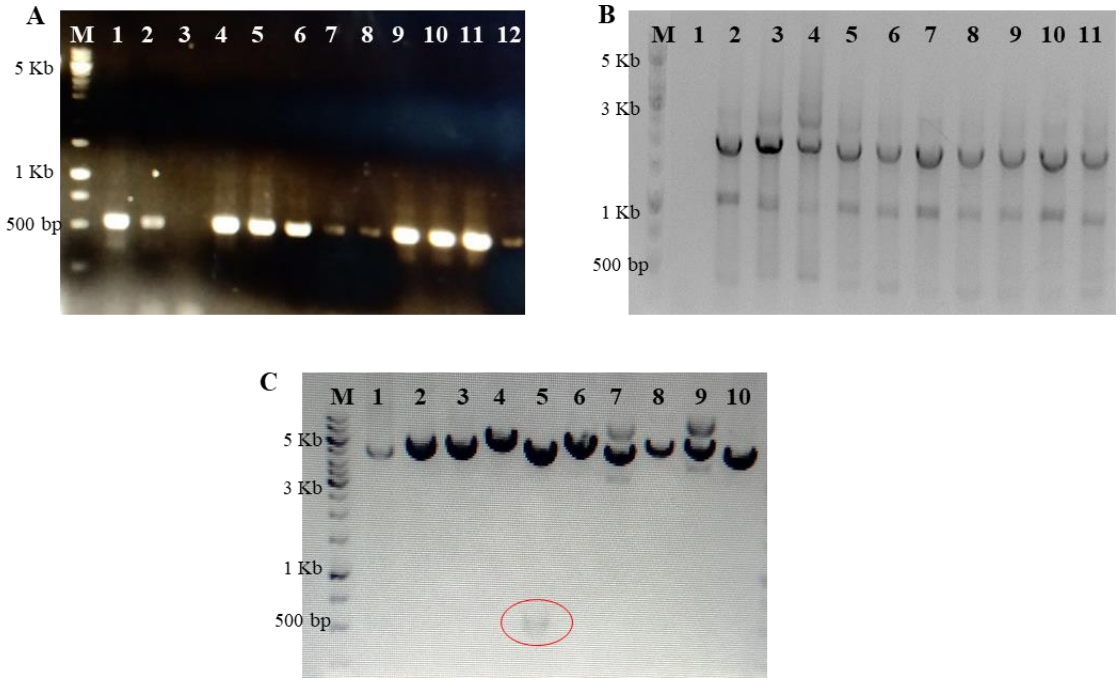


Plate 14. Confirmation of the clone ΔpGEX/BBTV CP A Colony PCR using primers designed for RF cloning M: 1 kb DNA ladder, Lanes 1-12: randomly selected colonies. Samples 1,2,4,5,6,9,10 and 11 are positive clones. **B** Plasmid isolated from randomly picked colonies used as template and amplified with M13 reverse and gene reverse primers. **C** Confirmation of clone by restriction digestion. The red circle indicates release of insert from clone that was not modified by RF cloning using megaprimers

vector, upstream to the insert, was eliminated by RF cloning and only that at the end of the gene was intact. The plasmid released the insert if RF cloning was unsuccessful (Plate 14C). Confirmed clone was designated as ΔpGEX/BBTV CP.

4.2.12 Overexpression of recombinant BBTV CP

4.2.12.1 Overexpression of pRSET/BBTV CP and pGEX/BBTV CP

BBTV CP cloned to pRSET-C vector, was overexpressed in BL21 pLysS and Rosetta pLysS strains of *E. coli*. Overexpression of 20 kDa recombinant CP was very less in both the hosts, however, comparatively, overexpression was highest at 16 ºC in both the hosts (Plate 15). Level of expression decreased with increased temperature in Rosetta pLysS whereas, the overexpression was least at 30 ºC in BL21 pLysS. Apparently, concentration of IPTG had no significant influence on the overexpression of the recombinant protein in both the hosts when the culture was induced at 16 ºC.

Since the overexpression of BBTV CP in pRSET-C clone was not efficient in both the hosts, at various temperatures of induction and concentrations of IPTG, the CP gene was cloned in to the Sma1 site of pGEX-4T-2. The pGEX/BBTV CP clone was transformed to BL21 pLysS and Rosetta pLysS so as to obtain ~45 kDa fusion protein (20 kDa BBTV CP +25 kDa GST) (Plate 16). However, the 25 kDa band corresponding to GST protein was overexpressed and the ~20 kDa BBTV CP was cleaved from the GST tag which was observed in the 12 % SDS-PAGE gel profile. However, the overexpression was highest in Rosetta pLysS at 16 ºC. When the temperature for induction was changed to 30 ºC the 20 kDa band was not observed in the gel indicating the stability of the cleaved recombinant protein at lower temperature only (Plate 17A).

In Western blot analysis using BBTV specific primary antibody, a band at ~20 kDa was visible in the pGEX/BBTV CP harbouring Rosetta cell lysate induced at 16 ºC and was absent in that induced at 30 ºC (Plate 17B). Western blot analysis confirms the degradation of BBTV CP cleaved from the GST tag at higher temperature.

4.2.12.2 Overexpression of pET/BBTV CP

Since it is tedious to purify the protein without the tag by affinity column chromatography, BBTV CP gene was cloned in to pET32a (+) at the EcoRV and BamH1 sites of the double

digested vector. The recombinant plasmid was transformed to C41 strain of *E. coli* apart from Rosetta pLysS and BL21 pLysS strains for expression of recombinant protein (Plate 18). Expression in Rosetta and BL21 with pLysS plasmid was similar. In all the strains, the protein was highly soluble.

Due to cloning, 14 amino acids from the vector were added to the N terminal of the recombinant protein apart from the thioredoxin (Trx) tag, hexa histidine (6x His) tag and S tag. The expected molecular weight of the recombinant fusion protein expressed using pET 32a (+) vector was ~37 kDa (20 kDa BBTV CP + ~17 kDa Trx- His-S tag protein and the extra amino acids added from the vector sequence).

Overexpression was standardised at various IPTG concentrations and time of induction. Temperature for induction was kept at 16 ºC as from the previous experiment it was evident that the protein is unstable at higher temperatures. Maximum overexpression was obtained in C41 strain of *E. coli* induced by 0.5 mM IPTG for 14 h (Plate 19A). Western blot was carried out using anti his monoclonal antibody with the induced and uninduced cell lysate of C41 harbouring the pET/BBTV CP recombinant plasmid and single band at 37 kDa in the induced fraction confirmed the overexpression of the fusion protein (Plate 19B).

4.2.12.3 Overexpression of ΔpGEX/BBTV CP

In the previous experiment, BBTV CP cloned to pGEX-4T-2 by conventional method and overexpressed in various *E. coli* hosts was unsuccessful. Proteolysis of the overexpressed fusion protein between the GST tag and the CP caused the separation of the target protein from its tag making purification by affinity column chromatography unfeasible. The problem was solved by RF cloning by modifying the region between the GST tag and the CP disposed to trypsin cleavage. The thrombin cleavage site was also replaced to TEV protease cleavage site in order to precisely cleave off the tag at a minimal cost.

The recombinant clone confirmed was transformed to all the three *E. coli* strains *viz.*, Rosetta, BL21 pLysS and C 41. Overexpression of the fusion protein of molecular weight slightly higher that 45 kDa (20 kDa BBTV CP + 25 kDa GST tag + the TEV protease cleavage site) was standardised in all the three hosts (Plate 20). No leaky expression of the fusion protein was observed and the uninduced fraction did not have 45 kDa band.

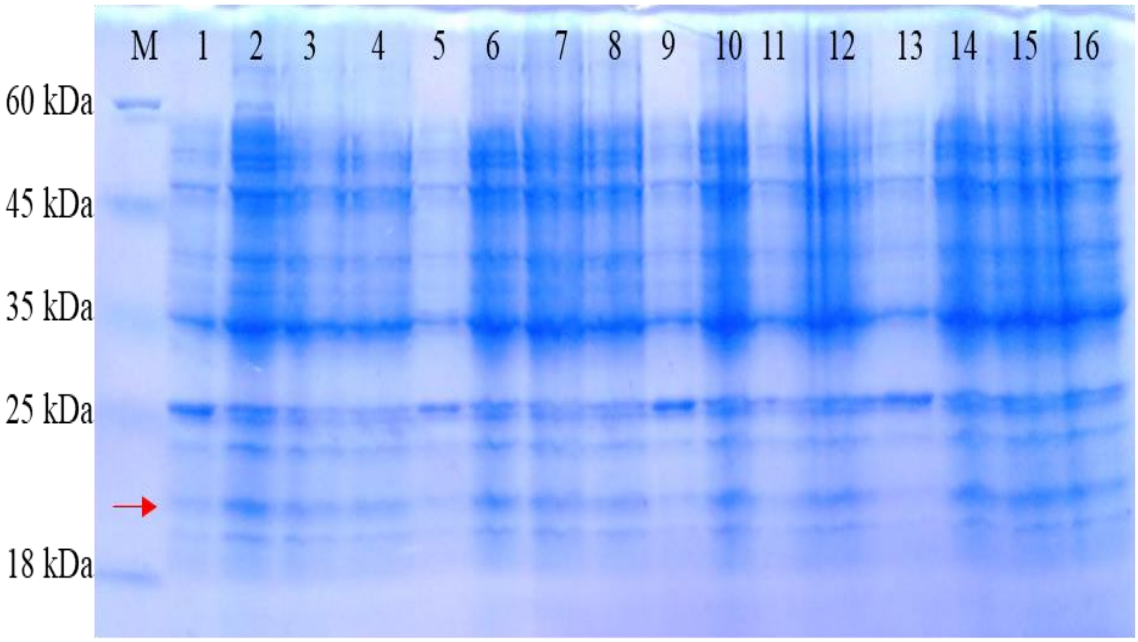


Plate 15 **Overexpression of pRSET/BBTV CP.** M: Protein molecular weight marker, Lanes 1-8: BBTV CP overexpressed in Rosetta harbouring pRSET/BBTVCP, The 20 kDa protein suspected to be rBBTV CP is marked using red arrow, Lane 1: uninduced, Lanes 2-4: induced overexpression of rBBTV CP at 16 ºC, 30 ºC and 37 ºC, Lane 5: uninduced, Lanes 6-8: overexpression of rBBTV CP induced by 0.3 mM, 0.5 mM and 1 mM IPTG at 16 ºC, Lanes 9-16: BBTV CP overexpressed in BL21 pLysS harbouring pRSET/BBTVCP, Lane 9: uninduced, Lanes 10-12: overexpression induced at 16 ºC, 30 ºC and 37 ºC, Lane 13: uninduced, Lanes 14-16: overexpression induced by 0.3 mM, 0.5 mM and 1 mM IPTG at 16 ºC

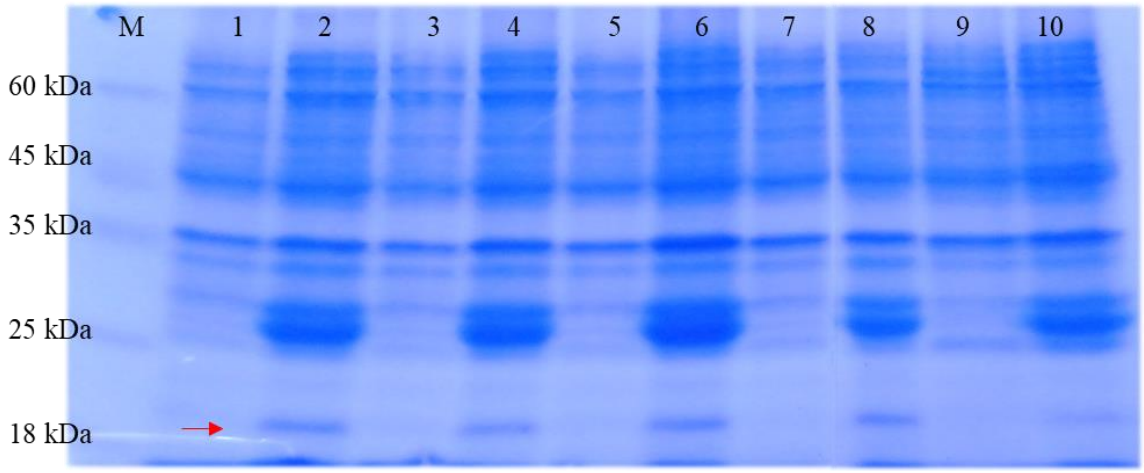


Plate 16 **Overexpression of pGEX/BBTV CP** M: Protein molecular weight marker, Lanes 1-6: BBTV CP overexpressed in Rosetta pLysS harbouring pGEX/BBTV CP, The 20 kDa protein suspected to be rBBTV CP is marked using red arrow, Lanes 1, 3 and 5: uninduced, Lanes 2, 4 and 6: induced overexpression of rBBTV CP by 0.3 mM, 0.5 mM and 1 mM IPTG at 16 ºC, Lanes 7-10: BBTV CP overexpressed in BL21 pLysS harbouring pGEX/BBTV CP, Lanes 7 and 9: uninduced, Lanes 8 and 10: overexpression induced at by 0.3 mM and 0.5 mM IPTG at 16 ºC

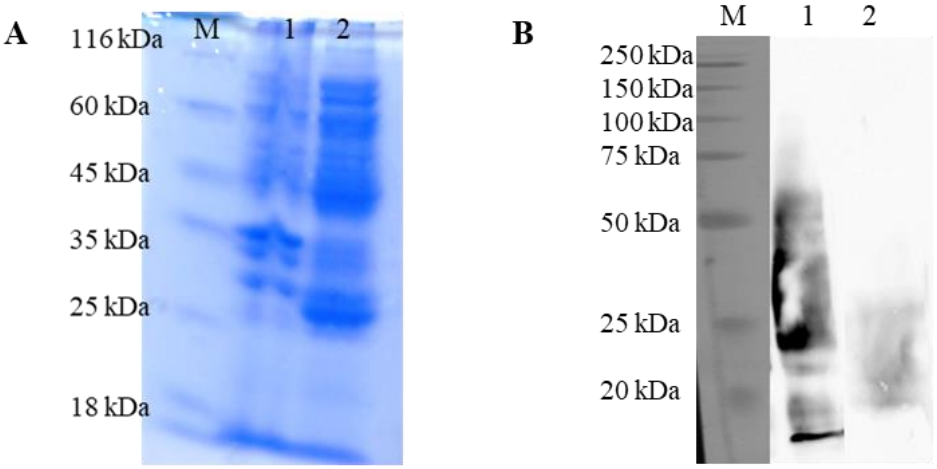


Plate 17 **Confirmation of overexpression A** Recombinant BBTV CP overexpressed in Rosetta pLysS harbouring pGEX/BBTVCP M: Protein molecular weight marker, Lane 1: uninduced, Lane 2: culture induced at 30 ºC by 0.5 mM IPTG. B Western blot analysis of cell lysate of Rosetta pLysS harbouring pGEX/BBTVCP Lane 1: induced at 16 ºC Lane 2: induced at 30 ºC

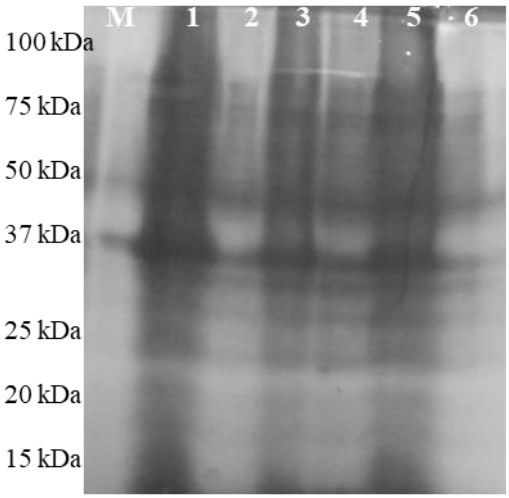


Plate 18 Standardisation of expression host for overexpressed pET/rBBTV CP M: Protein molecular weight marker, Lanes 1 and 2: uninduced and induced *E. coli* C41 harbouring pET/BBTV CP, Lanes 3 and 4: uninduced and induced *E. coli* BL21 pLysS harbouring pET/BBTV CP, Lanes 4 and 5: uninduced and induced *E. coli* Rosetta pLysS harbouring pET/BBTV CP

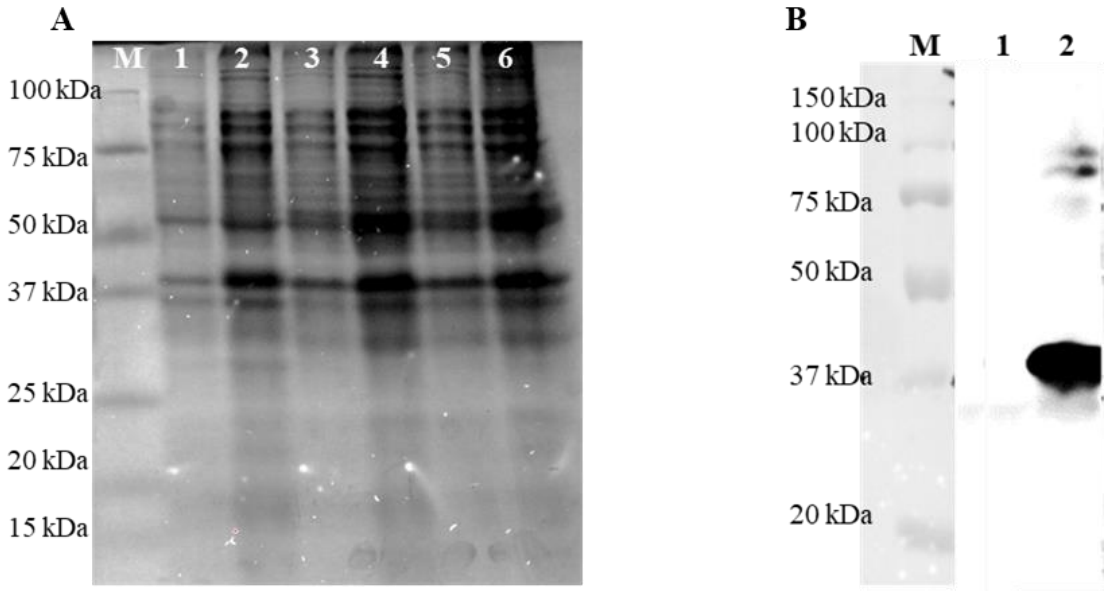


Plate 19 Overexpression of pET/BBTV CP and confirmation A Recombinant BBTV CP overexpressed in *E. coli* C41 harbouring pET/BBTV CP M: Protein molecular weight marker; Lanes 1, 3 and 5: uninduced; Lanes 2, 4 and 6: culture induced at 16 ºC by 0.5 mM IPTG for 8 h, 12 h and 14 h. B Western blot analysis of cell lysate of *E. coli* C41 harbouring pET/BBTV CP Lane 1: uninduced; Lane 2: induced at 16 ºC

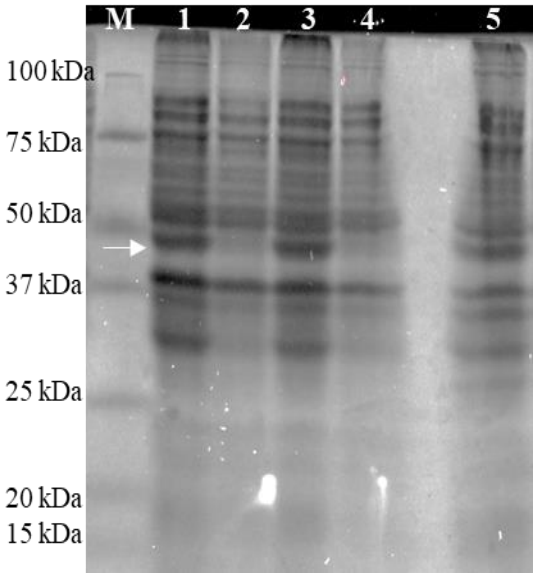


Plate 20 Standardisation of expression host for overexpressed ΔpGEX/rBBTV CP M: Protein molecular weight marker, Lanes 1 and 2: induced and uninduced *E. coli* C41 harbouring ΔpGEX/BBTV CP, Lanes 3 and 4: induced and induced *E. coli* Rosetta pLysS harbouring ΔpGEX/BBTV CP, Lanes 4 and 5: uninduced and induced *E. coli* BL21 pLysS harbouring ΔpGEX/BBTV CP. White arrow represents the overexpressed 45 kDa rCP

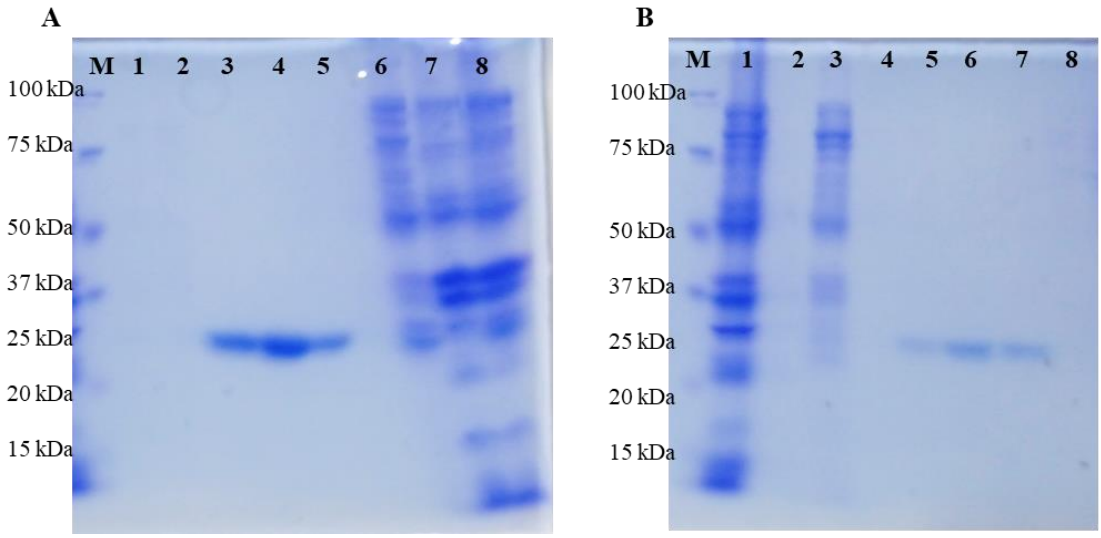


Plate 21. Purification of rBBTV CP A Purification of recombinant protein overexpressed in Rosetta pLysS transformed with pGEX/BBTV CP recombinant plasmid M: protein molecular weight marker, Lanes 1-5: elutions, Lane 6: flow through, Lane 7: pellet obtained after centrifugation, Lane 8: supernatant. **B** Purification of recombinant protein overexpressed in BL21 pLysS M: protein molecular weight marker, Lane 1: supernatant, Lane 2: pellet, Lane 3: flow through, Lanes 4-8: elutions

4.2.13 Purification of rBBTV CP

After checking the overexpression of the recombinant coat protein, in various *E. coli* hosts and standardising the parameters for maximum expression like the concentration of inducer, time and temperature for induction, purification of the overexpressed protein was carried out as per the protocol mentioned in methods section (3.18.1). Since pRSET/BBTV CP failed to overexpress the recombinant protein in the *E. coli* hosts used in the study, purification was not attempted. The pGEX/BBTV CP clone transformed to Rosetta and BL21 pLysS strains, produced fairly good amount of recombinant CP. However, due to the proteolysis between the tag and the CP, GSH sepharose affinity column chromatography was unsuccessful (Plate 21). Only the GST protein at 25 kDa was purified by the method.

Nevertheless, ΔpGEX/BBTV CP clone transformed to C41 exhibited high level of expression. This vector host pair was used for purification of BBTV rCP (Plate 22). The recombinant protein was highly soluble and was purified without any non-specific bands. Elutions were pooled and dialysed against 50X volume of dialysis buffer without glutathione.

BBTV CP gene also cloned to pET32a (+) to obtain high overexpression of recombinant protein in C 41 strain of *E. coli*. The protein was purified as mentioned in the methods section and designated as pET rCP. High concentration of fusion protein of molecular weight ~37 kDa was purified by Ni $^{2+}$-NTA column chromatography (Plate 23). Few high molecular weight protein bands of very low concentration were also observed in the gel in the second and third elutions. The elutions were pooled and dialysed against 50X dialysis buffer without imidazole to remove the same from the elution buffer. Protein was also concentrated by dialysis.

4.2.14 Cleavage of affinity tag from the fusion protein

In order to cleave the GST tag from ΔBBTV CP fusion protein (ΔpGEX/ BBTV CP), TEV protease was overexpressed in BL21 and purified by Ni^{2+}-NTA affinity column chromatography (Plate 24). As expected, 27 kDa protein was purified. Few high molecular weight proteins were also eluted along with the target protein in elutions 1-4 (Plate 24; lanes 6-9). The elutions without non-specific proteins were pooled and concentrated by dialysis.

During dialysis, imidazole was also removed from the protein. The TEV protease was quantified by Bradford's assay. From 1 L culture of BL21 harbouring recombinant plasmid with TEV protease gene, 1.261 mg/ ml of protease was obtained. The purified protease containing 50 % glycerol was flash freezed in liquid nitrogen and stored at -80 ºC as aliquots in order to avoid precipitation.

Cleavage of tag was attempted at various concentrations of the target protein as well as the protease (Plate 25). Despite using higher concentration of the protease, cleavage of the fusion protein was ineffective. As BBTV CP degrades at higher temperature (established in previous study) cleavage at room temperature or at 37 ºC were not evaluated. The 25 kDa GST protein was seen to dissociate from the fusion protein even if not treated with TEV protease and a similar SDS-PAGE gel profile was observed in the cleaved and uncleaved fusion proteins. However, 19 kDa band corresponding to the BBTV CP was not observed in the gel.

Subsequently, on column cleavage of pET rCP (BBTV CP fusion protein with Trx tag) was attempted at various concentration of thrombin. In an on-column cleavage the tag remains attached to the beads and only the target protein separates out. On the contrary, protein was seen to have cleaved as there was no band at 37 kDa, however, the 19 kDa band corresponding to BBTV CP was absent in the gel (Plate 26). In solution cleavage of pET rCP with Trx tag was also attempted. Similar to that observed in GST tagged protein, partial cleavage of the protein occurred with visible bands at both 37 kDa and light band at 17 kDa corresponding to the fusion protein and the Trx tag. However, 19 kDa protein was not present. To confirm the presence of the low molecular weight bands and to effectively separate then from one another, Tricine-SDS-PAGE was performed (Plate 27). The gel profile indicated that the fusion protein is completely digested by thrombin at 25 ºC within 8 h. However, the 19 kDa band of was not observed in this case as well.

4.2.15 Quantification of the rCP

The rBBTV CP fusion protein with GST and Trx tags were quantified by Bradford's assay. Standard plot was drawn and the concentration of rCP was deduced (Fig 15). After dialysis, 5 ml of ΔpGEX/rBBTV CP with concentration of 1.084 mg/ ml was obtained from 1 L culture. Whereas, 4 ml of 1.152 mg/ml of pET/rBBTV CP with Trx affinity tag was

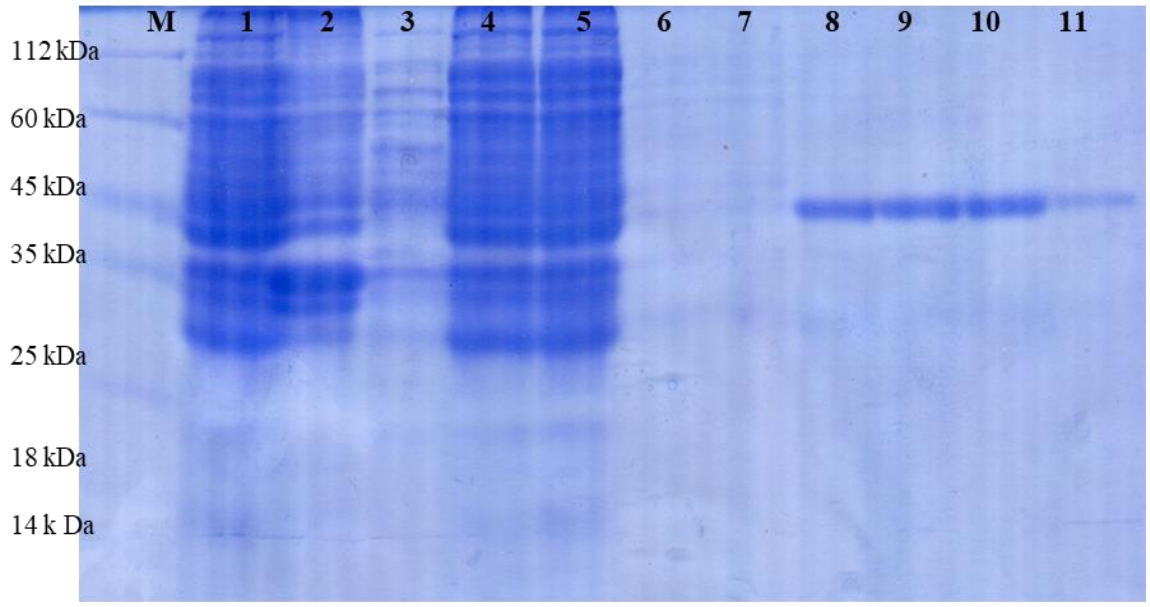


Plate 22 Purification of ΔpGEX/rBBTV CP (GST fusion protein) overexpressed in *E. coli* C41. M: protein molecular weight marker, Lane 1: supernatant, Lane 2: pellet, Lane 3: flow through, Lanes 4-6: wash, Lanes 7-11: elutions

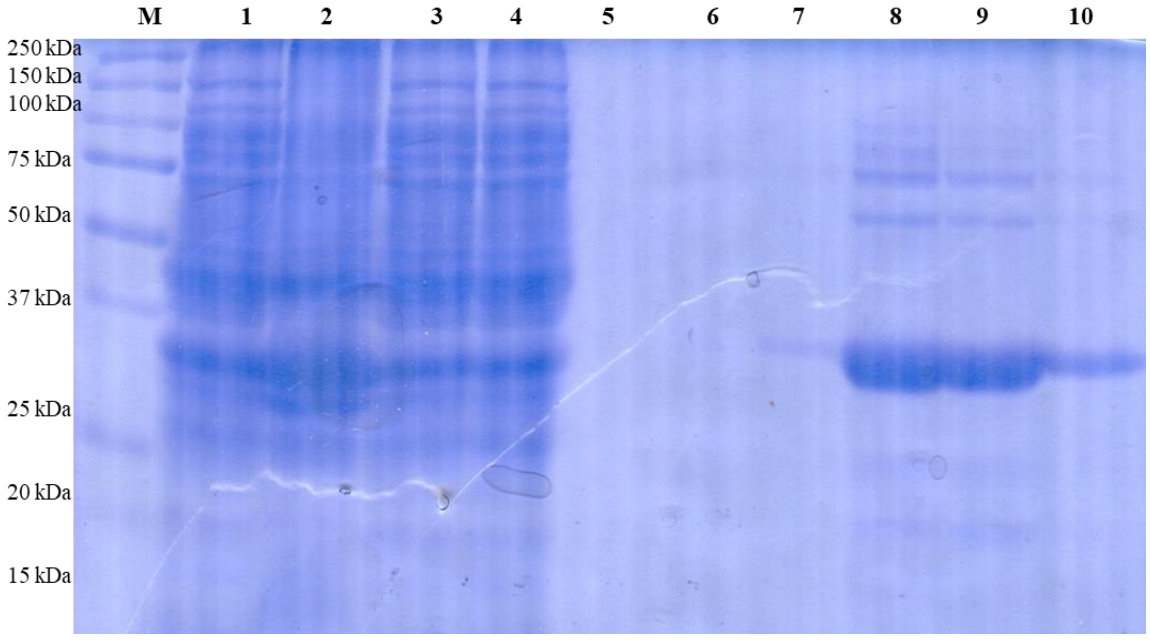


Plate 23 Purification of pET/rBBTV CP (Trx fusion protein) overexpressed in *E. coli* C41 by Ni^{2+}-NTA affinity column chromatography. M: protein molecular weight marker, Lane 1: supernatant, Lane 2: pellet, Lane 3: flow through, Lanes 4-6: wash, Lanes 7-10: elutions

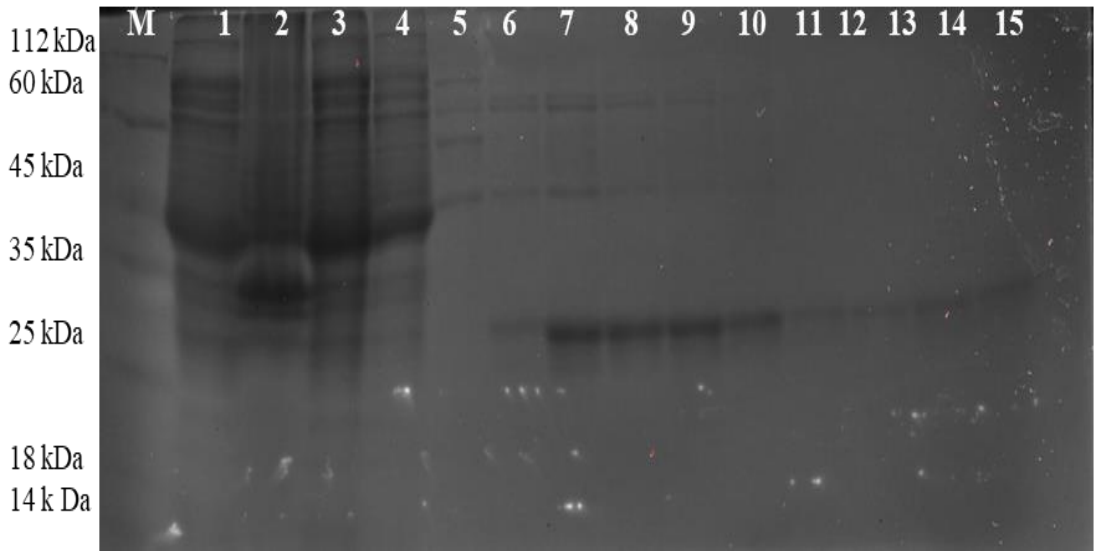


Plate 24 Purification of TEV protease. M: molecular weight marker. Lane 1: supernatant, Lane 2: pellet, Lanes 3-5: wash, Lanes 6-15: elutions

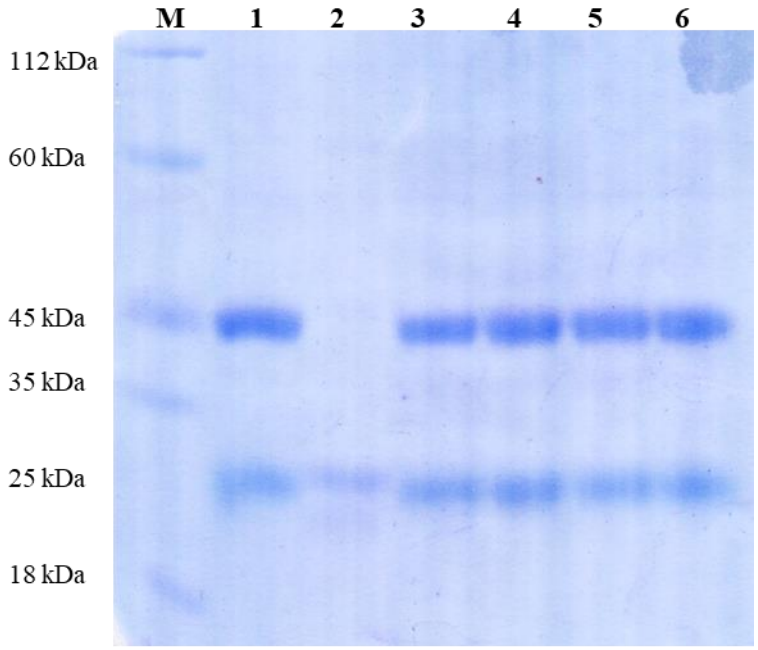


Plate 25 Cleavage of fusion protein with TEV protease. M: molecular weight marker. Lane 1: fusion protein not treated with TEV protease, Lane 2: GST protein, Lanes 3-6: 100 µg fusion protein cleaved with 2.0 µg, 1.0 µg, 0.5 µg and 0.25 µg TEV protease at 20 ºC for 12 h

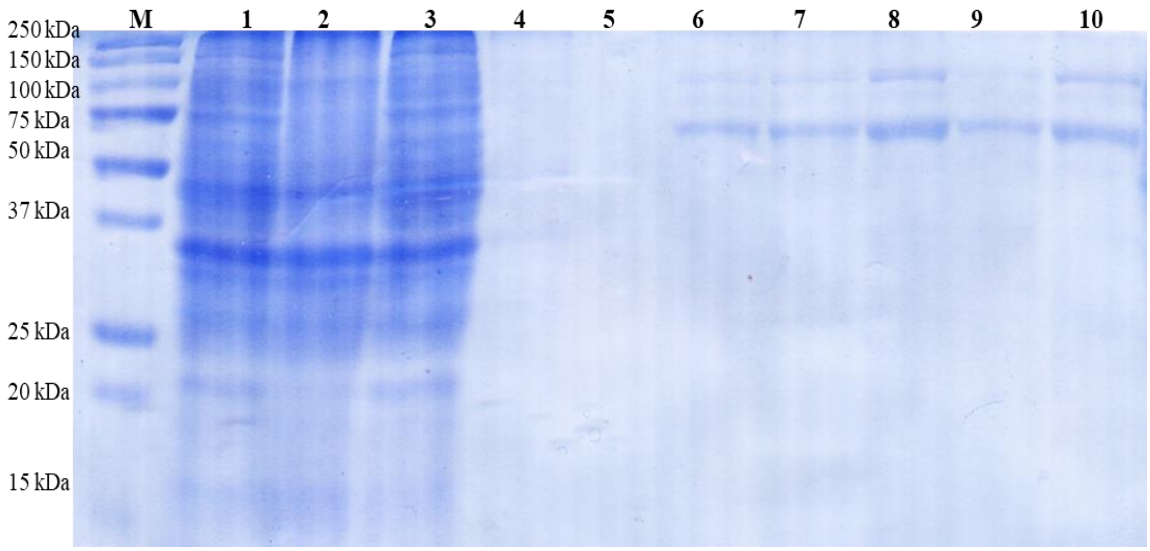


Plate 26 On column digestion of fusion protein with thrombin to cleave off the target protein from the tag. M: protein molecular weight marker, Lane 1: supernatant, Lanes 2: pellet, Lane 3: flow through, Lanes 4-5: wash, Lanes 6-10: BBTV fusion protein bound to Ni^{2+} NTA beads incubated with thrombin at 25 ºC for 2 h, 4 h, 6 h, 8 h, 12 h

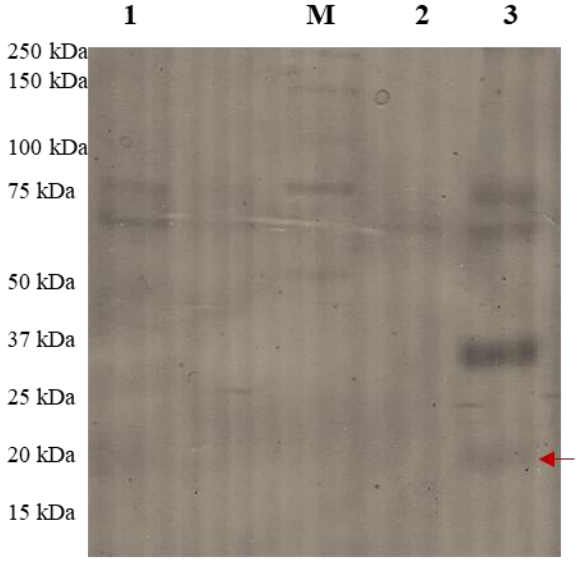


Plate 27 Tricine SDS PAGE profile of eluted BBTV CP fusion protein with Trx tag incubated with thrombin. M: molecular weight marker (corresponding molecular weights indicated outside the gel picture). Lane 1: on column digested sample for reference, Lane 2: fusion protein digested in solution for 8 h. Lane 3: fusion protein digested in solution for 4 h. The red arrow corresponds to the Trx tag released after partially digestion of fusion protein

obtained from 1 L after dialysis. Considering that Trx tag is less bulky than the GST tag, pET/rBBTV CP was used for immunisation of animals for raising polyclonal antiserum.

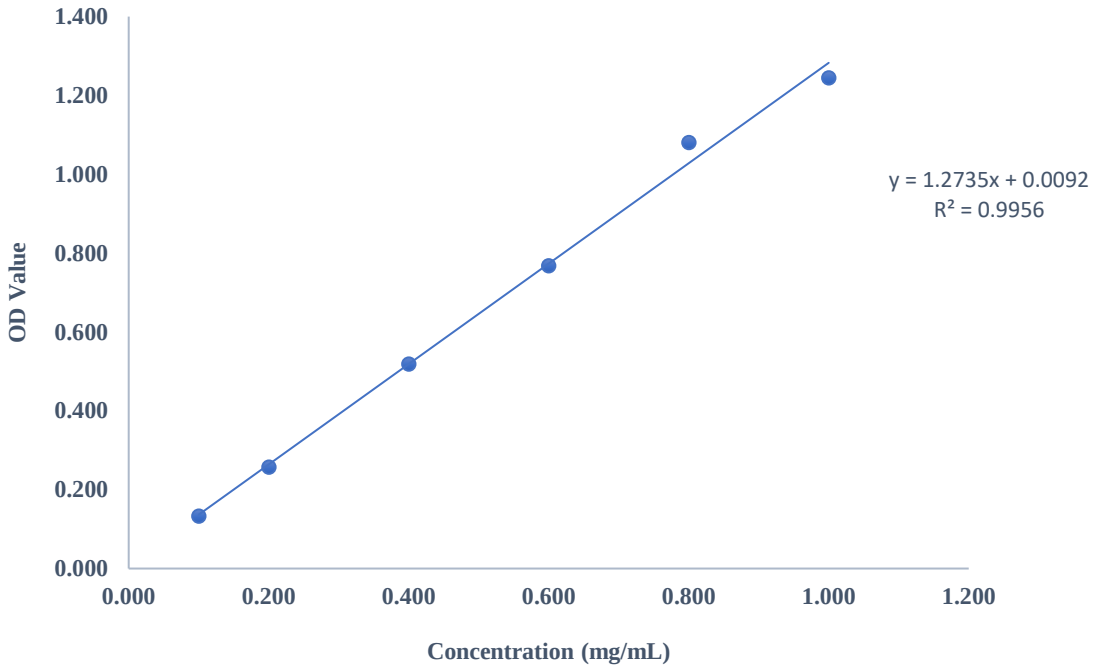


Fig 15. Standard graph of different concentration of BSA. Linear equation is displayed in the graph

4.2.16 *In silico* analysis of CP

4.2.16.1 Fold analysis

FoldIndex analysis of BBTV CP and BBTV fusion protein with GST tag and Trx tags were carried out. The results depicted that the arginine ® and lysine (K) N terminal of the BBTV CP was disordered (Fig 16A). Even in the fusion protein was partially disordered with two very ordered structures on either side (Fig 16 B and C). The C terminal of the GST and Trx proteins were disordered to facilitate TEV and thrombin respectively to cleave the tag off the fusion protein.

4.2.16.2 Analysis of protease cleavage sites in BBTV CP

In order to find the reason behind the dissociation of the BBTV rCP fusion protein from the GST tag, the fusion protein sequences were analysed in Peptide cutter (Fig 17). The

lysine and arginine rich N terminal of BBTV CP was predicted to be highly prone to trypsin cleavage. Unusually, a thrombin cleavage site was also predicted between the 55^{th} and 56^{th} amino acid of the BBTV CP. Thrombin, a serine protease, recognises amino acid sequence, LVPRGS and cleaves between arginine and glycine residue. These might be the probable reasons for the separation of the tag from the target protein in solution making it impossible to purify the target protein using affinity column chromatography.

Thrombin cleavage at a lower temperature might also be cleaving the protein at non specific site which is degrading the protein. However, in ΔpGEX/rBBTV CP fusion protein, the thrombin cleavage site was replaced by the TEV protease site which was more specific (Fig 17B). From the fold analysis, it is evident that the GST protein as well as BBTV CP, except for its N terminal, are highly folded and thus the trypsin cleavage sites might be inaccessible for the proteases.

4.2.16.3 Secondary structure prediction

Secondary structure of the BBTV CP was predicted using PSIPRED (Fig 18). The protein constituted mostly of coiled coil domains followed by extended strands and helices. The disordered N terminal of the protein constituted mostly of coiled-coil domain. The arginine rich N terminal consisted of a major coiled coil domain. This was predicted to be disordered region in the protein in the FoldIndex analysis.

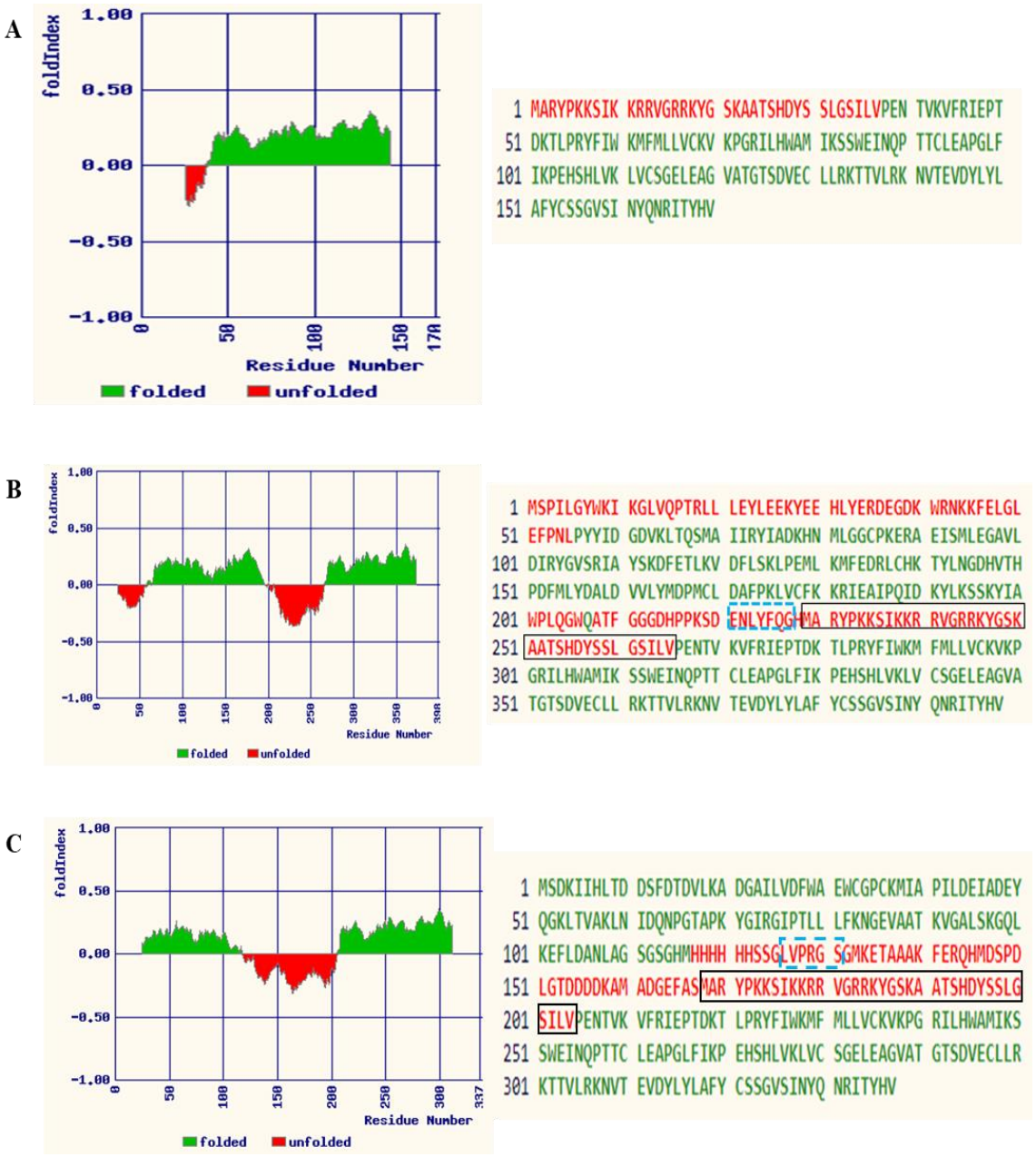


Fig 16. FoldIndex analysis of BBTV CP. Predicted ordered (green) and disordered (red) regions in **A** BBTV CP **B** ΔpGEX/rBBTV CP and **C** pET/rBBTV CP. The protease cleavage sites are marked with dotted rectangles and the N terminal disordered region of BBTV CP in fusion protein is marked in black rectangle.

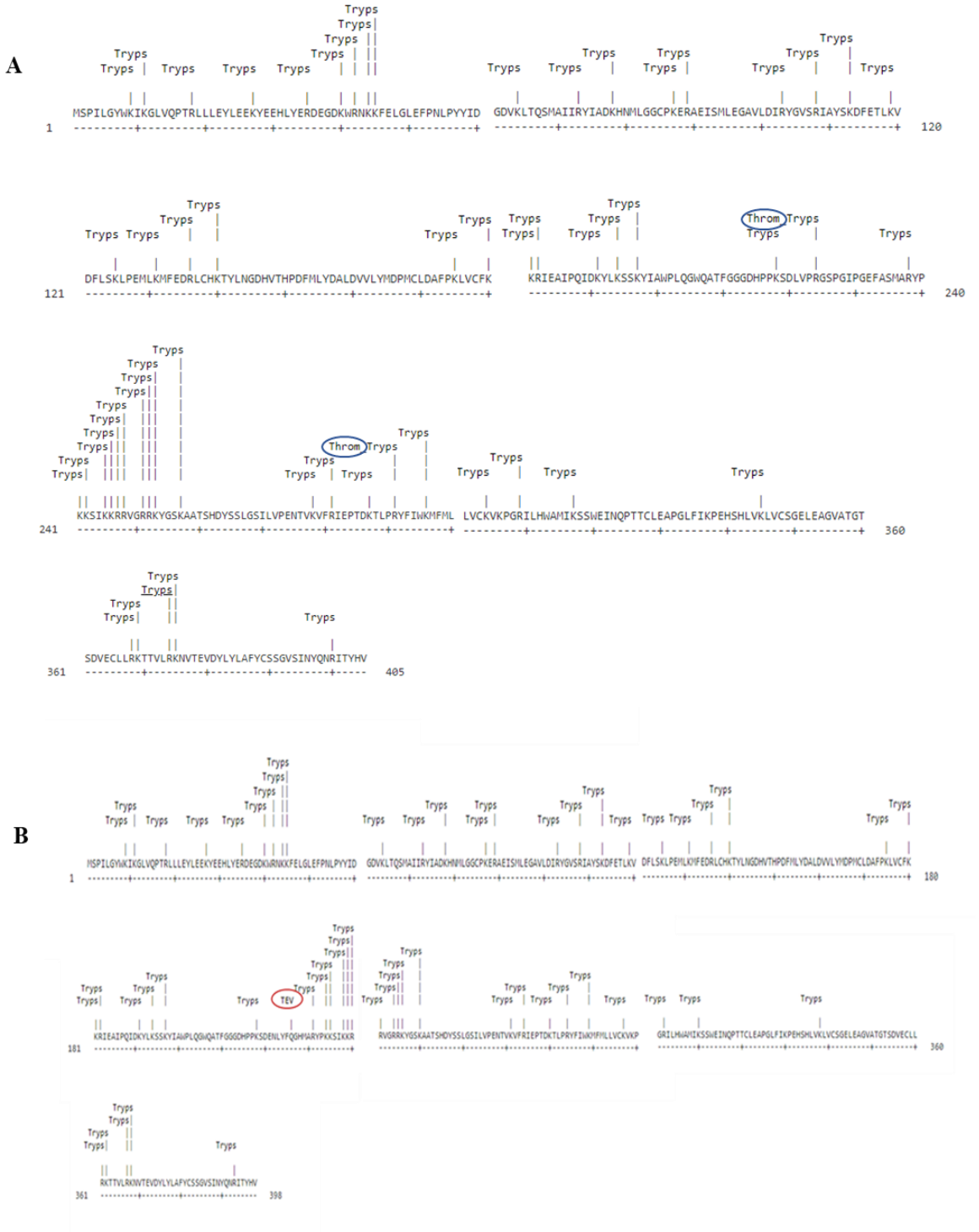


Fig 17. Predicted protease cleavage sites of A pGEX/rBBTV CP and B Δ pGEX/rBBTV CP. Thrombin cleavage site and TEV protease cleavage site are encircled in blue and red respectively.

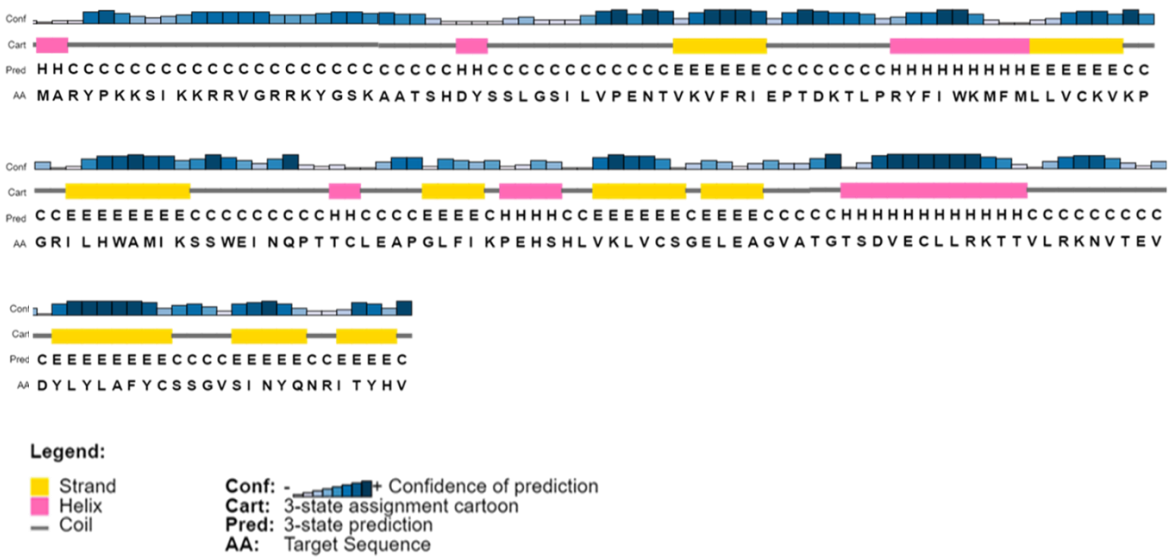


Fig 18. Secondary structure of BBTV CP predicted by PSIPRED

4.2.16.4 Protein structure prediction

Threading of BBTV CP was performed by i-TASSER suite comparing with the CP of *Ageratum yellow vein virus* (PDB ID- 6F2S) (Fig 19A). These residues as well as the N terminal lysine and arginine residues were predicted to be buried by i-TASSER. Eight β sheets (yellow) were found in the entire structure along with the coiled coil domains (green). The Ramachandran plot of the model drawn in MolProbity software proposed 51.8 % of all the residues were favoured and 79.2 % of all the residues were in the allowed region. However, 35 outliers were also estimated in the analysis (Fig 19B). Overall quality score of the model calculated by ERRAT was 75.92. The z score of the model was -4.34. The ProSA analysis showed that knowledge-based protein folding energy of the model was highest in the N terminal (Fig 20).

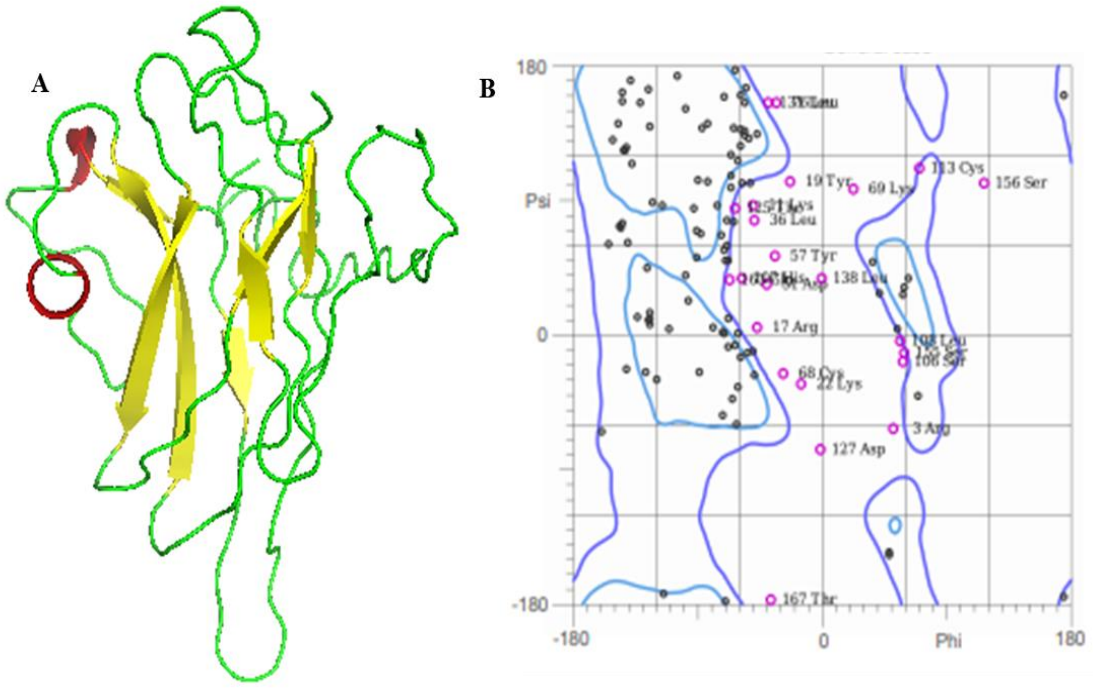


Fig 19. Homology modelling of BBTV CP. A Homology model of BBTV CP. **B** Ramachandran plot of the model

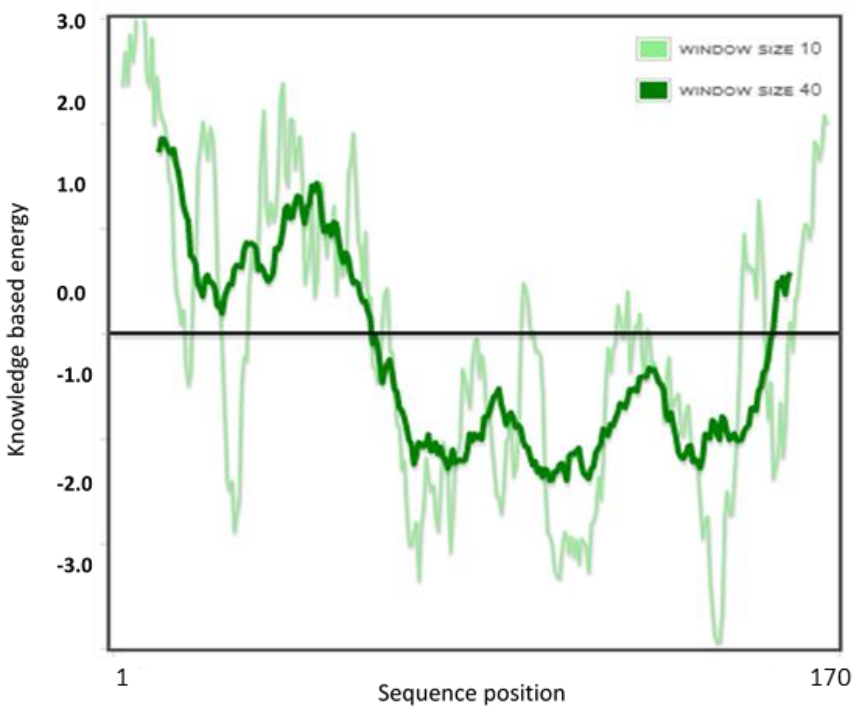


Fig 20. ProSA energy profile of BBTV CP modelled by iTASSER. Thin and thick lines indicate average energy over each 10 and 40 residue fragments respectively.

4.2.16.5 Epitope prediction

Major linear epitope for B cell was present at the N terminal of BBTV CP constituting 32 amino acid residues as per Bepipred Linear Epitope Prediction 2.0 (Fig 21A). Interestingly, a linear epitope of 80 residues was detected in pET/rBBTV CP consisting of C terminal of the affinity tag and the N terminal of BBTV CP (Fig 21B). This might prove to be more immunogenic than the rCP without the tag. Moreover, only minor epitopes were present in the affinity tag other than the one associated with the N terminal of BBTV CP. This would result in less antibodies produced against the tag.

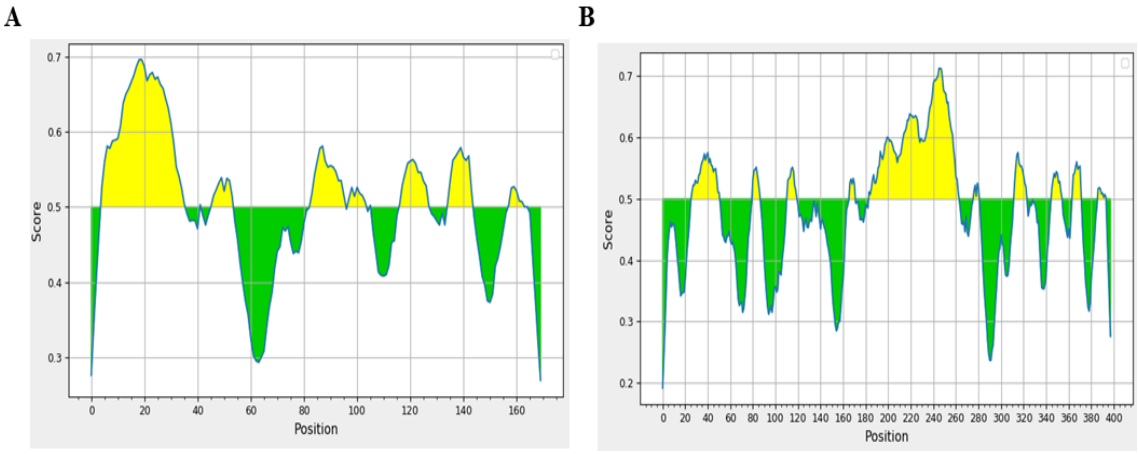


Fig 21. Epitope prediction of A. BBTV CP and **B.** pET/rBBTV CP (fusion protein) predicted by Bepipred Linear Epitope Prediction

4.2.17 Biophysical characterisation of rBBTV CP

4.2.17.1 Assessment of self-assembly

The ability of the fusion proteins to self assemble in sucrose density gradient was assessed. The SDS-PAGE gel images indicated that the ΔpGEX/rBBTV CP remained in the 10 per cent fraction of the density gradient (Plate 28A). Although a very conspicuous band at 37 kDa was seen in the ultra pellet, it was unable to be separated in sucrose density gradient (Plate 28B). In the gradient pET/rBBTV CP was present between 20 and 30 per cent of the gradient. A prominent band at 37 kDa was observed in the respective fractions but along with other *E. coli* proteins. These fractions containing maximum concentration of pET/rBBTV CP was pooled and final ultracentrifugation was carried out.

4.2.17.2 Electron microscopy

BBTV fusion protein, pET/BBTV CP apparently was self assembling in sucrose density gradient and this was validated by electron microscopy. The Transmission electron micrographs of suspected VLPs of pET/BBTV CP depicted icosahedral particles which were slightly higher than the size of the BBTV CP expected on the account of presence of the Trx-S-His tag (Plate 29). However, the preparation was not homogenous as the purification was partial and other *E. coli* proteins were present which made the results inconclusive.

4.2.17.3 Fluorescence spectroscopy

Fluorescence spectra of ΔpGEX/rBBTV CP and pET/rBBTV CP with GST and Trx tag respectively were deduced. This provided an experimental insight on the tertiary structure of the protein. The fluorescence spectra of ΔpGEX/rBBTV CP evidently had two emission maxima (λmax) at 325 nm and 345 nm (Fig 22). However, the intensity of the fluorescence was five fold higher in pET/rBBTV CP which depicted an emission spectrum comparable to that of globular proteins (Fig 23). The high intensity of fluorescence may be indicative of the partial disorder in the protein.

The contribution of the tag (Trx-S-His) to the fluorescence however, cannot be neglected. The number of tryptophan, tyrosine and phenylalanine residues predicted in pET/rBBTV CP fusion protein were 5, 11 and 11 and that in ΔpGEX/rBBTV CP were 7, 24 and 15 respectively. Both their spectra indicated that some of these residues are exposed which are contributing to the fluorescence. This validated the structure predictions carried out earlier.

4.2.17.4 Mass spectroscopy

To identify the proteins corresponding to 37 kDa (monomeric and VLP) in-gel digestion followed by MALDI-TOF analysis was carried out as described in method section. The MALDI-TOF of the monomeric 37 kDa BBTV CP fusion protein yielded 88 peptides and suspected VLP yielded 77 peptides. Both the analysis gave similar results. Surprisingly, 58 per cent of the query matched with viroplasm protein of *Cauliflower mosaic virus* (Table 12).

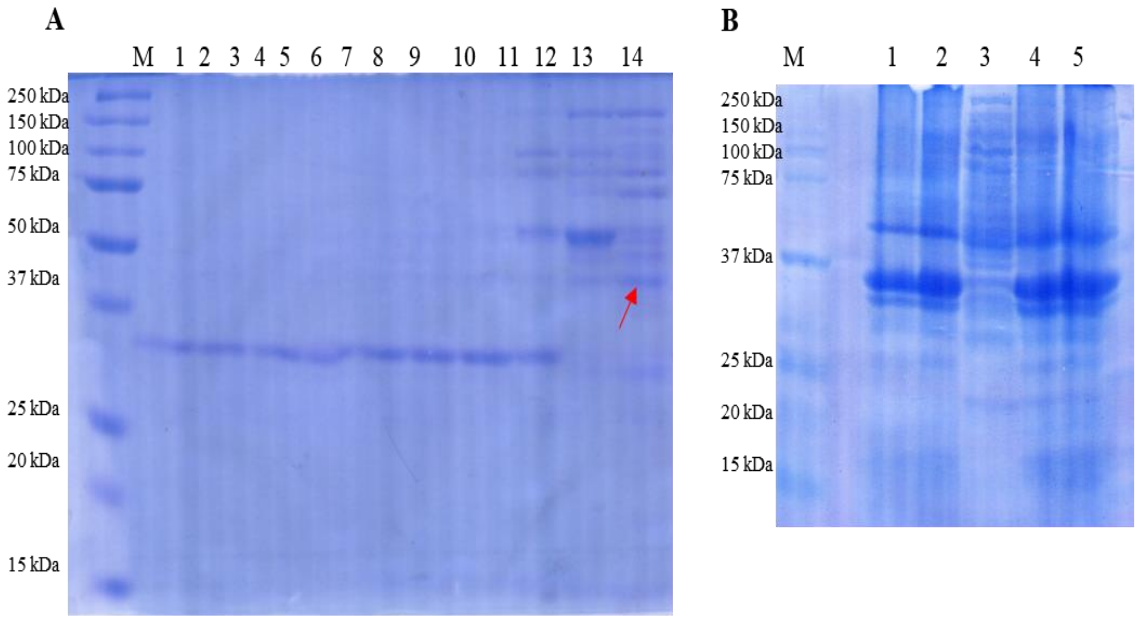


Plate 28 Assessment of self assembly of BBTV fusion protein. A Sucrose gradient centrifugation of ΔpGEX/rBBTVCP. M: Protein molecular weight marker; Fractions collected from sucrose gradient Lane 1-4: 40 %; Lanes 5-8: 30 %; Lane 9-11: 20 %; Lanes 12-14: 10 %. The 45 kDa band corresponding to ΔpGEX/rBBTV CP is indicated by red arrow. **B** Ultra centrifugation of pET/rBBTV CP. M: Protein molecular weight marker; Lane 1: Supernatant; Lane 2: pellet after lysis of induced culture; Lane 3: Supernatant after ultra centrifugation; Lane 4: ultra pellet; Lane 5: Fractions between 20-30 % after sucrose gradient ultra-centrifugation with maximum concentration of pET/rBBTV CP pooled

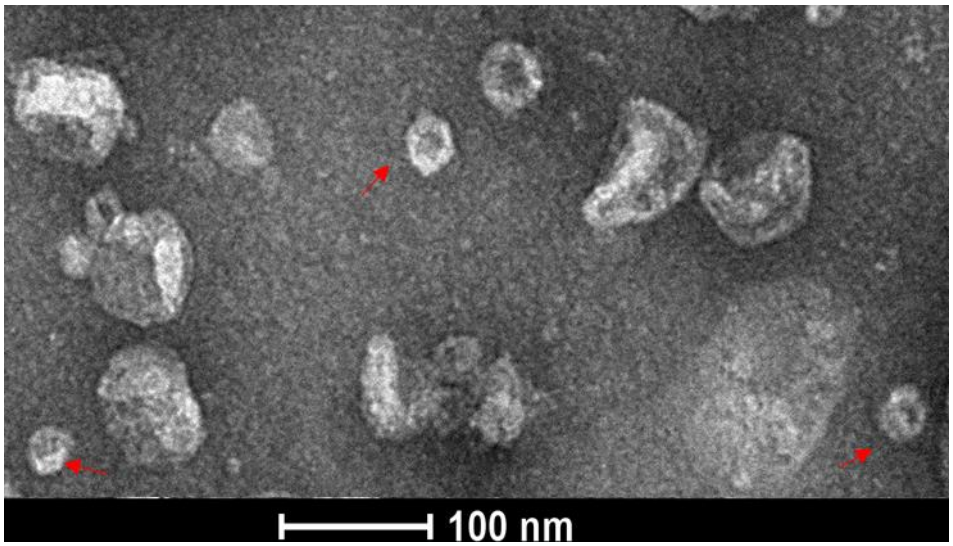


Plate 29 Electron micrograph of BBTV CP VLP. The red arrows indicate the icosahedral particles of the virus.

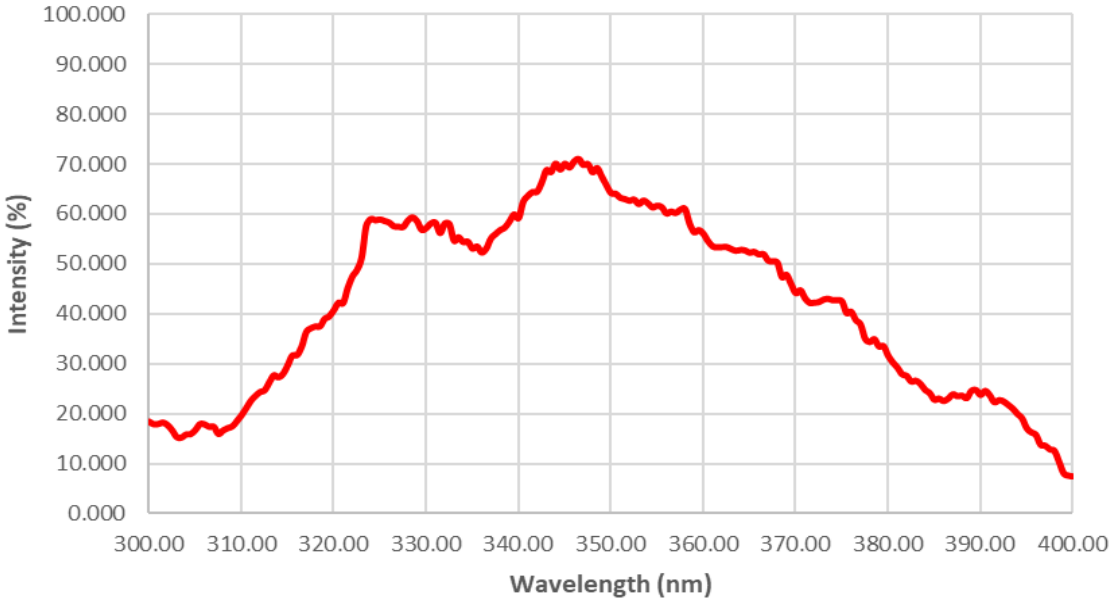


Fig 22. Fluorescence spectrum of rBBTV CP with GST tag

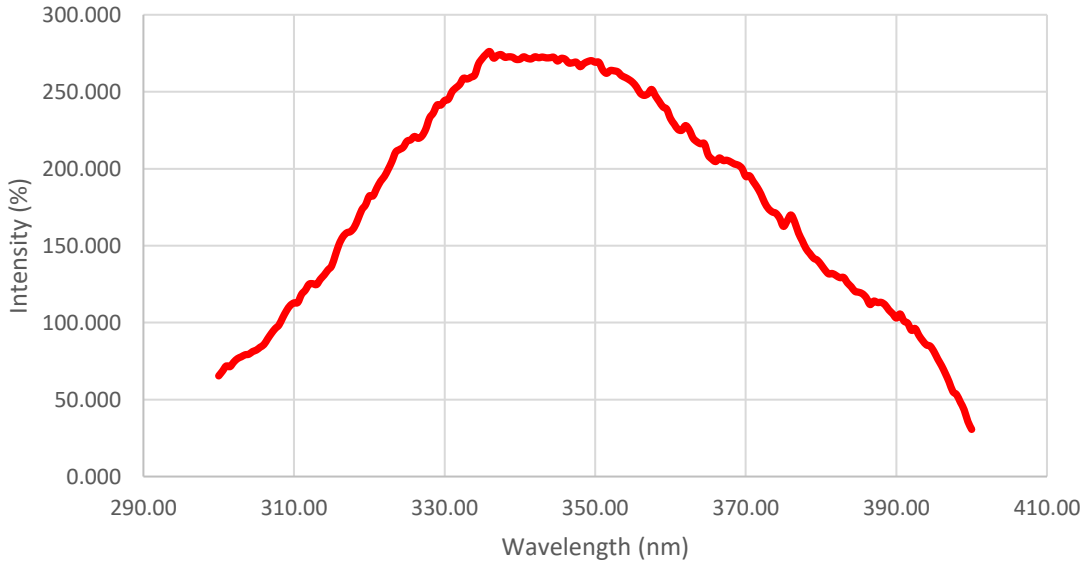


Fig 23. Fluorescence spectrum of rBBTV CP with Trx-H-S tag

Table 12. Summary of the proteome analysis used for the characterisation of the VLP and monomeric pET/BBTV CP

Protein ID	Protein description	Score	Per cent sequence coverage
VAP_CAMVD	*Cauliflower mosaic virus* virion-associated protein	33	58 %
SCAF_BPPH2	Capsid assembly scaffolding protein of Bacillus phage phi29	28	100 %
MP_MSVK	Movement protein of *Maize streak virus*	25	35 %
CAPSD_TCV	*Turnip crinkle virus* capsid protein	24	55 %

4.2.18 Antiserum production and titre confirmation

Polyclonal antibody was raised against pET/rBBTV CP fusion protein as mentioned in the methods section. Antiserum titre was estimated by DAC-ELISA and Dot Blot with the antigen used for immunisation. The titre after third bleed was estimated to be 1:128000 (Fig 24). Dot blot assay confirmed the reactivity of the antibody to the specific antigen (Plate 30).

However, field samples tested at 1:128000 dilution might give false negative results if the antigen present is very less (at early stage of the infection). Hence, the titre was confirmed by DAC-ELISA with field sample to be 1:10000 (Table 13). It was evident from the table that negative value in the commercial antiserum was higher when compared to that developed in the present study and correspondingly the OD values (absorbance at 405nm; A_{405}) of the infected samples also were more.

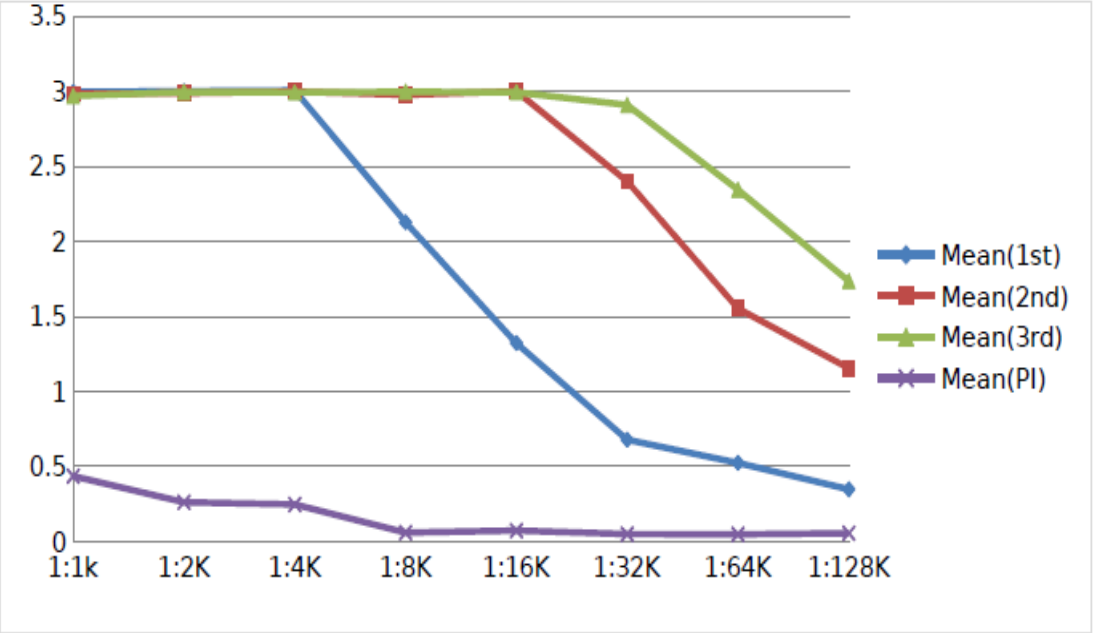


Fig 24. Titration by ELISA. ELISA results showed that the titre of recombinant antiserum from the 1st, 2nd and 3rd bleed were more than 128000

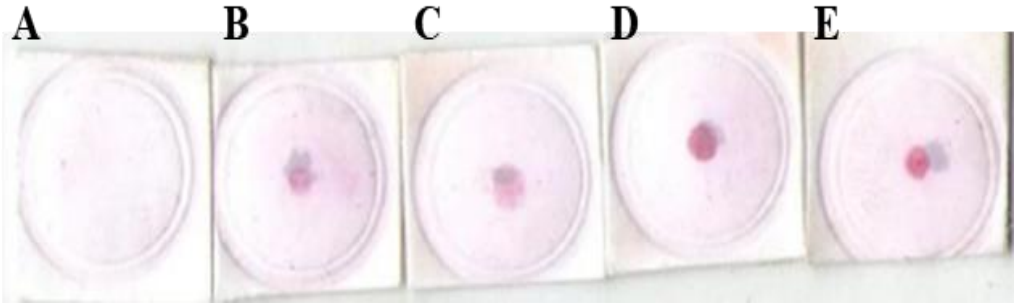


Plate 30. Dot Blot assay to study the reactivity of the antibody. A Pre-immune serum **B** first test bleed **C** first bleed **D** second bleed **E** third bleed.

Table 13. Determination of antibody titre by DAC-ELISA. The OD values are mean of 3 replications. For comparison, commercial antiserum (1:1000) was also used.

Antiserum dilution / **Sample Details**	**1:100000**	**1:50000**	**1:25000**	**1:10000**	**1:5000**	**1:1000**	**1:500**	**BBTV commercial antiserum (1:1000 dilution)**
Healthy sample	0.051	0.051	0.083	0.119	0.162	0.189	0.284	0.202
Infected sample 1	0.12	0.181	0.312	0.363	0.403	0.786	0.865	0.716
Infected sample 2	0.182	0.33	0.358	0.404	0.609	1.1375	1.027	1.011

Table 14. Interaction between primary and enzyme conjugate. The OD values are mean of 4 replications.

Primary antiserum dilution / **Secondary antibody dilution**	**1:25000**	**1:10000**	**1:5000**
1:20000	0.122	0.192	**0.355**
1:10000	0.180	**0.243**	**0.326**
1:5000	0.216	**0.293**	**0.469**

The interaction between primary and secondary antibodies at various dilutions of was also studied (Table 14). The A_{405} shown in the table are average of four replications. The A_{405} higher than double negative control is highlighted in the table. Satisfactory results were obtained at 1:10000 dilutions of primary and secondary antibodies. At lower dilutions of primary antiserum, BBTV was detected even when higher dilutions of secondary antibody (1:20000) was used. Noteworthy observation was that the reverse was not true.

4.2.19 Comparison of antiserum produced against rBBTV CP with BBTV specific commercially available antiserum

4.2.19.1 DAC-ELISA

The antiserum produced in the present study was compared with the commercially available antiserum procured from NRCB, Tiruchirapalli, Tamil Nadu. The titre value for detection of antigen from field samples using the antiserum raised against rBBTV CP was 10-fold higher (1:10000) than commercially available antiserum which was estimated to be 1:1000 earlier. This was confirmed by DAC-ELISA (Table 13). It was evident that in commercial antiserum non-specific binding was higher. Thus, the antibodies produced using purified rBBTV CP fusion protein with the higher titre value could be used for detecting the virus in the field samples using 10 fold lower concentration of the antiserum.

The sensitivity of the assay was tested by assessing the maximum dilution of antigen at which detection was possible. Antigen isolated in carbonate buffer was diluted to 1:10, 1:20, 1:40 and 1:80 using coating buffer and added to the microtitre plate wells for binding. Interestingly, at 1:10000 dilution standardised for primary and secondary antibodies antigen diluted up to 1:40 was detected by the rBBTV CP specific antiserum (Table 15).

The same concentration of antigen was detected by 1:1000 dilution of commercial antiserum. The A_{405} of diluted negative control (1:10, 1:20, 1:40 and 1:80) detected using 1:10000 v/v of rBBTV CP specific antiserum was 0.103, 0.09, 0.087, 0.057. The consistently low values of negative control at all the dilutions confirm the fact that non-specific binding of the antiserum is negligible. However, the value of negative control when detected using 1:1000 of commercial antiserum was 0.197 at 1:40 dilution and 0.254 in case of undiluted sample.

Table 15. Determination of sensitivity of DAC-ELISA. The A_{405} values are average of 4 replications. A_{405} values of the healthy sample is given in the bracket.

Antiserum dilution / Antigen concentration	1:25000	1:10000	1:5000	1:1000 (Commercial antiserum)
Undiluted	0.205 (0.108)	**0.230** (0.105)	**0.285** (0.118)	**0.862** (0.394)
1:10	0.143 (0.120)	**0.238** (0.103)	**0.280** (0.101)	**0.711** (0.351)
1:20	0.145 (0.091)	**0.185** (0.090)	**0.233** (0.104)	**0.668** (0.284)
1:40	0.124 (0.061)	**0.211** (0.087)	**0.294** (0.101)	**0.624** (0.242)
1:80	0.115 (0.057)	**0.185** (0.068)	**0.195** (0.087)	0.375 (0.189)

4.2.19.2 Immunocapture- Polymerase Chain Reaction

Immunocapture-Polymerase Chain Reaction (IC-PCR) was conducted using both the antisera. Equal concentration *ie.,* 10 ng/ml of antisera was used for capturing the antigen. However, the antigen diluted up to 1:4 was detected by rBBTV CP specific antiserum and that diluted up to 1:2 was detected by commercial antiserum. Virus extracted in TBS buffer performed best compared to other buffers. When heating was performed in presence of transfer buffers, DNA release was less and amplification was not observed (Plate 31). However, brief heating without transfer buffer amplified the gene. When heating was not done no amplification was observed. Amplification at 900 bp was obtained in infected plant samples.

4.2.19.3 Dot Immuno Binding Assay (DIBA)

Antigen was isolated in TBS buffer pH 7.5 for this assay. In order to remove chlorophyll pigment from the sample, chloroform was used. This reduced the interference in result interpretation. Secondary antibody used was anti-rabbit IgG HRP conjugate and detection of was carried out using DAB. Positive samples showed dark brown spots and in negative control, there was no colour developement. Evidently negative and positive samples could be differentiated using naked eye. This serodiagnostic assay is very useful in detection of

the virus in the field. On testing, up to 1:40 dilution of antigen was detected using 1:10000 v/v of new antiserum. The results were compared with 1:1000 v/v of commercial antiserum (Plate 32). Validation of the assay was conducted by screening 24 samples including field and tissue culture (TC) samples for presence of BBTV infection (Plate 33).

4.2.19.4 Standardisation of Triple antibody Sandwich ELISA

The sensitivity of the antiserum in TAS-ELISA was assessed. Mouse monoclonal antibody procured from Agdia (diluted to 1:500 v/v) was used as capture antibody to trap the antigen. In DAC-ELISA samples diluted up to 1:40 was detected at 1:10000 v/v of recombinant antiserum. However, TAS-ELISA detected up to 1:80 dilution of antigen with the same concentration of antibody (Table 16). Moreover, the experiment was completed within a day with higher specificity and sensitivity than DAC-ELISA.

Table 16. Triple Antibody Sandwich ELISA using recombinant antiserum

Antiserum dilution / Antigen dilution	**1:1000**	**1:5000**	**1:10000**
1:10	0.806	0.508	0.345
1:20	0.528	0.385	0.205
1:40	0.354	0.298	0.158
1:80	0.276	0.147	0.112

4.2.20 Virus indexing of field samples and TC samples by DAC-ELISA

A total of 10 field samples that were suspected to be infected by BBTV on basis of symptoms and 247 TC samples were indexed for presence of BBTV using the rBBTV CP specific antiserum for validation of serodiagnostics developed. The commercial antiserum was kept as control. All the samples tested positive for BBTV infection by commercial antiserum at 1:1000 was tested positive by new antiserum at 1:10000 dilution. The A_{405} of negative samples detected by commercial antiserum was higher than that detected by recombinant antiserum.

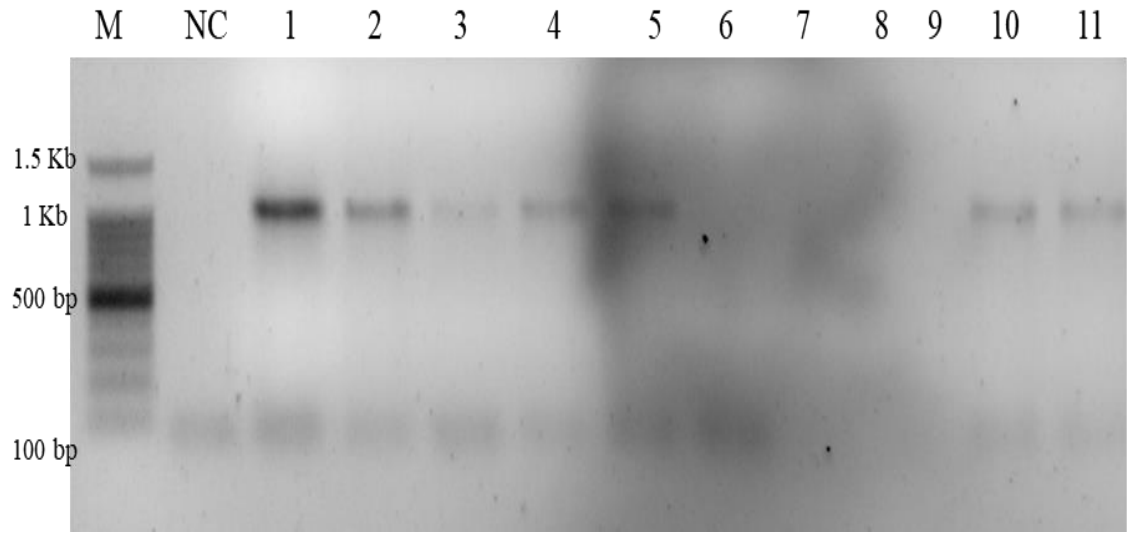


Plate 31 Standardisation of IC-PCR. M: 1 kb molecular weight marker; NC: healthy sample; IC-PCR using recombinant antiserum Lane 1: Antigen extracted in TBS; Lane 2: antigen extracted in extraction buffer; Lane 3: antigen extracted in PBS; Lane 4: antigen extracted in coating buffer; Lane 6: no heating prior to amplification. IC-PCR using commercial antiserum Lane 7: antigen extracted in TBS but no heating prior to amplification; Lane 8: antigen extracted in PBS; Lane 9: antigen extracted in coating buffer; Lane 10: antigen extracted in extraction buffer; Lane 11: antigen extracted in TBS

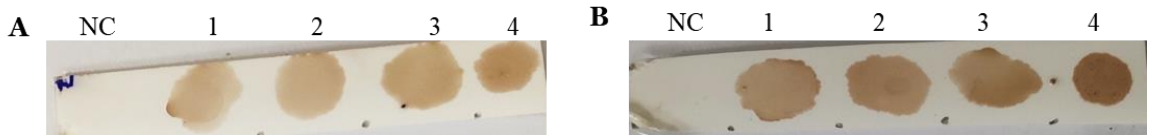


Plate 32 Comparison of **A** recombinant and **B** commercial antiserum by DIBA. NC: healthy sample; 1-4: different dilution of clarified antigen 1:80, 1:40, 1:20; undiluted unclarified antigen

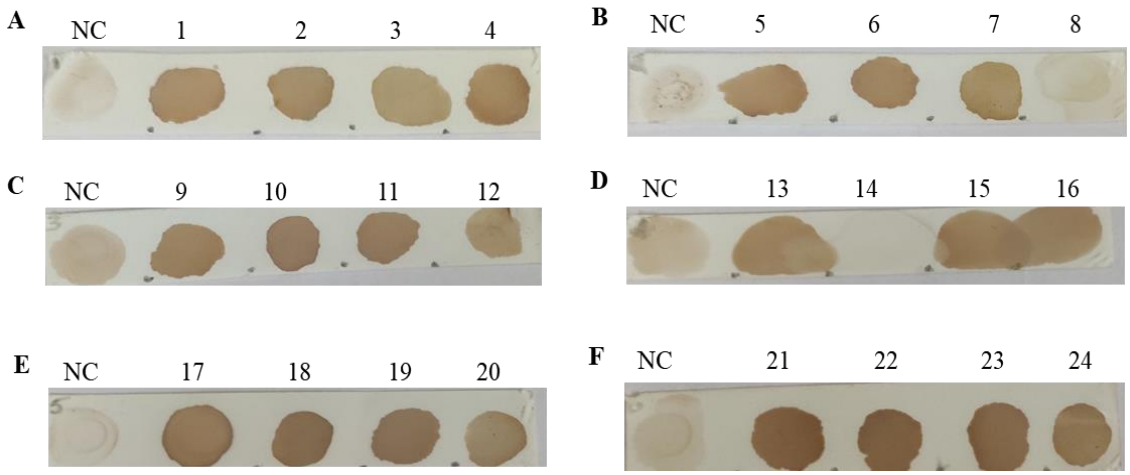


Plate 33 **Testing of field samples and tissue culture samples.** NC: healthy sample; 1: BRS11 (Njalipoovan); 2: BRS12 (Kunnan); 3: BRS13 (Palayankodan); 4: BRS14 (Nendran); 5: BRS15 (Chengalikodan); 6: TCR3 (Robusta); 7-10: TC Chengalikodan; 11-16: TC Grand Naine; 17-19: TC Njalipoovan; 20-21: TC Karpooravalli; 22: TC Myndoli; 23-24: TC Njalipoovan; Samples 8 and 14 were detected to be negative

4.3 STUDIES ON BANANA BRACT MOSAIC VIRUS

4.3.1 Symptomatology and disease incidence

A variety of symptoms on various parts of BBrMV infected plants at various growth phases were observed irrespective of the cultivar. Most prominent being, reddish parallel streaks on pseudostem, petiole, midrib and dark and light red/purple mosaic pattern on the bract (Plate 34). In the pseudostem of banana cv. Njalipoovan, red diamond shaped lesions were seen instead of reddish streaks. There was no significant variation in symptomatology spatially. Varietal variation on symptoms were also inconspicuous.

The disease incidence of BBrMD was calculated as per the observations at the time of the survey (Fig 25). The highest incidence of BBrMD was recorded at Naduthara (41.82 %) and Thekkumkara (36.58 %) in Wadakkanchery block, Thrissur.

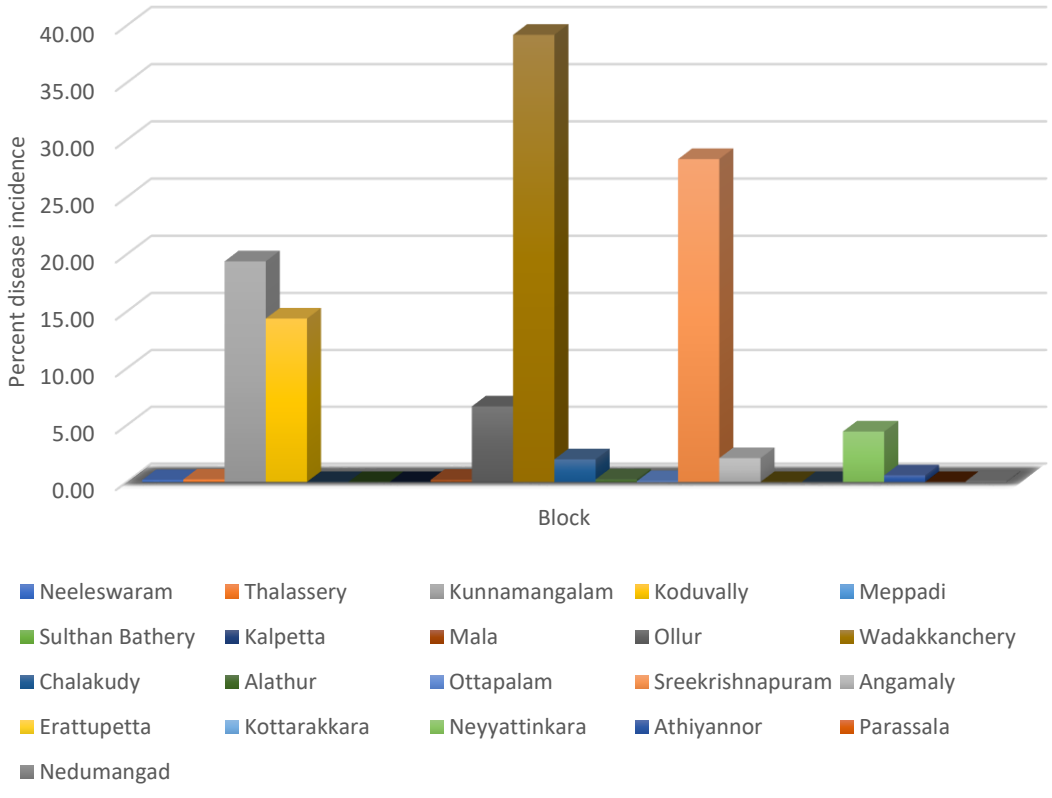


Fig 25. Block wise per cent disease incidence of BBrMD

Apart from these locations, Sreekrishnapuram block in Palakkad district also showed moderate incidence of the disease (28.33 %). Maximum incidence was recorded in

Kozhikode (18.56 %) followed by Thrissur (15.65 %) district. During the period when survey was conducted, BBrMD incidence was not recorded in any of the locations in Wayanad, Kollam and Kottayam districts.

4.3.2 Preliminary assay by DAC-ELISA

The BBrMV specific antiserum was procured from NRC Banana, Tiruchirapalli. Antiserum titre was determined to be 1:1000 v/v through DAC-ELISA (Plate 35). Virus (antigen) isolated from the samples suspected to be infected by BBrMV were coated on to ELISA plate for DAC-ELISA. Absorbance value at 405 nm more than twice that of negative control (NC) was only considered positive (Plate 35).

4.3.3 Total RNA isolation and RT-PCR

Twelve representative BBrMV infected samples that gave positive result in ELISA were selected and total RNA was isolated. A double band corresponding to host 25 S and 18 S RNA and other low molecular weight band were observed in the gel (Plate 36A). RNA was incubated with Reverse transcriptase enzyme at appropriate conditions to produce cDNA. The PCR was performed with the cDNA to amplify BBrMV CP gene (Plate 36B). An amplicon of 1062 bp was obtained after RT-PCR corresponding the complete coding region of the CP gene which is about 900 bp and 3' UTR.

4.3.4 Sanger dideoxy sequencing of CP gene and submission to GenBank

The PCR products after purification and sequencing were pre-processed and subjected to nucleotide BLAST. The sequence identity ranged from 96-97 per cent with BBrMV CP sequences present in the NCBI nucleotide repository with 100 per cent query coverage (Fig 26). Twelve BBrMV CP gene sequences were deposited under the accession numbers MT818176- MT818187 (Appendix III).

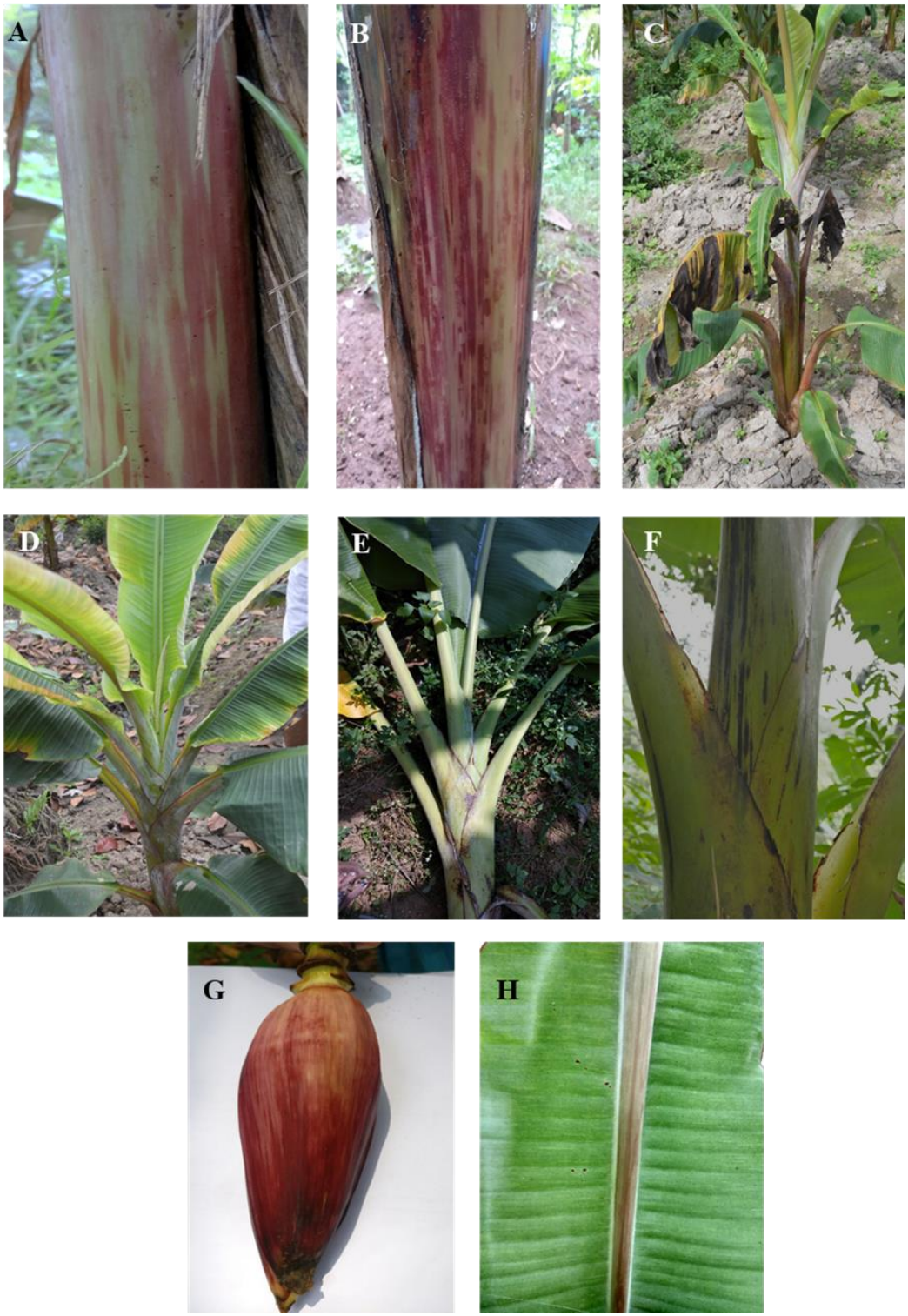


Plate 34 **Symptoms of BBrMD recorded during the survey. A** Diamond shaped dark reddish lesions on banana cv. Njalipoovan **B** Reddish streaks on pseudostem of infected plant **C** separation of old sheath from the pseudostem **D-E** Traveller's palm like arrangement of leaves at various stages of the crop **F** Black streaks on the pseudostem of infected plant **G** Mosaic patterns of purple/red on the bract H Reddish streaks on the midrib of leaf

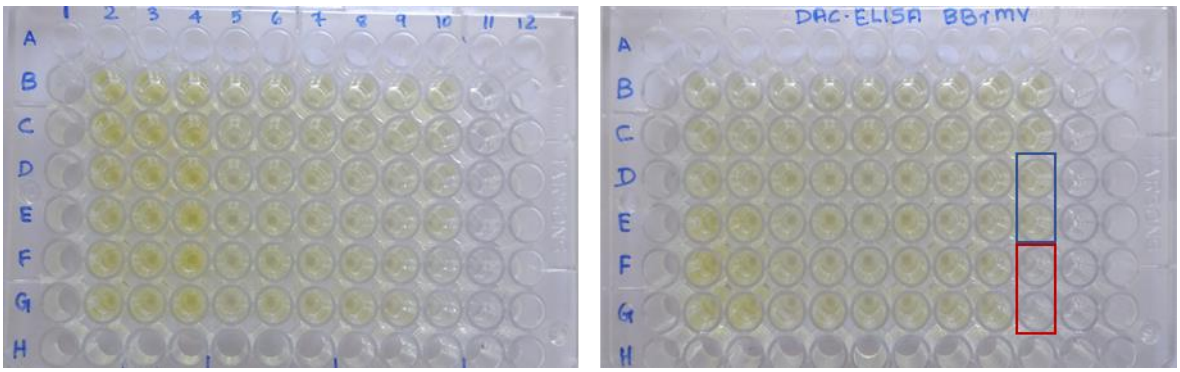


Plate 35 Determination of antiserum titre and detection of BBrMV by DAC-ELISA

A Determination of antiserum titre B1-G4: 1:200; B5-G7: 1:500; B8-G10: 1:1000

B DAC-ELISA of samples collected from field suspected to have BBrMV infection. Red squares denote negative control (NC) and blue squares denote positive control (PC)

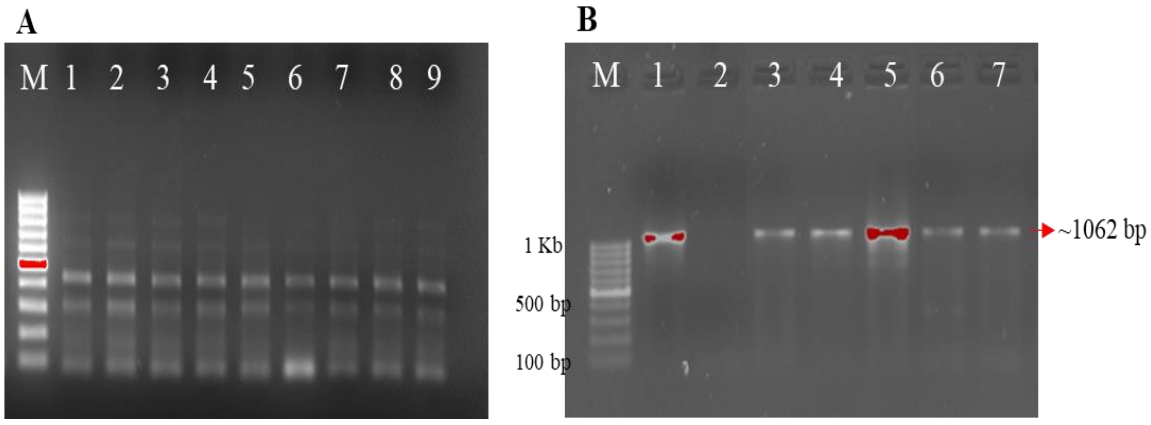
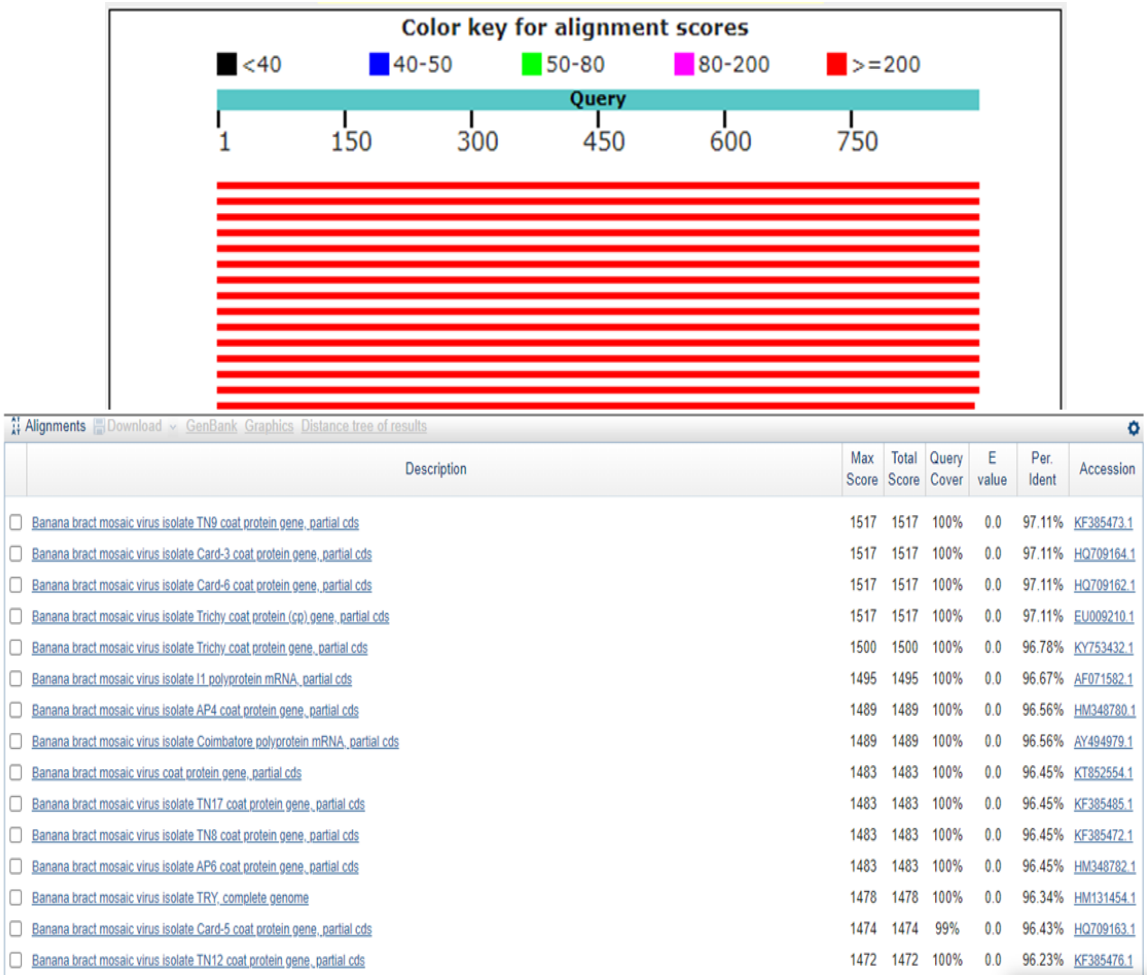


Plate 36 Total RNA isolation and RT-PCR A Total RNA isolated from the infected banana leaves electrophoretically separated in 0.8 % agarose gel **B** Reverse transcriptase-PCR using BBrMV CP specific primers to obtain 1062 bp amplicon. M: 1 kb DNA ladder Lane 1: PC, Lane 2: NC, Lanes 3-7: infected samples



Alignments | Download | GenBank | Graphics | Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Banana bract mosaic virus isolate TN9 coat protein gene, partial cds	1517	1517	100%	0.0	97.11%	KF385473.1
Banana bract mosaic virus isolate Card-3 coat protein gene, partial cds	1517	1517	100%	0.0	97.11%	HQ709164.1
Banana bract mosaic virus isolate Card-6 coat protein gene, partial cds	1517	1517	100%	0.0	97.11%	HQ709162.1
Banana bract mosaic virus isolate Trichy coat protein (cp) gene, partial cds	1517	1517	100%	0.0	97.11%	EU009210.1
Banana bract mosaic virus isolate Trichy coat protein gene, partial cds	1500	1500	100%	0.0	96.78%	KY753432.1
Banana bract mosaic virus isolate I1 polyprotein mRNA, partial cds	1495	1495	100%	0.0	96.67%	AF071582.1
Banana bract mosaic virus isolate AP4 coat protein gene, partial cds	1489	1489	100%	0.0	96.56%	HM348780.1
Banana bract mosaic virus isolate Coimbatore polyprotein mRNA, partial cds	1489	1489	100%	0.0	96.56%	AY494979.1
Banana bract mosaic virus coat protein gene, partial cds	1483	1483	100%	0.0	96.45%	KT852554.1
Banana bract mosaic virus isolate TN17 coat protein gene, partial cds	1483	1483	100%	0.0	96.45%	KF385485.1
Banana bract mosaic virus isolate TN8 coat protein gene, partial cds	1483	1483	100%	0.0	96.45%	KF385472.1
Banana bract mosaic virus isolate AP6 coat protein gene, partial cds	1483	1483	100%	0.0	96.45%	HM348782.1
Banana bract mosaic virus isolate TRY, complete genome	1478	1478	100%	0.0	96.34%	HM131454.1
Banana bract mosaic virus isolate Card-5 coat protein gene, partial cds	1474	1474	99%	0.0	96.43%	HQ709163.1
Banana bract mosaic virus isolate TN12 coat protein gene, partial cds	1472	1472	100%	0.0	96.23%	KF385476.1

Fig 26. NCBI BLAST to identify the gene. Graphical summary and descriptions of BBrMV CP gene blasted against NCBI database

4.3.5 Phylogenetic analysis of BBrMV based on CP nucleotide and amino acid sequences

From the phylogenetic analysis of the CP gene of the Kerala isolates generated in the present study, it was evident that there is no relatedness between them based on the geography (Northern, Central and Southern zones) (Fig 27A). The isolates MT818183 and MT818184 clustered together with 99 per cent bootstrap support. MT818177 and MT818187 formed outgroups with maximum branch lengths indicating that they were evolved later than all other isolates.

Phylogeny of all the 56 isolates in the dataset was analysed. Apparently, clades formed were irrespective of the host and geography, a trend similar to the Kerala isolates observed earlier (Fig 27B). Apart from MT818177 and MT818187, MT818180 three Tamil Nadu isolates *viz.*, KF385476.1, KF385478.1 and KF385484.1 formed outgroup from the other

isolates. KT456531.1 isolated from ginger clustered together with BBrMV isolates infecting *Musa* sp. from Philippines (EU414267.1, AF071585.1, AF071586.1 and AF071590.1), Thailand (AF071589.1), Vietnam (AF071588.1), India (MT818186, AF071584.1 and KY369923.1) and West Samoa (AF071587.1) with high degree of statistical confidence (Group II) signifying divergence from a common ancestor.

The tree constructed by amino acid sequence also yielded similar results but with only one major cluster (Fig 27C). Apparently, all the isolates considered in this study apart from the new Kerala isolates, MT818177, MT818187 and MT818180 are closely related.

Fig 27A. Phylogeny of Kerala isolates evaluated in MEGA X. The tree with the highest log likelihood (-3831.83) is shown. The percentage of trees in which the associated taxa clustered together (bootstrap) is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Outgroups are marked in red colour.

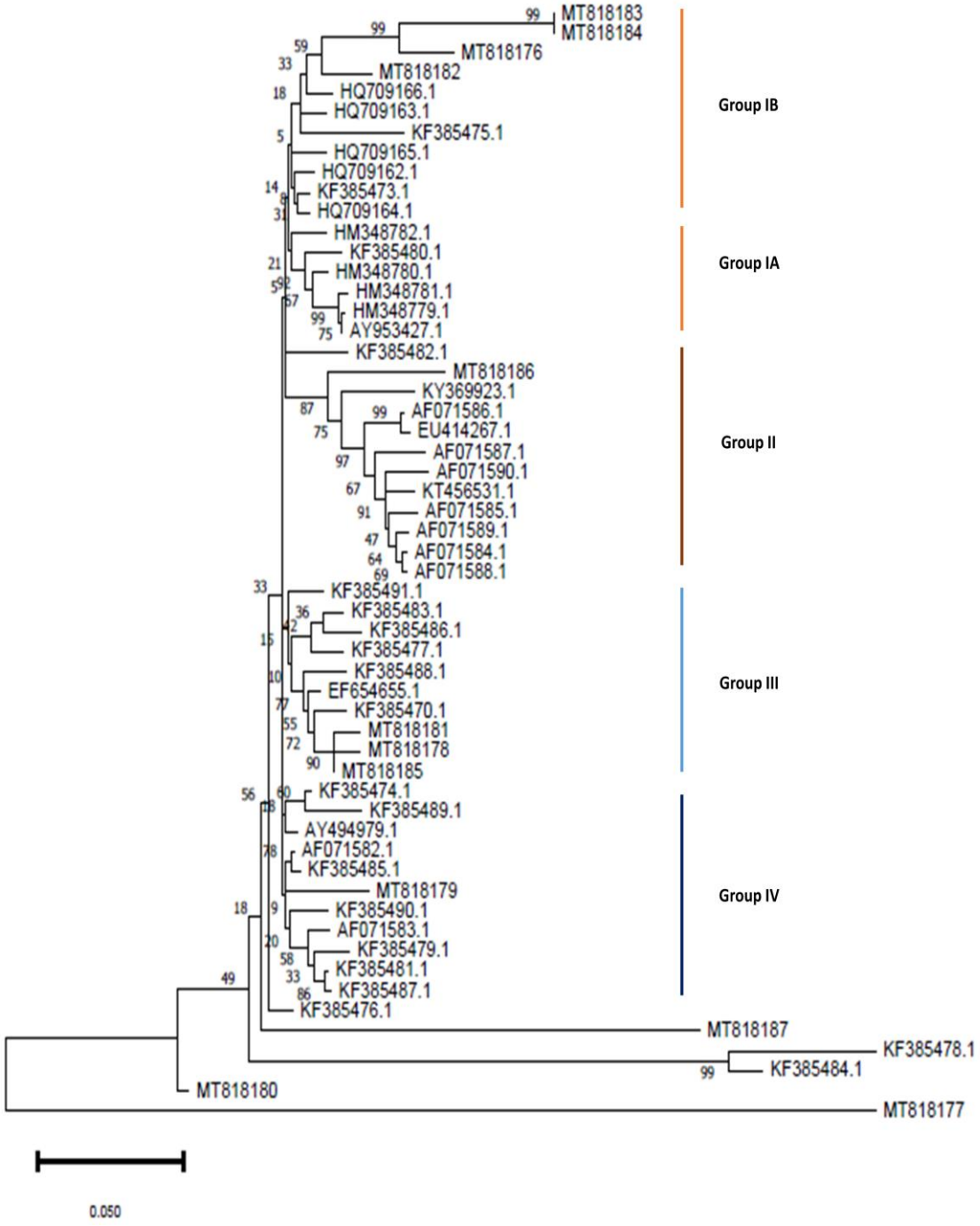


Fig 27B. Evolutionary analyses of BBrMV isolates in the dataset. The evolutionary history was inferred by Maximum Likelihood method and Kimura 2-parameter model. The tree with the highest log likelihood (-7410.50) is shown. The bootstrap values are shown next to the branches. This analysis involved 56 nucleotide sequences and a total of 917 positions in the final dataset.

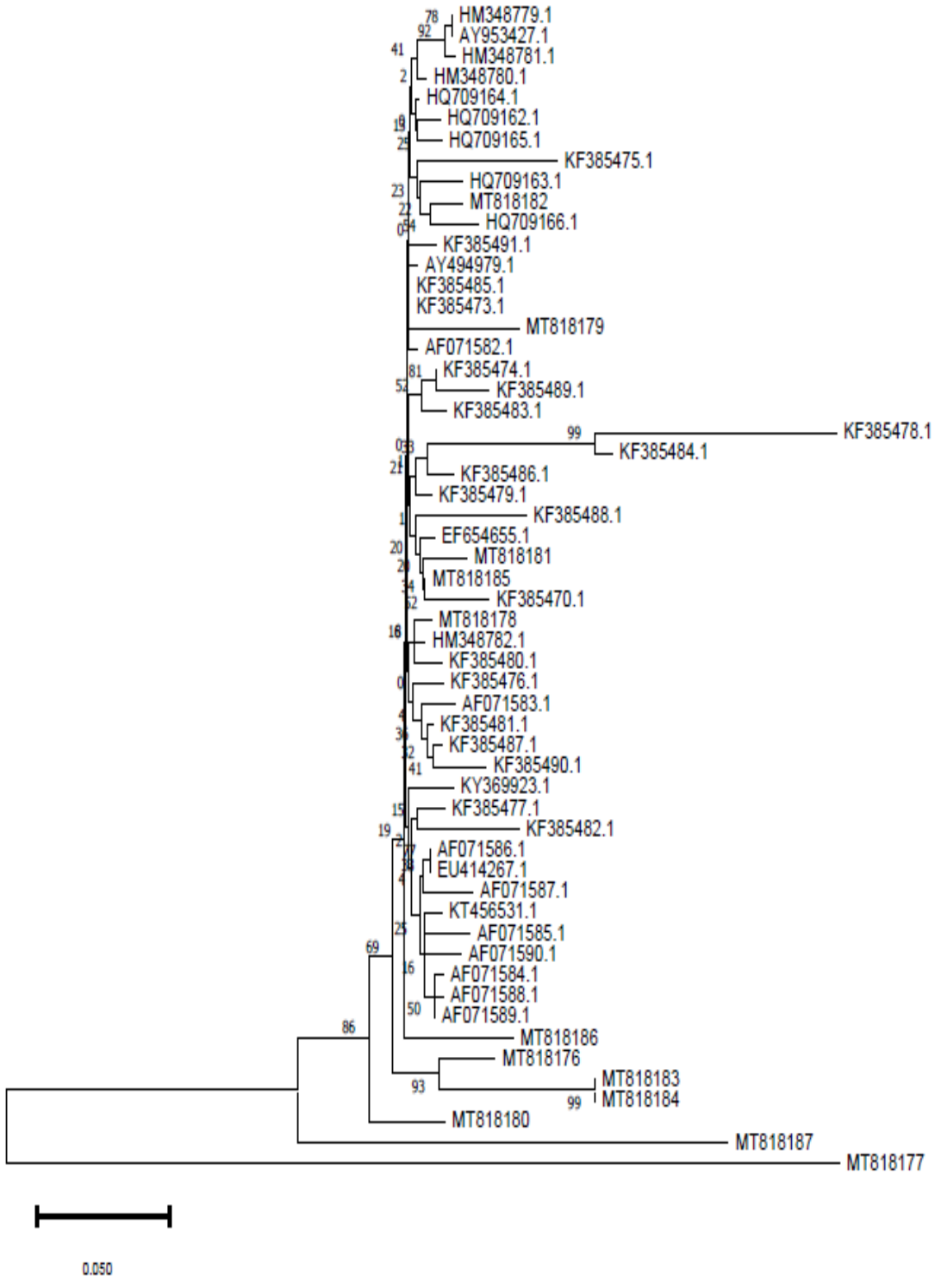


Fig 27C.Phylogenetic tree constructed using amino acid sequences of BBrMV CP. The statistical method used was Minimum evolution with 1000 bootstrap replications. Initial tree was obtained by Neighbor joining method with the sum of branch length = 1.631

4.3.6 Principal co-ordinate analysis

The result of phylogenetic analysis was substantiated by PCoA (Fig 28). The BBrMV isolates grouped irrespective of their host and geography into one major cluster. Curiously, two of the most divergent Kerala isolates were in separate cluster MT818177 (2) MT818187 (12). Apart from these, KF385478.1 (29) and KF385484.1 (35) were extremely distinct from the rest of the population that clustered together. These were collected from Arunachal Pradesh and Andaman and Nicobar Islands as part of a study conducted earlier at the same period of time. It is also evident that isolates collected from Old world and New world are not diverse rather, formed a single cluster.

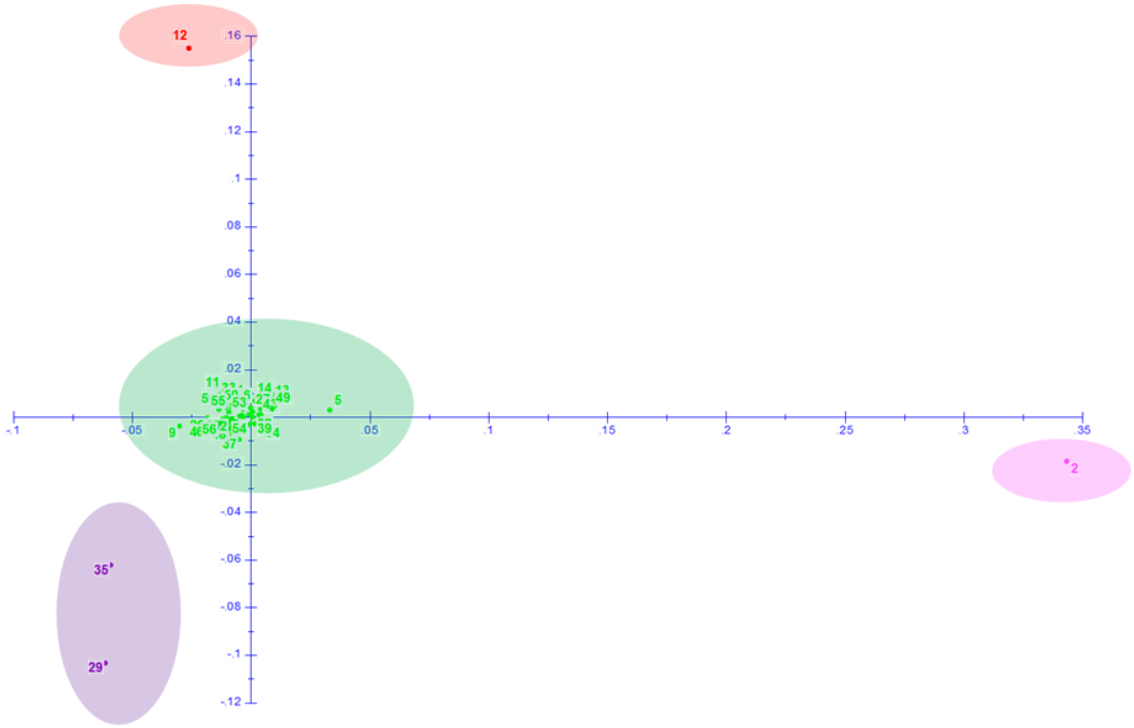


Fig 28. Principal co-ordinate analysis of 56 BBrMV isolates. Different clusters are represented in different coloured ellipses.

4.3.7 Assessment of genetic variability of BBrMV based on CP gene

4.3.7.1 Sequence homology of BBrMV CP gene

Nucleotide sequence homology between the Kerala isolates generated in this study ranged between 88-100 % (Table 17) and the amino acid sequence identity varied from 63-100 % (Fig 29). A cent per cent nucleotide and amino acid identity was observed between MT818183 and MT818184, collected from Kozhikode, a district in Northern Kerala.

However, homology of the isolates MT818177 and MT818187 was the least among the 56 isolates considered ranging between 91- 92 per cent indicating higher variability.

A high degree of divergence of these isolates from KF385478 and KF385484 (Tamil Nadu isolates) was observed. The similarity of MT818177 with KF385478 was only 76.2 %. Interestingly, the diversity of CP gene of BBrMV isolates infecting ginger and cardamom were lesser than those infecting *Musa* sp. The nucleotide sequence identity of KT456531.1, infecting flowering ginger ranged from 93-100 per cent and that infecting cardamom fell between the same range (Fig 29).

Pairwise amino acid sequence diversity was between MT818177 and MT818187 (62 %) most variable being the former with 62-66 per cent identity with other isolates. Amino acid identity of 56 sequences ranges from 62-100 % (Fig 30). This clearly indicated that the mutations in the nucleotide have been translated to the protein level indicating a low rate of amino acid conservation. The nucleotide sequences of isolates deposited earlier from Kerala were diverse from the isolates generated in this study indicating fixation of mutations over time. Thus, as expected in RNA viruses, the genetic variability between the isolates were higher.

4.3.7.2. Nucleotide diversity and mutation rate

Nucleotide diversity, Pi (π) signifies the genetic diversity in the sequence. Overall nucleotide diversity (π) of the CP gene was 0.068 and mutation rate was calculated as 0.192. From Fig 31, it is evident that the N and C terminal of the BBrMV CP gene is hypervariable irrespective of the geography or host.

Table 17. Homology matrix of BBTV Kerala isolates colour coded to differentiate most similar and dissimilar sequence pairs

	MT818180	MT818178	MT818181	MT818185	MT818186	MT818176	MT818183	MT818184	MT818182	MT818179	MT818177	MT818187
MT818180	100.00	96.60	96.10	96.80	93.40	96.00	96.80	96.80	95.60	95.20	92.20	93.10
MT818178	96.60	100.00	98.20	99.10	94.30	96.00	96.90	96.90	95.80	95.10	91.30	91.20
MT818181	96.10	98.20	100.00	99.10	93.90	95.20	96.10	96.10	95.80	94.70	91.30	91.20
MT818185	96.80	99.10	99.10	100.00	94.80	96.00	96.90	96.90	96.10	95.60	92.10	92.00
MT818186	93.40	94.30	93.90	94.80	100.00	93.40	94.30	94.30	92.80	93.10	89.20	89.20
MT818176	96.00	96.00	95.20	96.00	93.40	100.00	97.00	97.00	95.80	94.80	91.30	90.90
MT818183	96.80	96.90	96.10	96.90	94.30	97.00	100.00	100.00	96.60	95.70	92.20	91.80
MT818184	96.80	96.90	96.10	96.90	94.30	97.00	100.00	100.00	96.60	95.70	92.20	91.80
MT818182	95.60	95.80	95.80	96.10	92.80	95.80	96.60	96.60	100.00	94.80	91.30	90.90
MT818179	95.20	95.10	94.70	95.60	93.10	94.80	95.70	95.70	94.80	100.00	90.70	90.40
MT818177	92.20	91.30	91.30	92.10	89.20	91.30	92.20	92.20	91.30	90.70	100.00	88.20
MT818187	93.10	91.20	91.20	92.00	89.20	90.90	91.80	91.80	90.90	90.40	88.20	100.00

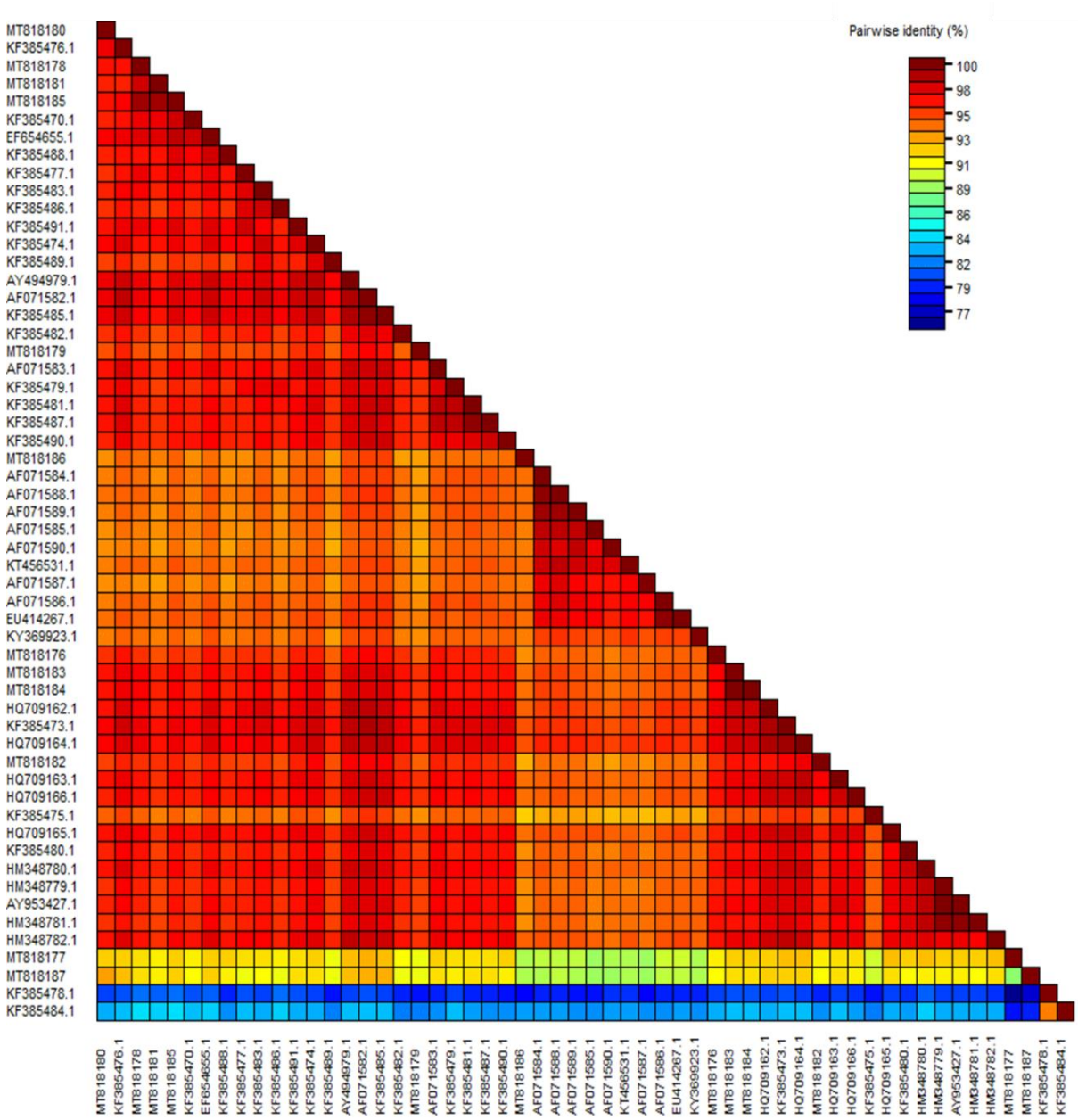


Fig 29. Pairwise sequence identity heat map of BBrMV CP gene of all the isolates in the dataset

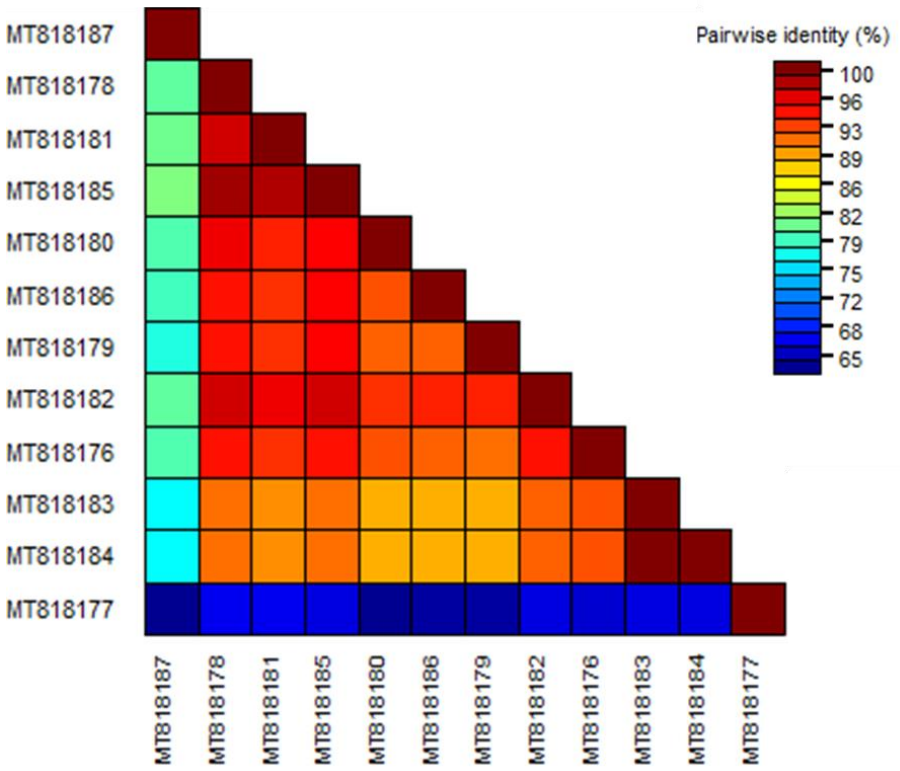


Fig 30. Amino acid sequence identity heat map of coding region of BBrMV CP Kerala isolates

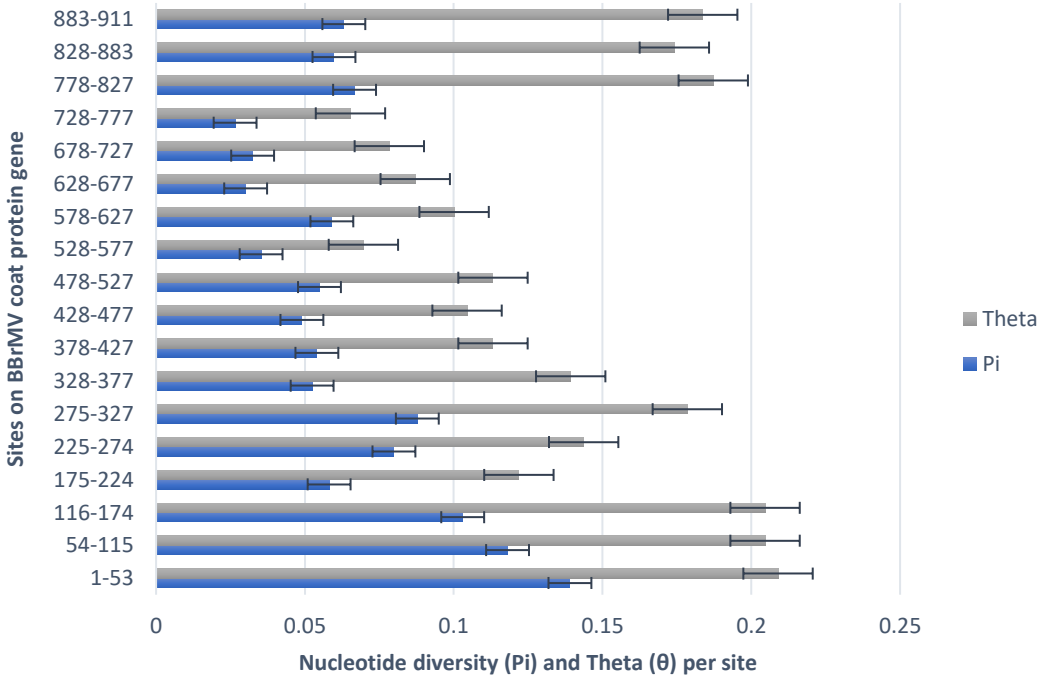


Fig 31. Nucleotide diversity and mutation rate. Estimation was based on total number mutations per site. The result is interpreted in window size 50. Error bar at 5 % is displayed on the corresponding bar

The core region is apparently stable. The Pi value for N terminal alone (1- 60 bp) was 0.142 and that of C terminal (884-900 bp) was 0.063. The mutation rate in N and C terminal was 0.194 and 0.184 respectively. The nucleotide diversity of the core region was 0.058 and mutation rate was calculated to be 0.126.

A total number of 782 mutations were detected throughout the dataset. The core region between 363- 832 bp was detected to be highly conserved with statistically significant conservation value of 0.545. Sixty In-Del sites and 552 polymorphic sites were detected throughout the CP gene. The nucleotide diversity among the Kerala isolates generated in the study was compared to old submissions in NCBI. It was evident that the former was more diverse than the latter with Pi 0.119 which was much larger than that of the latter (0.031). The π and θ values at the N terminal were higher than the average of the total dataset indicating a higher level of mutation (Fig 32). Remarkably, the nucleotide diversity among the cardamom isolates were very less (0.024) indicating less diversity in the population when compared to those infecting *Musa* sp.

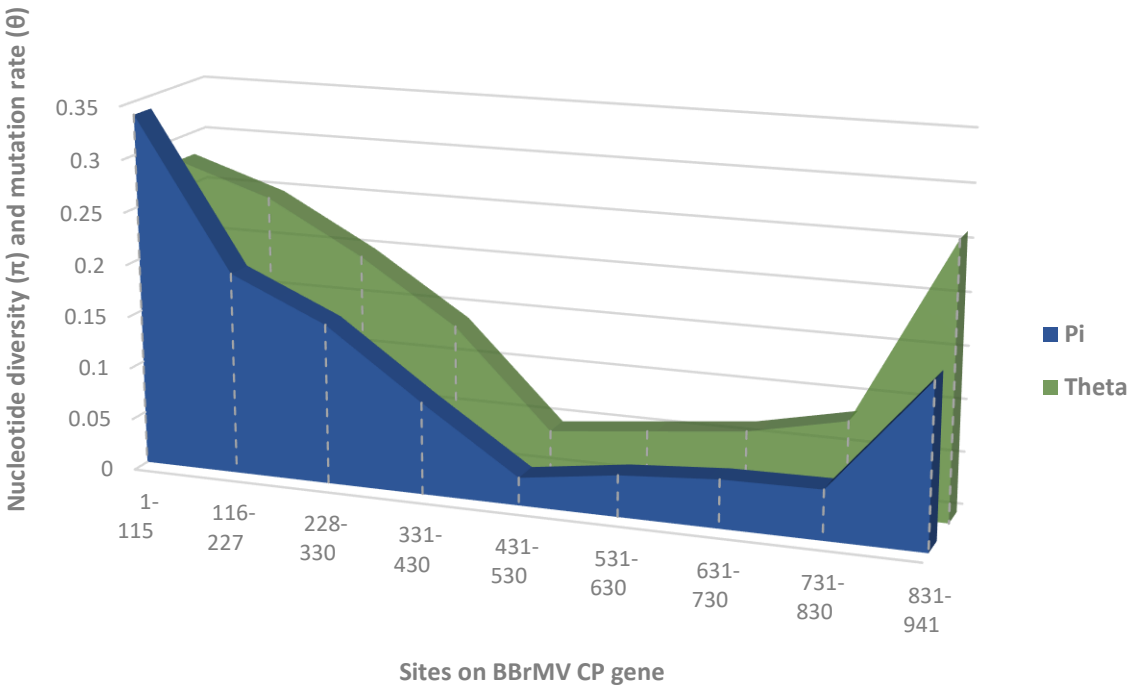


Fig 32. Nucleotide diversity and mutation rate of coat protein gene of BBTV Kerala isolates

4.3.7.3 Synonymous and non-synonymous substitutions

Most of the substitutions were non-synonymous and Ka/Ks of most of the sequence pairs were lesser than 1 indicating negative or purifying selections (Fig 33). Few values were close to 1 indicative of neutral selection of amino acids in BBrMV CP. The three sequence pairs *viz.*, MT818181-MT818185 (Kerala isolates generated in the present study), KF385488.1 (China) - KF385470.1 (India) and KF385488.1 -EF654655.1 (Pakistan) had Ka/Ks value more than 1. This is indicative of positive selection and that the new genotypes are having fitness advantage.

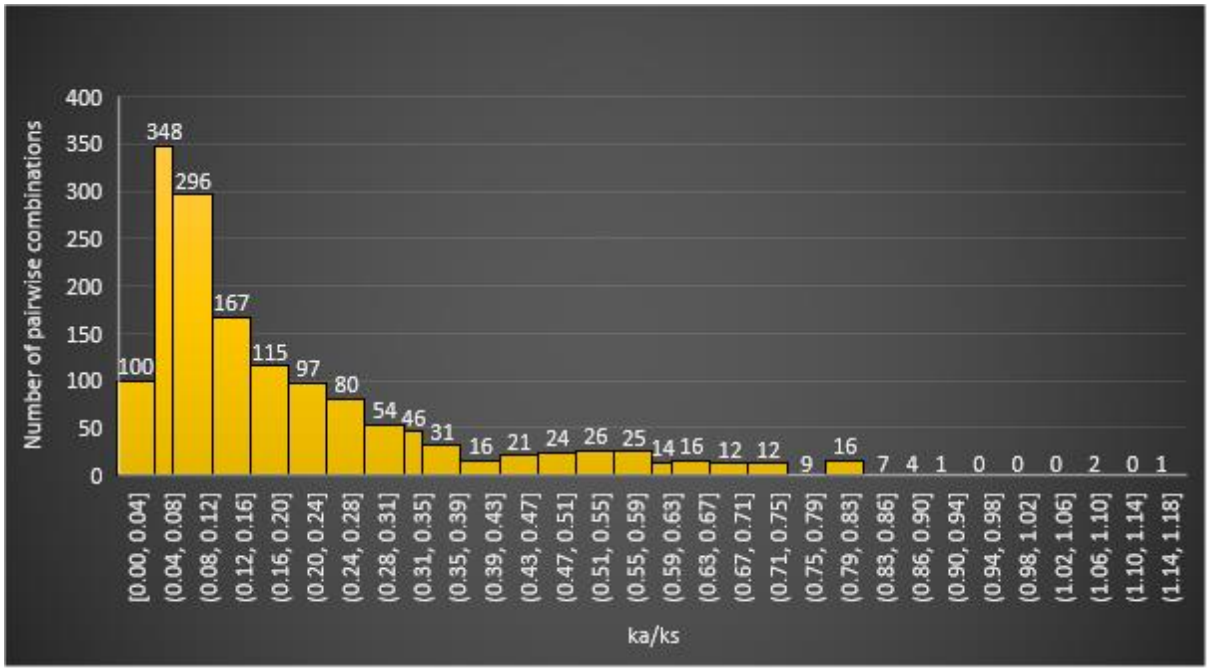


Fig 33. Histogram depicting the range of Ka/Ks in pairwise combinations of 56 BBrMV isolates in the dataset

All the three statistical tests *viz.,* Tajima's D test, Fu and Li's F and D tests were performed using total number of mutations. Tajima's D test gave a statistically significant negative value (-2.33) which signifies abundance of rare alleles in the population. The result implied that the population may have been subjected to a recent selective sweep, population expansion after a recent bottleneck. Few codons with multiple evolutionary paths were detected during the analysis. Tajima's D value at synonymous and non-synonymous sites were negative but non significant in former and highly significant in the latter. Fu and Li statistics (Fu and Li' s D and F test) also substantiated this result as the values were also negative and highly significant.

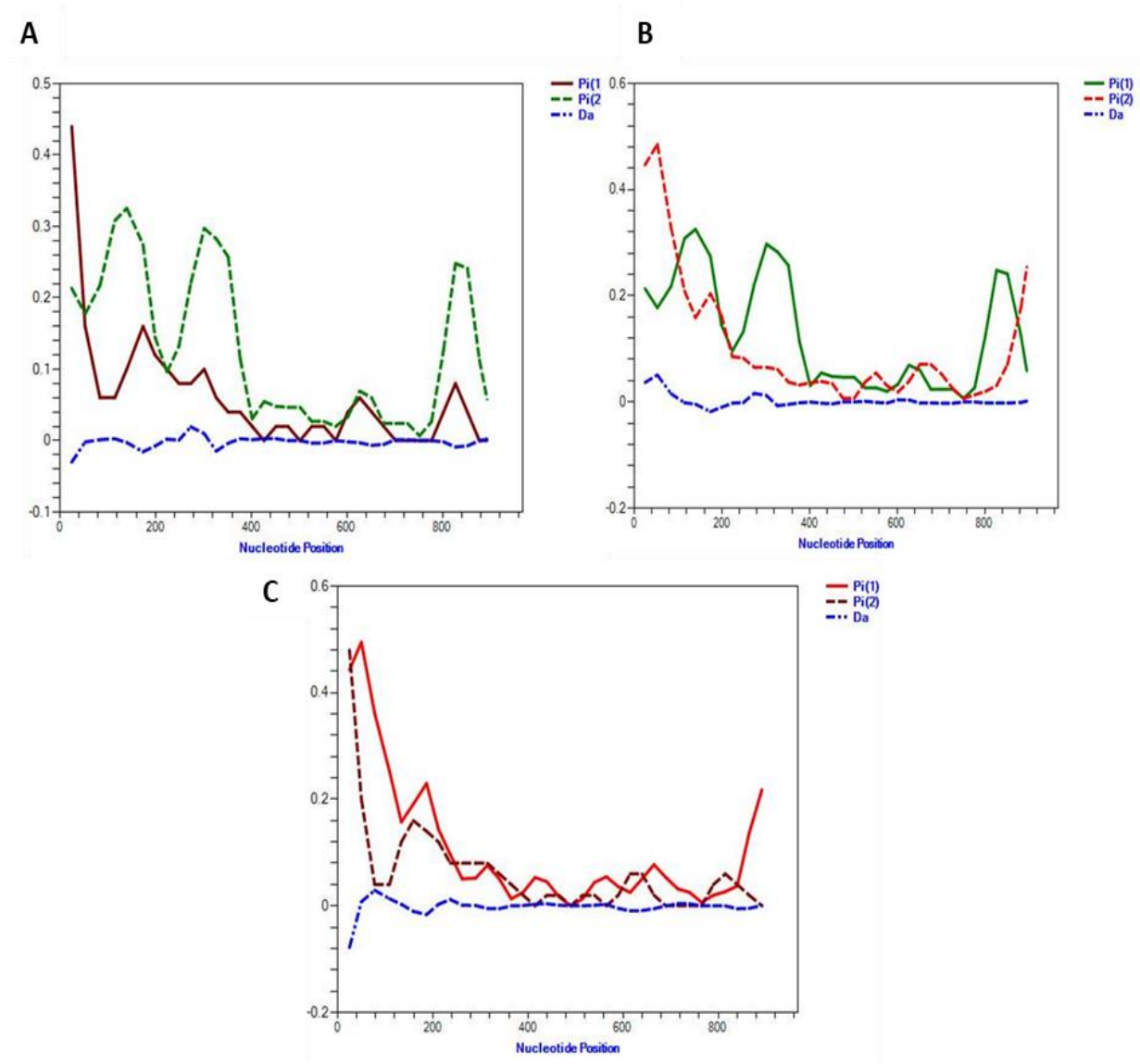


Fig 34. Comparing the divergence of BBrMV population infecting banana in Kerala based on the CP gene. Comparison between isolates from **A** Southern (Maroon) and Central (Green) zones **B** Central and Northern (Red) zones and **C** Northern and Southern zones of Kerala.

The DNA diversity between subpopulations of Kerala based on geography (North, South and Central zones) were calculated. The number of net nucleotide substitutions per site between populations (Da) was also represented (Fig 34). From the figure it was evident that N terminal of BBrMV CP was diverse between the population and among the population itself. Core region is highly conserved except for central zone isolates. Number of shared mutations between isolates from Southern- Central, Central-Northern and Northern- Southern zones were 19, 55 and 32 respectively.

4.3.8 Gene flow between population subsets

The gene flow between subpopulations made based on geography and host were calculated. Since only single sequences were available in Northern Kerala subset and flowering ginger subset, they were excluded in the analysis. Pairwise low fixation index (F_{ST}) indicated that there is no genetic divergence between the subpopulations of Kerala (Table 18). Fixation index (F_{ST}) between Southern states of India *viz.*, Kerala, Tamil Nadu, Karnataka and Andhra Pradesh were also ~zero indicative of frequent gene flow. Remarkably, even the isolates infecting cardamon and banana showed similar trend. The G_{ST} of Andhra Pradesh and Karnataka was zero indicating complete panmixis. The isolates collected from North and Central Kerala depicted a very low value of F_{ST} and N_{ST}. Similar trend was observed in the case of Kerala and Tamil Nadu isolates implying that there is a frequent gene flow through the planting materials. This substantiates the clustering of most of the South Indian isolates together as they seem to be diverging from a common ancestor.

Although the values were much lesser than 1, the highest F_{ST} and N_{ST} values were depicted by BBrMV isolates infecting banana in India and Philippines. The highest net nucleotide substitutions (Da) observed were also between these subpopulations.

4.3.9 Codon usage bias

Nucleotide composition of BBrMV CP gene was calculated in CAI-CAL web server (Table 19). Considering the entire data set, the gene was predominantly AU rich (55.44 ± 0.699). The percentage of adenine (% A) and uracil (% U) were highest with an overall average of 30.93 ± 0.420 and 24.50 ± 0.744 respectively followed by % G and % C (24.04 ± 0.349 and 20.55 ± 0.538) (Table 18). Individually, the isolates infecting banana and cardamom showed a similar trend in the nucleotide composition of the CP gene. Although, CP of flowering ginger isolate, % G was higher than % U since only one isolate is available, validation of the deviation from the common pattern cannot be ascertained with statistical support. Pima facie, the nucleotide constitution was observed to play an important role in the codon usage. The percentage of A+U at the wobble position (AU3) was the highest with an average of 63.90 ± 1.817 %. Per cent G+C at the wobble position (GC3) was the least (36.10 ± 1.817 %).

Table 18. Gene flow and differentiation between the population subsets of BBrMV isolates based on geography and host

Population 1	Population 2	Hs	Ks	G_{ST}	N_{ST}	F_{ST}	Dxy	Da
North Kerala	Central Kerala	0.957	107.773	0.026	0.044	0.034	0.126	0.004
Kerala	Tamil Nadu	0.996	71.967	0.002	0.037	0.035	0.086	0.003
Kerala	Andhra Pradesh	0.994	67.703	0.013	0.128	0.143	0.069	0.010
Kerala	Karnataka	0.994	69.394	0.013	0.030	0.040	0.065	0.003
Tamil Nadu	Karnataka	1.000	48.880	0.013	0.028	0.026	0.047	0.001
South India	North India	0.843	2.412	0.020	0.016	0.016	0.006	0.000
South India	North East India	0.891	4.441	0.038	0.185	0.184	0.020	0.004
North India	North East India	0.857	8.071	0.077	0.183	0.181	0.021	0.004
Tamil Nadu	Andhra Pradesh	1.000	47.263	0.013	0.162	0.167	0.051	0.009
Andhra Pradesh	Karnataka	1.000	19.900	0.000	0.217	0.214	0.029	0.006
Banana India	Philippines	0.999	61.963	0.045	0.297	0.300	0.073	0.022
BBrMV infecting cardamom	BBrMV infecting Banana	0.999	60.028	0.036	0.076	0.084	0.050	0.004
Kerala isolates infecting Banana	South Indian isolates infecting cardamom	0.999	76.241	0.016	0.090	0.107	0.073	0.008

Hs denotes the genetic diversity within the population, **Ks** denotes the rate of synonymous substitutions per site, genetic differentiation is denoted by **N_{ST}** and **G_{ST}** values and gene flow is denoted by fixation index, **F_{ST}**. **Dxy** is the average number of nucleotide substitutions per site between populations and **Da** denotes the number of net nucleotide substitutions per site between populations. All the values were estimated using the DnaSP6 program with 1,000 permutation tests

Table 19. Nucleotide composition, effective number of codons and codon adaptability index of CP gene of BBrMV isolates

Parameter / Sequences	Length	%A	%C	%U	%G	%GC	%AU	%GC3	%GA3	%GU3	%AC3	%CU3	%AU3	Nc	CAI
BBrMV isolates infecting banana															
MT818176	900	31.00	20.56	24.33	24.11	44.70	55.33	37.33	52.00	56.67	43.33	48.00	62.67	56.50	0.715
MT818177	888	29.73	19.26	27.14	23.87	43.10	56.87	41.22	49.32	54.73	45.27	50.68	58.78	58.40	0.720
MT818178	900	30.78	20.33	24.89	24.00	44.30	55.67	36.00	50.67	57.33	42.67	49.33	64.00	52.70	0.717
MT818179	900	30.56	20.11	25.78	23.56	43.70	56.33	35.00	50.33	57.33	42.67	49.67	65.00	51.10	0.716
MT818180	903	31.34	20.04	25.03	23.59	43.60	56.37	35.22	50.83	55.15	44.85	49.17	64.78	53.50	0.718
MT818181	900	31.33	20.11	24.78	23.78	43.90	56.11	35.00	50.67	57.00	43.00	49.33	65.00	52.50	0.720
MT818182	900	30.78	21.00	24.22	24.00	45.00	55.00	38.00	51.00	55.67	44.33	49.00	62.00	55.60	0.717
MT818183	903	31.45	20.49	24.47	23.59	44.10	55.92	36.54	53.49	55.81	44.19	46.51	63.46	54.60	0.704
MT818184	903	31.45	20.49	24.47	23.59	44.10	55.92	36.54	53.49	55.81	44.19	46.51	63.46	54.60	0.704
MT818185	900	31.00	20.33	24.89	23.78	44.10	55.89	35.00	50.67	57.00	43.00	49.33	65.00	52.30	0.721
MT818186	900	31.33	20.56	24.22	23.89	44.40	55.56	36.33	52.00	56.33	43.67	48.00	63.67	54.40	0.705
MT818187	909	30.03	18.92	27.28	23.76	42.70	57.32	38.61	47.19	55.78	44.22	52.81	61.39	55.60	0.722
AF071582.1	900	30.78	20.33	24.56	24.33	44.70	55.33	35.67	51.67	56.67	43.33	48.33	64.33	53.60	0.713
AF071583.1	900	30.67	20.33	24.67	24.33	44.70	55.33	35.00	51.33	56.33	43.67	48.67	65.00	54.10	0.712
AF071584.1	900	30.67	21.67	23.22	24.44	46.10	53.89	39.33	51.67	53.67	46.33	48.33	60.67	54.40	0.724
AF071585.1	900	30.44	21.11	23.78	24.67	45.80	54.22	38.00	51.67	55.00	45.00	48.33	62.00	56.00	0.714
AF071586.1	900	30.78	20.89	23.89	24.44	45.30	54.67	37.67	52.00	56.33	43.67	48.00	62.33	55.60	0.716
AF071587.1	900	31.33	21.44	23.44	23.78	45.20	54.78	38.00	51.67	53.67	46.33	48.33	62.00	55.50	0.712
AF071588.1	900	30.89	21.56	23.22	24.33	45.90	54.11	39.00	51.67	54.00	46.00	48.33	61.00	54.10	0.722
AF071589.1	900	30.67	21.67	23.11	24.56	46.20	53.78	40.00	52.00	54.00	46.00	48.00	60.00	53.80	0.726

Parameter / Sequences	Length	%A	%C	%U	%G	%GC	%AU	%GC3	%GA3	%GU3	%AC3	%CU3	%AU3	Nc	CAI
AF071590.1	900	30.67	21.67	23.22	24.44	46.10	53.89	40.00	51.00	53.00	47.00	49.00	60.00	57.10	0.725
KY369923.1	900	31.11	21.00	23.89	24.00	45.00	55.00	36.00	51.67	55.00	45.00	48.33	64.00	53.70	0.708
KF385470.1	900	30.67	21.00	24.67	23.67	44.70	55.33	36.00	50.33	55.00	45.00	49.67	64.00	55.00	0.722
KF385473.1	900	30.78	20.44	24.44	24.33	44.80	55.22	36.33	51.67	56.67	43.33	48.33	63.67	54.60	0.716
KF385474.1	900	30.78	20.11	24.89	24.22	44.30	55.67	35.33	51.33	57.33	42.67	48.67	64.67	52.30	0.712
KF385475.1	900	31.00	20.56	24.33	24.11	44.70	55.33	35.67	50.67	55.67	44.33	49.33	64.33	57.00	0.714
KF385476.1	900	31.33	19.78	25.11	23.78	43.60	56.44	33.00	51.33	57.00	43.00	48.67	67.00	52.80	0.714
KF385477.1	900	30.56	20.44	24.78	24.22	44.70	55.33	36.00	51.00	57.00	43.00	49.00	64.00	50.40	0.727
KF385478.1	900	31.56	20.67	24.22	23.56	44.20	55.78	32.00	52.00	51.33	48.67	48.00	68.00	50.50	0.709
KF385479.1	900	31.11	20.56	24.22	24.11	44.70	55.33	35.33	51.67	55.00	45.00	48.33	64.67	51.50	0.718
KF385480.1	900	30.89	20.33	25.00	23.78	44.10	55.89	34.33	51.00	56.67	43.33	49.00	65.67	51.40	0.714
KF385481.1	900	30.89	20.78	24.22	24.11	44.90	55.11	36.33	51.33	55.00	45.00	48.67	63.67	54.80	0.719
KF385482.1	900	30.56	19.78	24.89	24.78	44.60	55.44	35.33	51.00	57.67	42.33	49.00	64.67	53.30	0.711
KF385483.1	900	30.67	20.67	24.33	24.33	45.00	55.00	36.00	51.00	56.33	43.67	49.00	64.00	53.90	0.715
KF385484.1	900	32.78	21.11	23.44	22.67	43.80	56.22	31.00	51.67	49.33	50.67	48.33	69.00	51.60	0.696
KF385485.1	900	30.78	20.56	24.33	24.33	44.90	55.11	36.67	51.67	56.33	43.67	48.33	63.33	53.40	0.718
KF385486.1	900	31.00	20.44	24.44	24.11	44.60	55.44	35.00	51.00	56.00	44.00	49.00	65.00	51.60	0.719
KF385487.1	900	30.67	20.89	24.11	24.33	45.20	54.78	37.00	51.33	55.00	45.00	48.67	63.00	55.30	0.720
KF385488.1	900	31.11	19.89	24.78	24.22	44.10	55.89	36.00	51.33	56.00	44.00	48.67	64.00	53.00	0.726
KF385489.1	900	31.11	20.11	24.89	23.89	44.00	56.00	33.67	51.00	56.67	43.33	49.00	66.33	51.90	0.706
KF385490.1	900	30.67	20.22	24.89	24.22	44.40	55.56	35.00	51.67	56.00	44.00	48.33	65.00	54.10	0.713
KF385491.1	900	30.67	20.22	24.89	24.22	44.40	55.56	35.00	51.00	58.00	42.00	49.00	65.00	53.10	0.715
HM348780.1	900	31.22	20.56	24.44	23.78	44.30	55.67	35.67	51.00	55.33	44.67	49.00	64.33	53.60	0.719
HM348779.1	900	31.33	20.11	24.78	23.78	43.90	56.11	35.00	51.00	56.00	44.00	49.00	65.00	52.60	0.715
HM348782.1	900	30.89	20.78	24.22	24.11	44.90	55.11	36.33	51.33	55.00	45.00	48.67	63.67	54.80	0.720

Parameter / Sequences	Length	%A	%C	%U	%G	%GC	%AU	%GC3	%GA3	%GU3	%AC3	%CU3	%AU3	Nc	CAI
HM348781.1	900	31.33	20.22	24.56	23.89	44.10	55.89	34.67	51.00	56.33	43.67	49.00	65.33	53.90	0.710
AY953427.1	900	31.33	20.00	24.89	23.78	43.80	56.22	34.67	51.00	56.33	43.67	49.00	65.33	52.30	0.713
EF654655.1	900	31.22	20.56	24.33	23.89	44.40	55.56	35.67	51.33	55.67	44.33	48.67	64.33	54.40	0.718
AY494979.1	900	30.89	20.22	24.78	24.11	44.30	55.67	35.00	51.33	57.67	42.33	48.67	65.00	53.70	0.713
EU414267.1	900	30.78	20.44	24.33	24.44	44.90	55.11	36.33	52.00	57.67	42.33	48.00	63.67	55.10	0.713
Average		**30.947**	**20.506**	**24.515**	**24.032**	**44.540**	**55.461**	**36.076**	**51.260**	**55.726**	**44.274**	**48.740**	**63.924**	**53.832**	**0.716**
SD		**0.440**	**0.552**	**0.775**	**0.361**	**0.716**	**0.717**	**1.888**	**0.889**	**1.583**	**1.583**	**0.889**	**1.888**	**1.706**	**0.006**
BBrMV isolates infecting cardamom															
HQ709162.1	900	30.89	20.56	24.33	24.22	44.80	55.22	36.33	51.33	57.00	43.00	48.67	63.67	54.80	0.722
HQ709163.1	900	30.89	20.44	24.67	24.00	44.40	55.56	35.67	51.33	56.33	43.67	48.67	64.33	54.00	0.712
HQ709164.1	900	30.89	20.89	24.11	24.11	45.00	55.00	37.33	51.33	55.33	44.67	48.67	62.67	55.20	0.718
HQ709165.1	900	30.78	20.00	25.00	24.22	44.20	55.78	35.00	51.33	57.67	42.33	48.67	65.00	53.30	0.707
HQ709166.1	900	31.00	20.78	24.44	23.78	44.60	55.44	35.67	50.67	55.67	44.33	49.33	64.33	53.40	0.719
Average		**30.89**	**20.53**	**24.51**	**24.07**	**44.60**	**55.40**	**36.00**	**51.20**	**56.40**	**43.60**	**48.80**	**64.00**	**54.14**	**0.72**
SD		**0.07**	**0.31**	**0.30**	**0.17**	**0.28**	**0.27**	**0.79**	**0.27**	**0.85**	**0.85**	**0.27**	**0.79**	**0.75**	**0.01**
BBrMV isolate infecting flowering ginger															
KT456531.1	900	30.56	21.22	23.78	24.44	45.70	54.33	38.00	51.33	54.67	45.33	48.67	62.00	55.50	0.718

The effective number of codon (Nc) of the gene was estimated to be 53. 89 ± 1.645. Although a high number, this value indicated presence of codon bias in the gene. The Nc of isolates infecting banana, cardamom and flowering ginger were 53.83 ± 1.706, 54.14 ± 0.75 and 55.50 respectively. There was no significant difference between the Nc values of the isolates infecting different hosts. However, Nc values ranged from 50.40 to 58.40, lowest value recorded in KF385477.1 and the highest in MT818177. The latter, Kerala isolate from Ernakulam (Central zone) was found to be the isolate that has been diverged from most common ancestor and according to the branch length, apparently was evolved recently. This can be indicative of the isolates evolving in such a way to not show codon bias and all the synonymous codons are used equally in the gene. The least Nc value among the Kerala isolates was recorded in the MT818179, collected from Thrissur (51.10).

Correlation analysis was conducted to statistically prove the role of nucleotide composition in CUB (Table 20). From the analysis it is evident that Nc value is positively correlated to percentage of C and G in the wobble position individually and together and negatively correlated to percentage of A and U in the wobble position and AU3. The nucleotide composition, as expected, had a determining role in the codon bias of the BBrMV CP.

Another parameter to assess codon bias was relative synonymous codon usage (RSCU). Evidently, RSCU values of codons corroborated the results from Nc (Table 21). From the data, apparently, bias is formed by 30 desirable codons out of which 12 high were frequency codons with RSCU value greater than 1.40 (UUA, UUG, AUU, UCA, CCA, CGU, GGU, GGA, ACA, GCU, GCA, GAU). Interestingly, 11 out of the high frequency codons has A or U in the wobble position indicative of the role of nucleotide composition in codon bias (result shown earlier).

The RSCU values of the Kerala isolates with Nc value highest and lowest (MT818177 and MT818179) were compared to confirm the codon bias pattern. Most of the high frequency codons were similar in both the isolates (Fig 35). However, codons UAU, AAA and UGU were not Hfc in the former. Curiously, ACU, AGA and CGU had become high frequency codons in MT818177. The data suggest a different pattern of evolution in this isolate which is also found to be one of the most genetically diverse isolates in the entire dataset.

Another diverse isolate generated in this study, MT818187, collected from Kasaragod, also showed very interesting codon usage pattern although the Nc value of the isolate was close

Table 20. Correlation of nucleotides compositions and effective number of codons of BBrMV CP

	Nc	*%GC*	*%A*	*%C*	*%U*	*%G*	*%A3*	*%C3*	*%U3*	*%G3*	*%GC3*	*%GA3*	*%GU3*	*%AU3*
Nc	1.000													
%GC	0.306	1.000												
%A	-0.406	-0.188	1.000											
%C	0.195	0.871**	0.177	1.000										
%U	-0.062	-0.833	-0.387	-0.919**	1.000									
%G	0.319	0.656**	-0.652	0.204	-0.249*	1.000								
%A3	-0.472**	-0.058	0.840**	0.265	-0.417**	-0.530**	1.000							
%C3	0.581**	0.352	-0.362	0.387	-0.130	0.114	-0.299	1.000						
%U3	-0.542**	-0.633	0.064	-0.707**	0.562**	-0.184	-0.010	-0.792	1.000					
%G3	0.477**	0.381	-0.596**	0.065	-0.025	0.670**	-0.763**	0.088	-0.218	1.000				
%GC3	0.721**	0.495**	-0.637	0.321	-0.110	0.505**	-0.696**	0.779**	-0.710**	0.693**	1.000			
%GA3	-0.068	0.429**	0.465*	0.489**	-0.666**	0.106	0.481*	-0.333	-0.311	0.200	-0.115	1.000		
%GU3	-0.105	-0.253	-0.389	-0.551**	0.458**	0.343	-0.577*	-0.607*	0.686**	0.560**	-0.086	-0.115	1.000	
%AU3	-0.721**	-0.495**	0.637**	-0.321	0.110	-0.504**	0.696*	-0.778*	0.710**	-0.693*	-1.000**	0.115	0.086	1.000

* Correlation is significant at the 0.05 level;

** Correlation is significant at the 0.01 level (two-tailed). Significant values are highlighted with blue colour.

Table 21. Relative Synonymous Codon Usage of BBrMV CP. High frequency codons are indicated by (*) next to RSCU value. Desirable codons are indicated by (#) next to the codon.

Sl No.	Amino Acid	Codon	RSCU	Sl No.	Amino Acid	Codon	RSCU
1.	Phenylalanine	UUU#	1.385	10.	Proline	CCU#	1.211
		UUC	0.615			CCC	0.722
2.	Leucine	UUA#	1.518*			CCA#	1.785*
		UUG#	2.186*			CCG	0.282
		CUU#	1.201	11.	Arginine	CGU#	1.409*
		CUC	0.445			CGC	0.406
		CUA	0.230			CGA	0.775
		CUG	0.420			CGG	0.767
3.	Isoleucine	AUU#	1.544*			AGA#	1.315
		AUC	0.578			AGG#	1.327
		AUA	0.878	12.	Glycine	GGU#	1.514*
4.	Valine	GUU#	1.358			GGC	0.234
		GUC#	1.124			GGA#	1.519*
		GUA#	1.008			GGG	0.732
		GUG	0.509	13.	Tryptophan	ACU#	1.331
5.	Serine	UCU#	1.136			ACC	0.379
		UCC	0.298			ACA#	1.471*
		UCA#	2.511*			ACG	0.819
		UCG	0.054	14.	Alanine	GCU#	1.647*
		AGU	0.819			GCC	0.485
		AGC#	1.183			GCA#	1.810*
6.	Tyrosine	UAU#	1.300			GCG	0.058
		UAC	0.700	15.	Aspartic acid	GAU#	1.544*
7.	Histidine	CAU#	1.193			GAC	0.456
		CAC	0.807	16.	Glutamic acid	GAA	0.729
8.	Asparagine	AAU#	1.266			GAG#	1.271
		AAC	0.734	17.	Cysteine	UGU#	1.167
9.	Lysine	AAA#	1.292			UGC	0.833
		AAG	0.708	18.	Glutamine	CAA#	1.188
						CAG	0.812

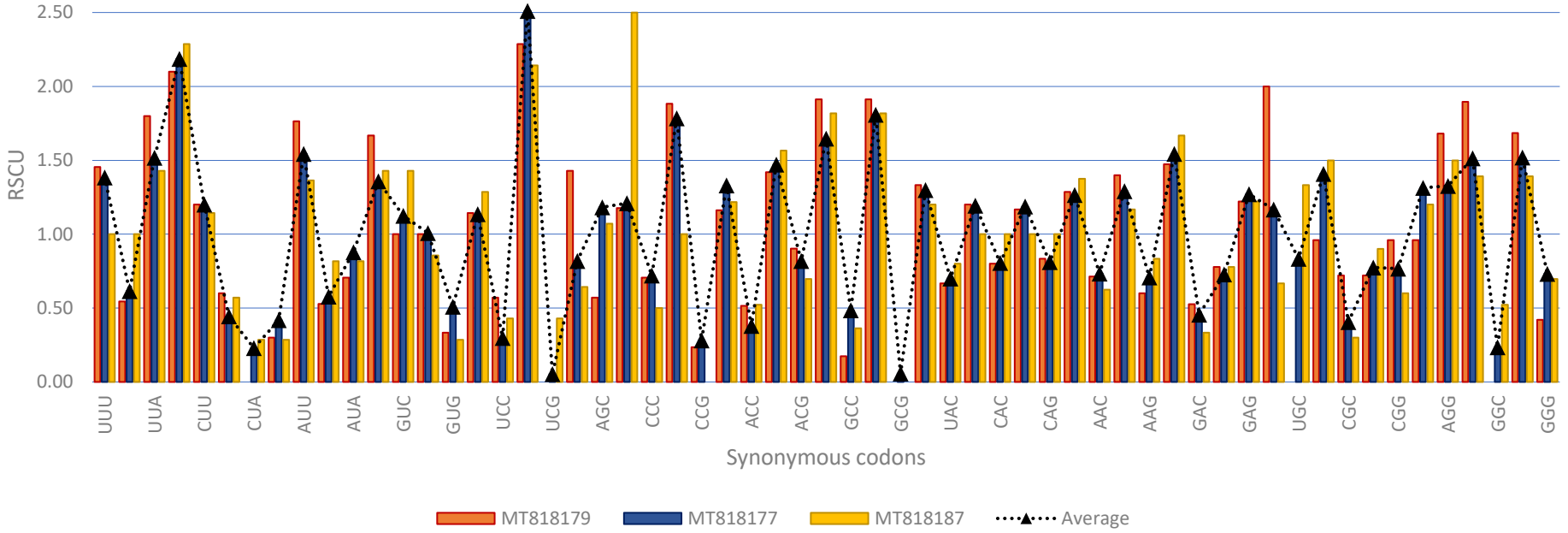


Fig 35. Comparison of RSCU of MT818177 and MT818187 with MT818179. The average RSCU of the dataset is depicted as line graph in the plot.

to the average Nc value of the dataset (55.60). RSCU of CCU coding for proline used in a higher frequency which otherwise was considered desirable but low frequency codon (Fig 35). It was interesting to observe that codon UCG which was not used in MT818177 and MT81819 was found in the coding region of BBrMV CP of MT818187. The average RSCU of this codon in the entire dataset was also close to zero. Whereas, CCG which was present in the CP gene of other isolates was absent in MT818187. Apart from this, non-synonymous codons were used in higher frequency in this isolate compared to the average RSCU.

To understand host in which the virus is most adapted, the codon adaptability index (CAI) was calculated (Table 19). It was surprising to note that CAI values were the same irrespective of the hosts. The Nc plot obtained imparted knowledge on codon usage patterns and demonstrated the role of mutation and natural selection on codon bias (Fig 36). Points below the standard curve, is indicative of the relatively equal contribution of nucleotide composition, mutational pressure and natural selection. If the CUB is influenced only by G+C content of the genome, the points would have been on the stand curve.

The neutrality plot (Fig 37) emphasises the role of natural selection on CUB. The average G+C percentage in the first and second position of the codon (GC12) in all the isolates ranged between 48-50 per cent except for MT818177 and MT818187 in which the values were 44 and 45 per cent respectively. However, percentage of G+C in the third position of the codon (GC3) of the dataset was comparatively broader and ranged from 31- 41 %. There was a significant negative correlation between GC12 and GC3 ($r = 0.385$, $p < 0.01$) suggesting mutational pressure have a prominent role in shaping CUB of BBrMV CP.

4.3.10 Recombination analysis

It was clear from molecular evolutionary analysis that all the BBrMV isolates tested, have diverged from a common ancestor and that the mutations in CP gene in the population is the result of negative or stabilising selection. Thus, it is imperative to understand the role of recombination in the variability of BBrMV CP of Kerala isolates.

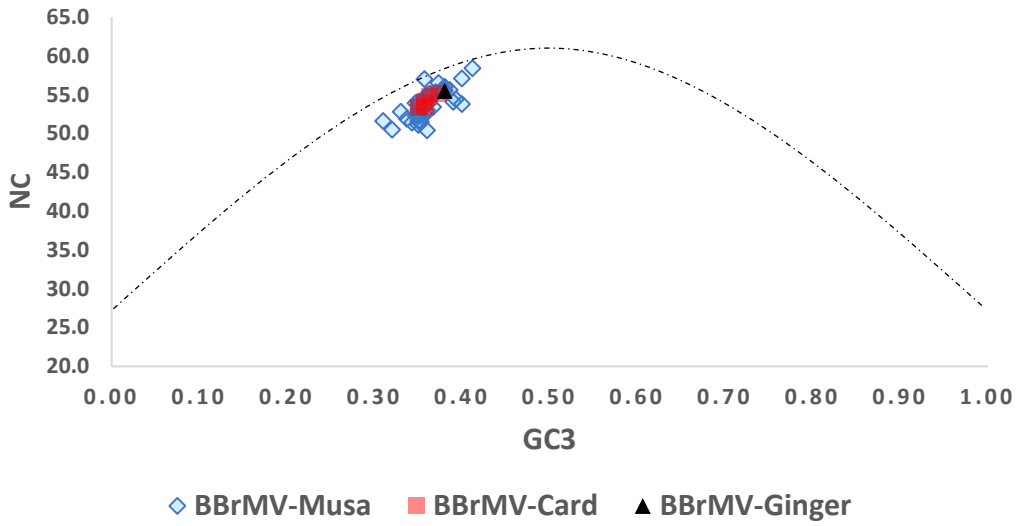


Fig 36. Nc plot of BBrMV CP. The relationship between number of codons (Nc) and GC content at the wobble position (GC3) of coat protein of BBrMV is represented in the graph. The dotted line represent the standard curve of expected GC3 content and Nc value under random codon usage

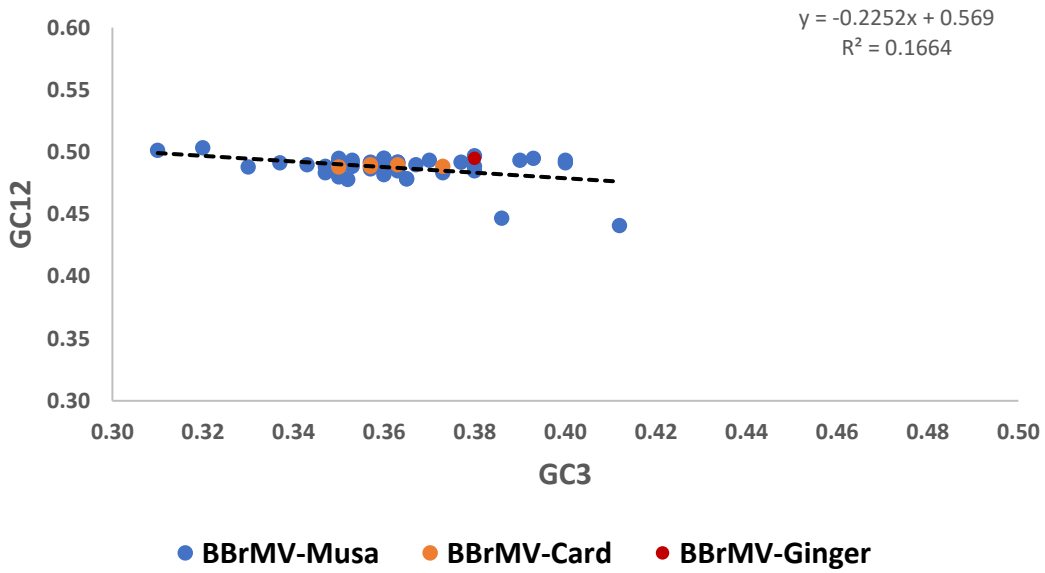


Fig 37. Investigation of magnitude of natural selection and mutational pressure in codon usage bias of BBrMV CP by Neutrality plot. The linear regression of GC12 against GC3 is represented in the graph by the black dotted line

A total of 12 recombinants were detected in the dataset consisting of 56 isolates, out of which, two were isolates from the present study. These two isolates were also the most divergent in the population and showed variations in the codon usage patterns.

Recombination was detected by RDP, GENECONV, MaxChi and 3Seq in MT818177 (Fig 38A). Recombination in MT818187 was reinstated by MaxChi, Chimaera and SiScan (Fig 38B). The major parents of both the recombinants were undetected and the minor parents were HM348782.1 and KF385489.1 respectively, both from Southern part of India (Andhra Pradesh and Tamil Nadu). Recombination breakpoint distribution plots of both suspected recombinants indicated distributions of actual detectable breakpoints in relation to MT818177 (Fig 38C).

Distribution of detectable breakpoint pairs in the analysed gene is represented in the breakpoint pair matrix (Fig 39A). Two breakpoint pairs were recognised between ~400-600 nucleotides in BBrMV coat protein gene.

It was evident that the CP gene had significantly fewer ($p < 0.05$) breakpoint pairs. The recombination region count matrix gave an overview of all the unique recombination events detected in the analysis (lower hemimatrix). Six unique events were detected while analysing 56 sequences of BBrMV coat protein (Fig 39B). Concurrent statistical test of this matrix indicated that few pairs of sites were separated more or less frequently by recombination than can be accounted for by chance (upper hemimatrix).

The McVean's LD matrix detected pairs of sites amidst analysed gene with unusual linkage disequilibrium pattern. Low marginal likelihood values (> -4.0) were indicative of recombinant cold spots (Fig 39C). The encircled region in the matrix indicates a low rate of recombination than expected in recombinantion hot spots within CP gene (corresponding to the region inidctaed as hot spots from breakpoint pair matrix).

The Robinson-Fould's matrix implies the degree to which phylogenetic tress differ from each other which are constructed from different parts of the alignment. Like RF matrix Shimodaira-Hasagawa (SH) matrix is also a useful tool to visualise the effect of recombination on phylogeny. Although, the orange/red regions in matrices are indicative of the fact that the CP of BBrMV mostly have low degrees of phylogenetic compatibility, recombination hot spots cannot be defined (Fig 39D). However, two recombination cold spots, one between ~1-200 nucleotides and another one between ~600-800 nucleotides were distinguished. This is corroborated by the results from McVean's LD matrix.

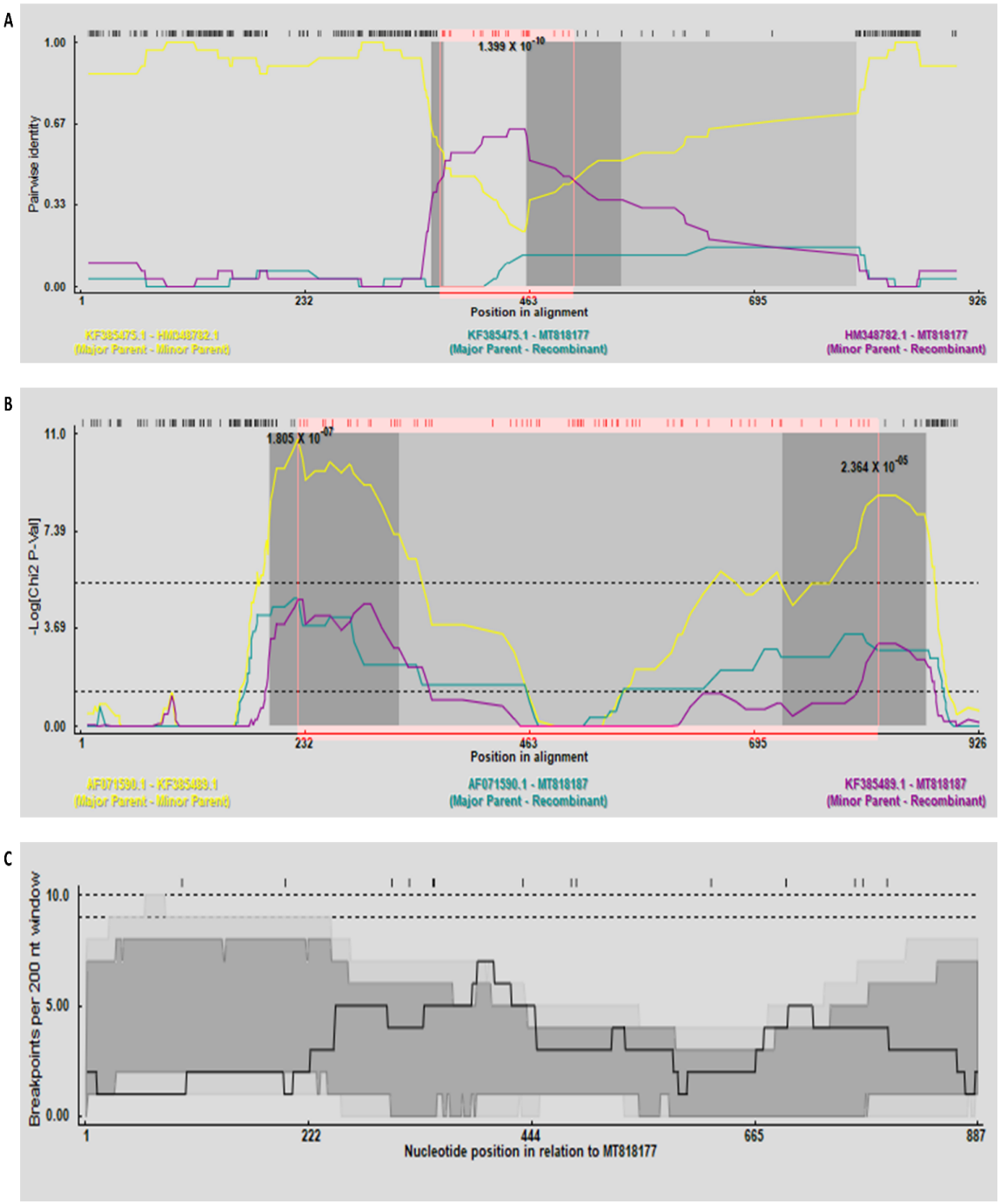


Fig 38. RDP analysis elucidating MT818177 and MT818187 as recombinants. Recombination plots of BBrMV Kerala isolates **A** RDP Plot of MT818177 **B** MaxChi plot of MT818187. Left and right of the pink line indicates recombinant break points. C Recombination breakpoint distribution plot in relation with type sequence, MT818177. The upper and lower discontinuous lines represent 99 % and 95 % confidence intervals respectively. Small vertical lines at the top of the graph indicates detectable breakpoint positions. The light and dark grey areas indicated 99 % and 95 % confidence intervals for recombinant cold/ hot spot test

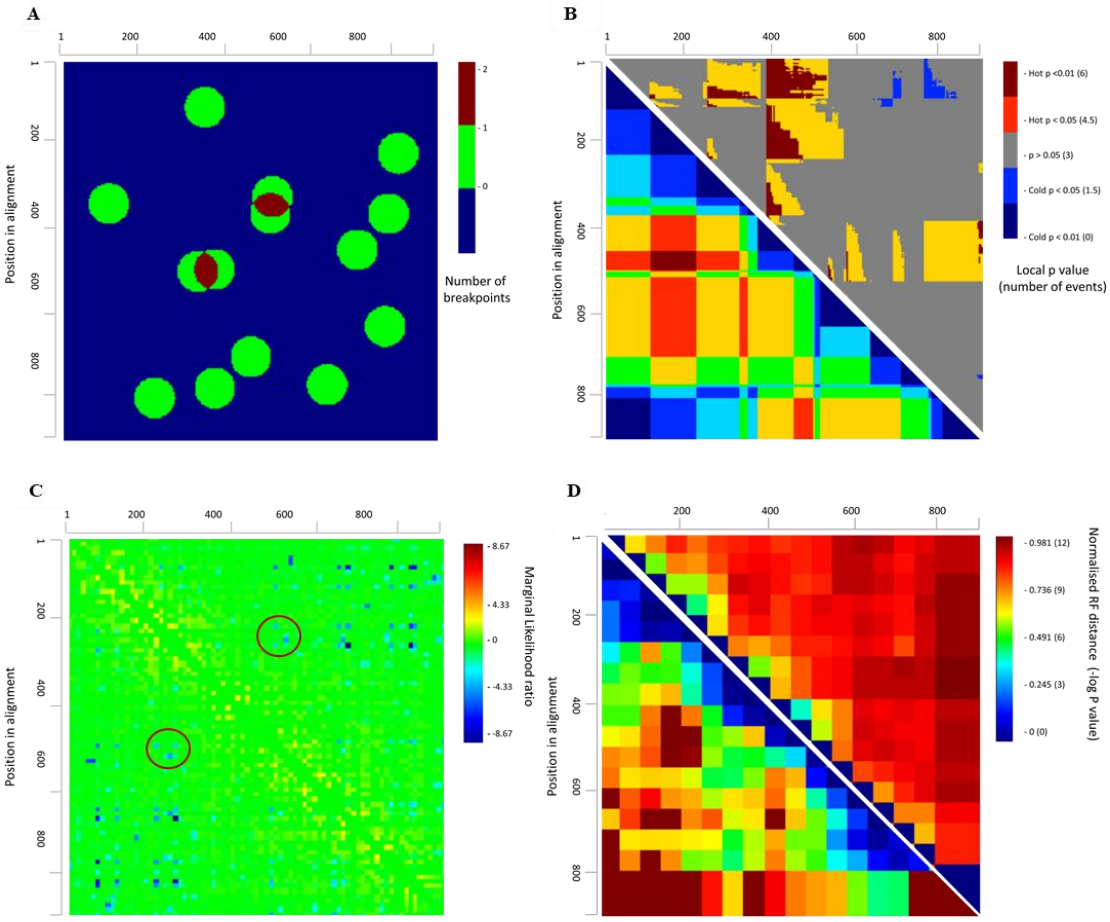


Fig 39. Recombination matrices based on BBrMV CP A Recombination breakpoint matrix **B** Recombination region count matrices indicate the presence of a recombination breakpoint pair hotspots. Statistically significant cold and host spots are depicted in the upper hemimatrix **C** McVean's Linkage disequilibirum (LD) matrix. The red circles indicate recombination rate lower than expected in recombinantion hot spots within coat protein gene **D** Robinson-Fould's (RF) (upper) and Shimodaira-Hasagawa (SH) (lower) compatibility matrices

4.3.11 Validation of designed primers for virus detection

Evidently, from the diversity analysis, BBrMV CP gene is evolving by means of mutation and recombination. The selection pressure is acting upon the codon usage bias and many interesting patterns were observed in the newly evolved isolates collected during the present study. These inferences were indicative of the importance of designing primers for nucleic-acid based detection of BBrMV considering the molecular diversity present in the population. Since it was observed that the N and C terminal are variable compared to the core region of the CP gene, forward primer was designed from comparatively conserved

NIb region and reverse primer targeted the 3' UTR (BBrMV NIb-CP FP and RP). The annealing temperature of the designed primer pair specific amplifying partial NIb region and complete CP region of BBrMV CP was standardised by gradient PCR. Amplicon of 1306 bp as expected was obtained in temperatures 51.6 ºC, 52.3 ºC, 54.9 ºC, 58.3 ºC, 64.1 ºC, 64.7 ºC. Among these, 54.9 ºC was selected as annealing temperature for the primer set. cDNA from infected samples collected from various parts of Kerala was used as template for validation of the primers (Plate 37). Few samples that did not amplify using the reported primers also amplified using the designed primers (Plate 38).

4.3.12 Molecular cloning

Cloning of BBrMV CP gene was carried out to various vectors as described in the methods section. The CP gene amplified using Taq polymerase was purified using ThermoFisher purification kit by following manufacturer's protocol (Plate 39A). Transformants were selected on LB supplemented with ampicillin. The clone was selected by blue white screening. Randomly selected white colonies were subjected to colony PCR (Plate 39B). Out of 10 colonies selected 6 were positive clones.

The positive clones that gave an amplicon of 1058 bp were selected and sub cultured. Plasmid from the clone was isolated by alkaline lysis method. Restriction digestion with EcoR1 released the insert from the plasmid as recognition sites of the enzymes are present at both the upstream and downstream of the insert (Plate 39C).

The pGEM-T clone was used as the template for amplifying BBrMV CP gene using designed primers (BBrMV CP sen and anti) to incorporate Nhe1 and BamH1 sites at the 5' and 3' end of the gene respectively. The annealing temperature of the primers were standardised to be 55 ºC (Plate 40A). Approximately 1080 bp gene was amplified using high fidelity Pfu polymerase in order to avoid any errors during multiplication (Plate 40B). The gene digested with Nhe1 and BamH1 was then ligated to double digested pRSET-C expression vector and transformed to DH5α (Plate 40C). The digested plasmid showed only the one linear band corresponding to 3 kb in contrast to three bands shown by an undigested plasmid.

Since blue white screening was absent in pRSET-C vector, plasmid was isolated from randomly selected white colonies (Plate 41A) and the plasmids that showed a difference in mobility due to increase in size was subjected to PCR (Plate 41B) and restriction digestion

(Plate 41C). Plasmid was used as template for PCR using designed primers to amplify BBrMV CP gene. Restriction digestion using BamH1 of the plasmid that gave amplification of the gene was also observed. Only one plasmid became linear after digestion with BamH1. The confirmation of this clone was also carried out by sequencing with T7 forward primer. On subjecting to NCBI-BLAST, the sequence showed maximum homology to BBrMV CP submitted in NCBI-GenBank database. The confirmed pRSET-C vector with BBrMV CP insert, was transformed to DH5α to amplify the clone. Plasmid was later transformed to two *E. coli* strains *viz.,* BL21 pLysS and Rosetta pLysS for overexpression of gene.

The BBrMV CP gene was also cloned to Sma1 site of pGEX-4T-2. The clone confirmation was done by plasmid isolation and PCR (Plate 42A and B). Since blunt cloning was performed, orientation of the gene was confirmed by restriction digestion with BamH1 (Plate 42C). As BamH1 site was incorporated at the 3' end of the gene by primer and also present in pGEX-4T-2 vector at the upstream to the cloning site of BBrMV CP, in positive clone with correct orientation, the insert was released on digestion with BamH1 enzyme. The positive clone was amplified in DH5α and subsequently transformed to BL21 pLysS and Rosetta pLysS for overexpression of the protein.

4.3.13 Overexpression of rBBrMV CP

The rBBrMV CP was overexpressed in two strains of *E. coli viz.,* BL21 pLysS and Rosetta pLysS. These strains harbouring pRSET-C/BBrMV CP was grown in LB broth until it reached the log phase and induced with 1 mM IPTG at 16 ºC. Subsequently the cell lysate of uninduced and induced culture was loaded on to 12 per cent SDS-PAGE for separation and to understand the protein profile. In the gel, a doublet at 35 kDa corresponding to the BBrMV CP was obtained in the induced fraction which was less in the uninduced fraction (Plate 43A and B). In both BL21 and Rosetta cell lines, the overexpression of BBrMV CP was similar. Nevertheless, since rare codons were present in the protein, Rosetta was used for further standardisations.

Due to cloning of BBrMV CP between Nhe1 and BamH1 sites of the vector, three extra amino acids after hexa histidine tag were added to the protein. Confirmation of the protein was done by Western blotting as explained in the methods section. From the blot, only the induced fraction gave a doublet band at 35 kDa when *Pepper vein banding virus* (PVBV),

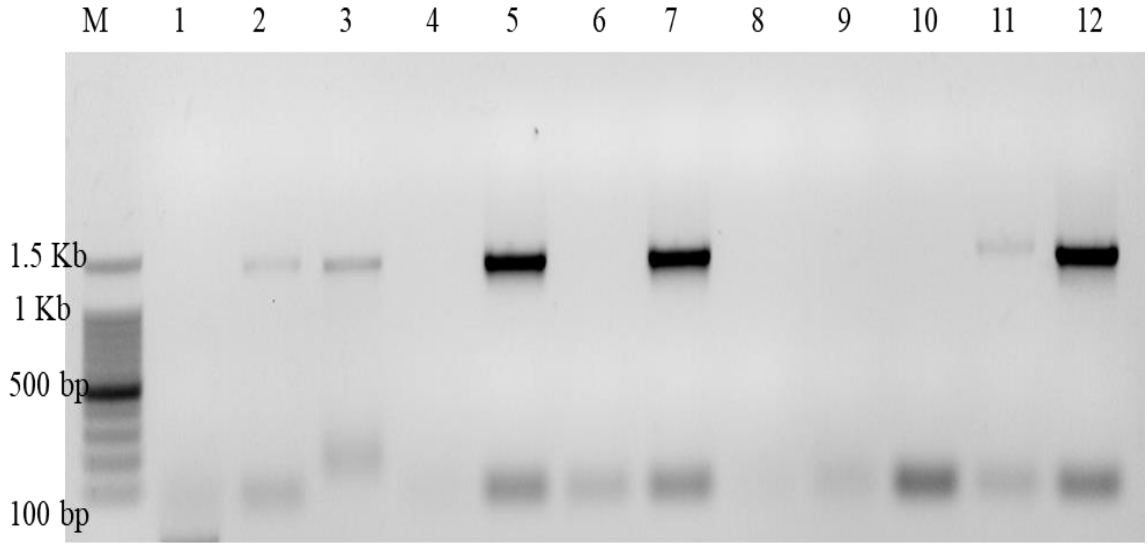


Plate 37 Standardisation of annealing temperature of BBrMV specific primer. M: 1 kb DNA ladder; Lanes 1-12: Amplification at annealing temperatures 51.4 ºC, 51.6 ºC, 52.3 ºC, 53.4 ºC, 54.9 ºC, 56.6 ºC, 58.3 ºC, 60.1 ºC, 61.7 ºC, 63.3 ºC, 64.1 ºC, 64.7 ºC

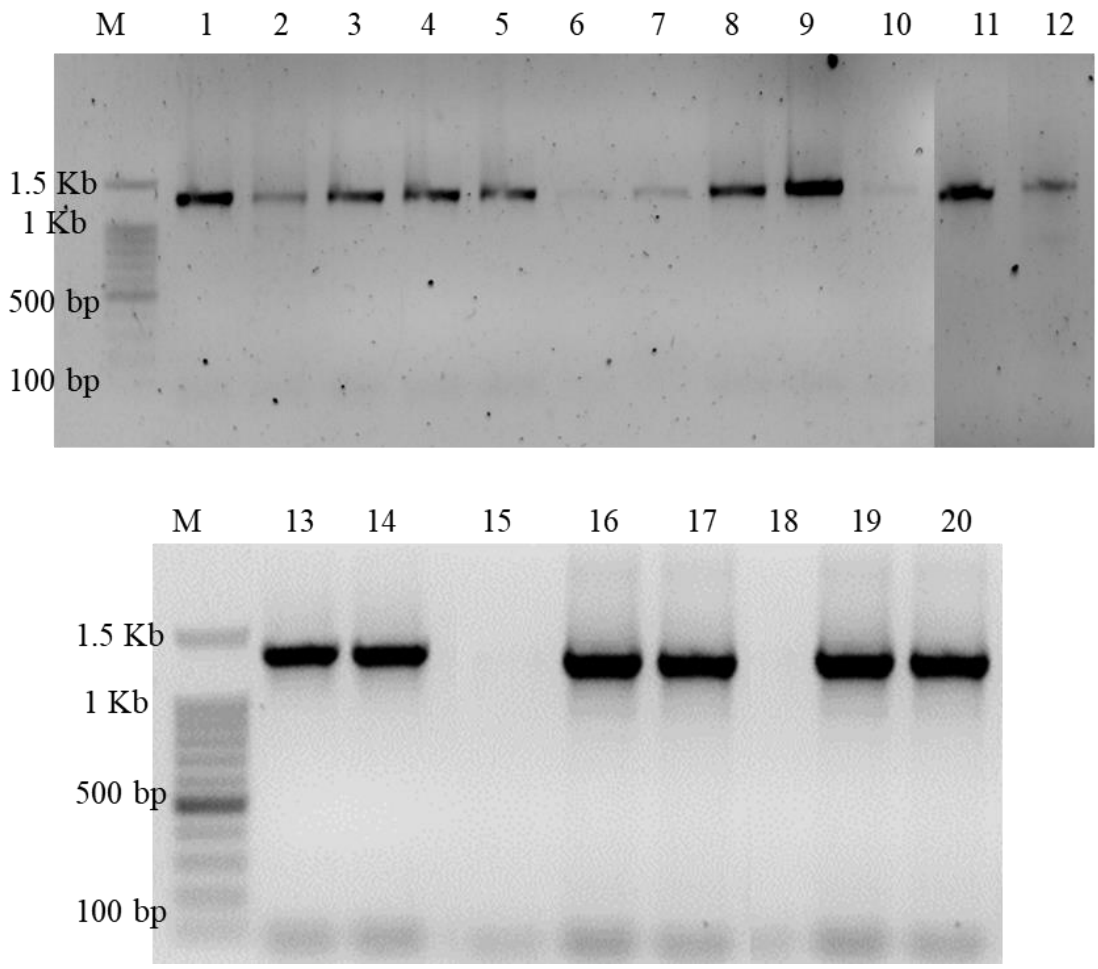


Plate 38 Validation of primer for nucleic-acid based detection of BBrMV in banana

M: 1 kb DNA ladder; Lane 1: TCRBr1 (Kunnan), Lane 2: KKDBr15 (Mysore poovan), Lane 3: KKDBr19 (Nendran), Lane 4: KASBr1 (Chengalikodan), Lane 5: TCRBr1 (Kunnan), Lane 6: WYDBrR1 (Nendran)*, Lane 7: TCRBr2 (Nendran), Lane 8: EKMBr1 (Attunendran), Lane 9: TCRBr3 (Nallabontha), Lane 10: KASBr1 (Nendran), Lane 11: KANBr1 (Nendran), Lane 12: TCRBr5 (Njalipoovan), Lane 13: BRSBr12 (Kunnan), Lane 14: PKDBr2 (Nendran)*, Lane 15: BRSBr14 (Phia), Lane 16: KKDBr20 (Mysore poovan), Lane 17: KOLBr1 (Nendran)*, Lane 18: BRSBr17 (Chengkadali), Lane 19: BRSBr18 (Njalipoovan); Lane 20: KKDBr21 (Njalipoovan)*

* Samples in which CP gene amplified in PCR using the designed primers and not in the reported primers

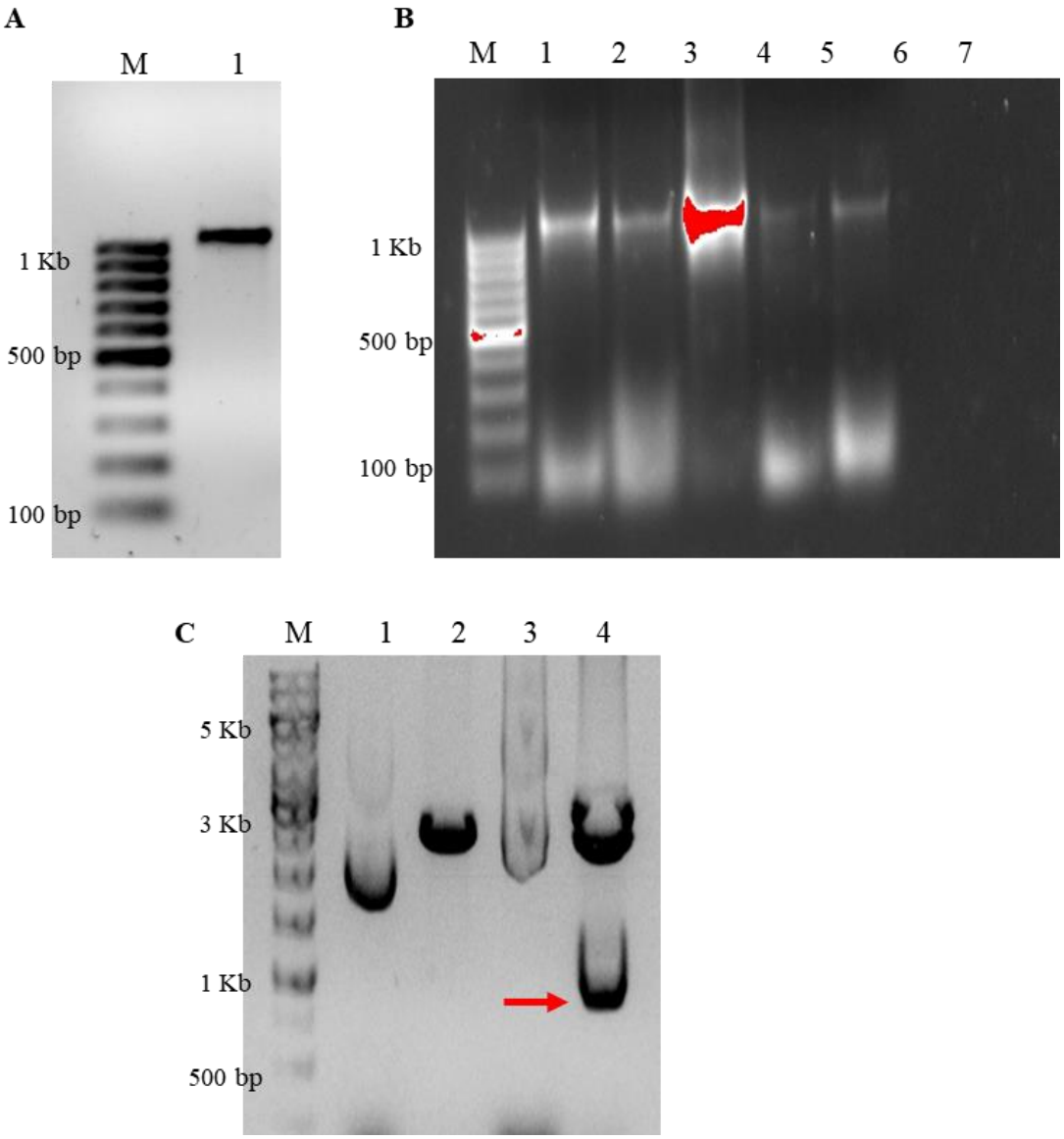


Plate 39 Cloning of BBrMV CP to pGEM-T easy vector. A Purification of PCR product. M M: 100 bp DNA ladder; 1: Purified BBrMV CP B Colony PCR of randomly selected white colonies. M: 100 bp DNA ladder; 1-5: positive clones; 6-7: negative C Restriction digestion of positive clone after screening by -colony PCR. M: 1 kb DNA ladder; 1: undigested pGEM-T empty vector; 2: digested pGEM-T vector; 3: undigested clone 4: digested clone. The 1 kb insert released is marked with red arrow

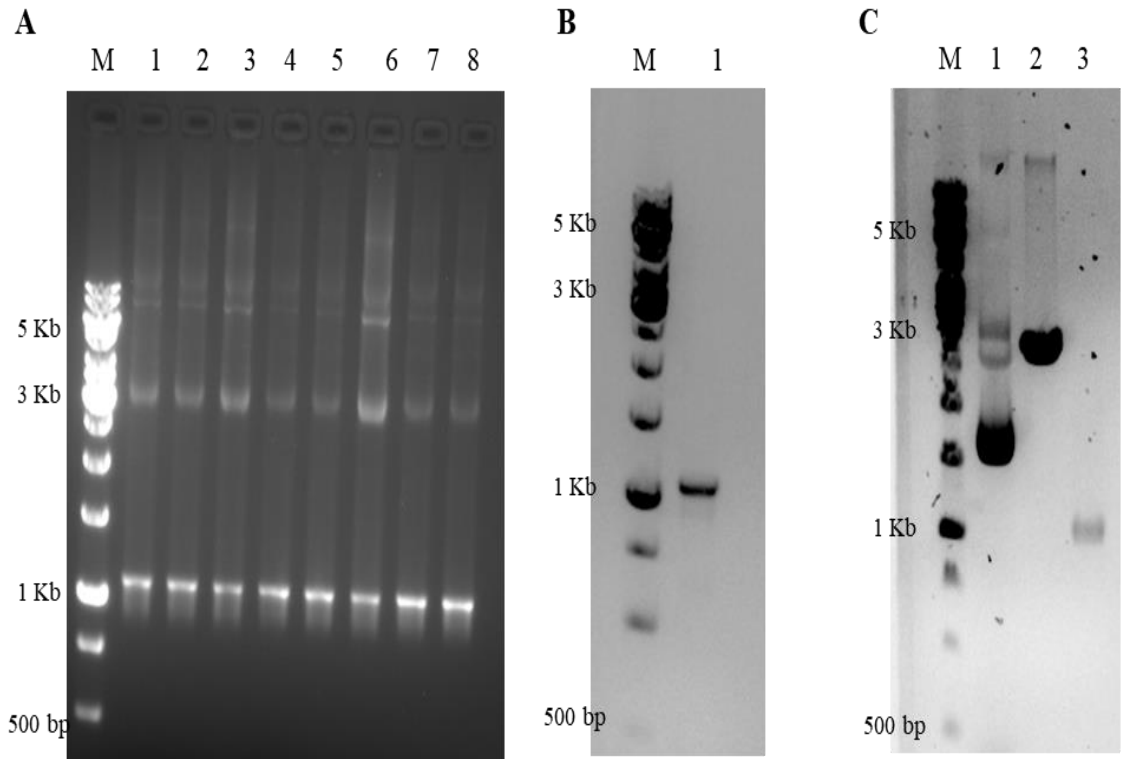


Plate 40 Cloning of BBrMV CP gene to pRSET-C. A Confirmation of annealing temperature of designed primer. M: 1 kb DNA ladder; Lanes 1-8: PCR product amplified at various annealing temperatures ***viz.*****, 65 °C, 64.3 °C, 63.1 °C, 61.3 °C, 59 °C, 57.3 °C, 56 °C, 55 °C. B** Amplification of 1080 bp BBrMV CP at 55 ^{o}C. **C** Digestion of pRSET-C and PCR product using Nhe1 and BamH1. M: Marker; Lane 1: pRSET-C empty vector; Lane 2: digested pRSET-C; Lane 3: digested PCR product

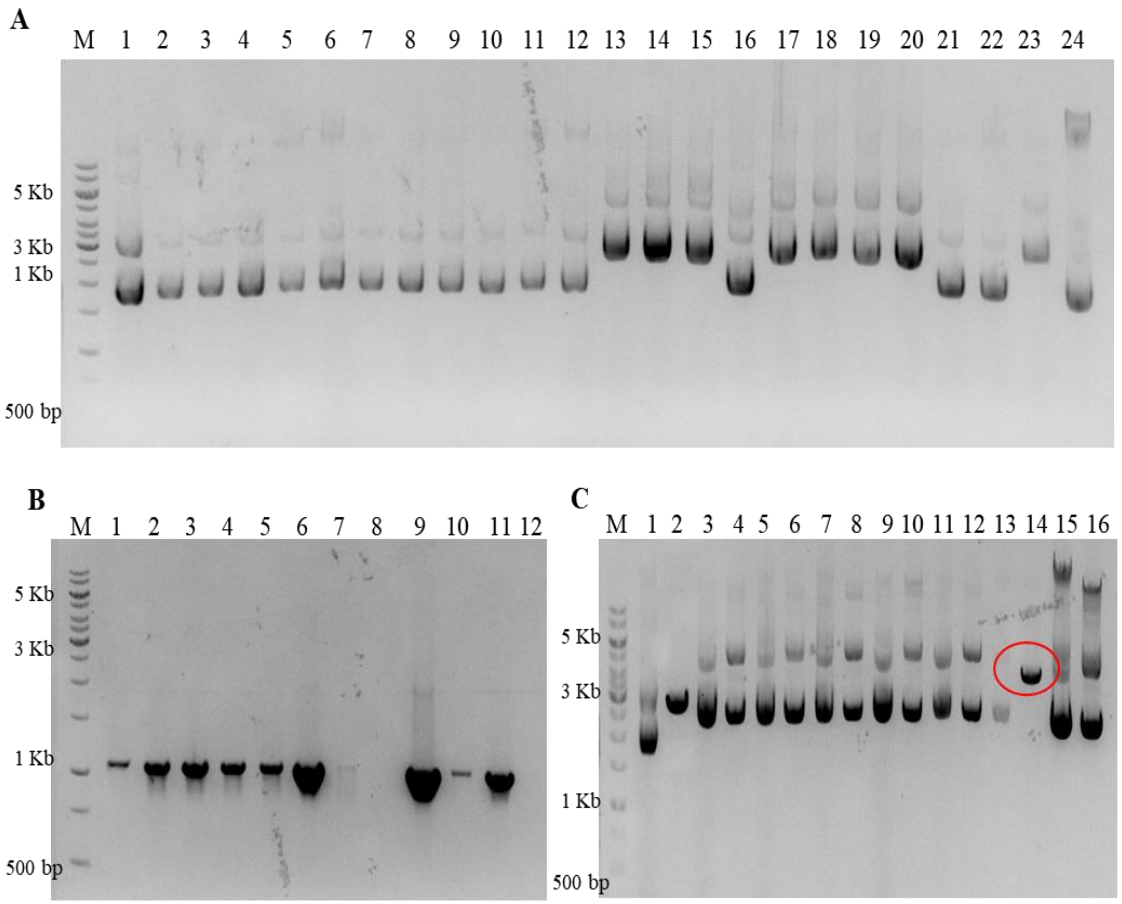


Plate 41 **BBrMV CP clone confirmation. A** isolation of plasmid from randomly selected white colonies. M: 1 kb DNA molecular weight marker; Lane 1: pRSET-C empty vector; Lanes 2-12, 16, 21, 22, 24: plasmid with no insert; Lanes 13-15, 17-20, 23: clones with change is mobility supposedly positive. **B** Confirmation of clone by PCR. M: 1 kb DNA molecular weight marker; Lanes 1-6, 9-11: positive clones that gave amplification of the gene using designed primers; Lanes 7-8, 12: negative clones with no amplification **C** restriction digestion of the clones that gave positive amplification in PCR. M: 1 kb DNA molecular weight marker; Lane 1: undigested pRSET-C empty vector; Lanes 2-13, 15-16: negative clones; Lane 14: Positive clone. The single band in the positive clone is encircled in the gel image

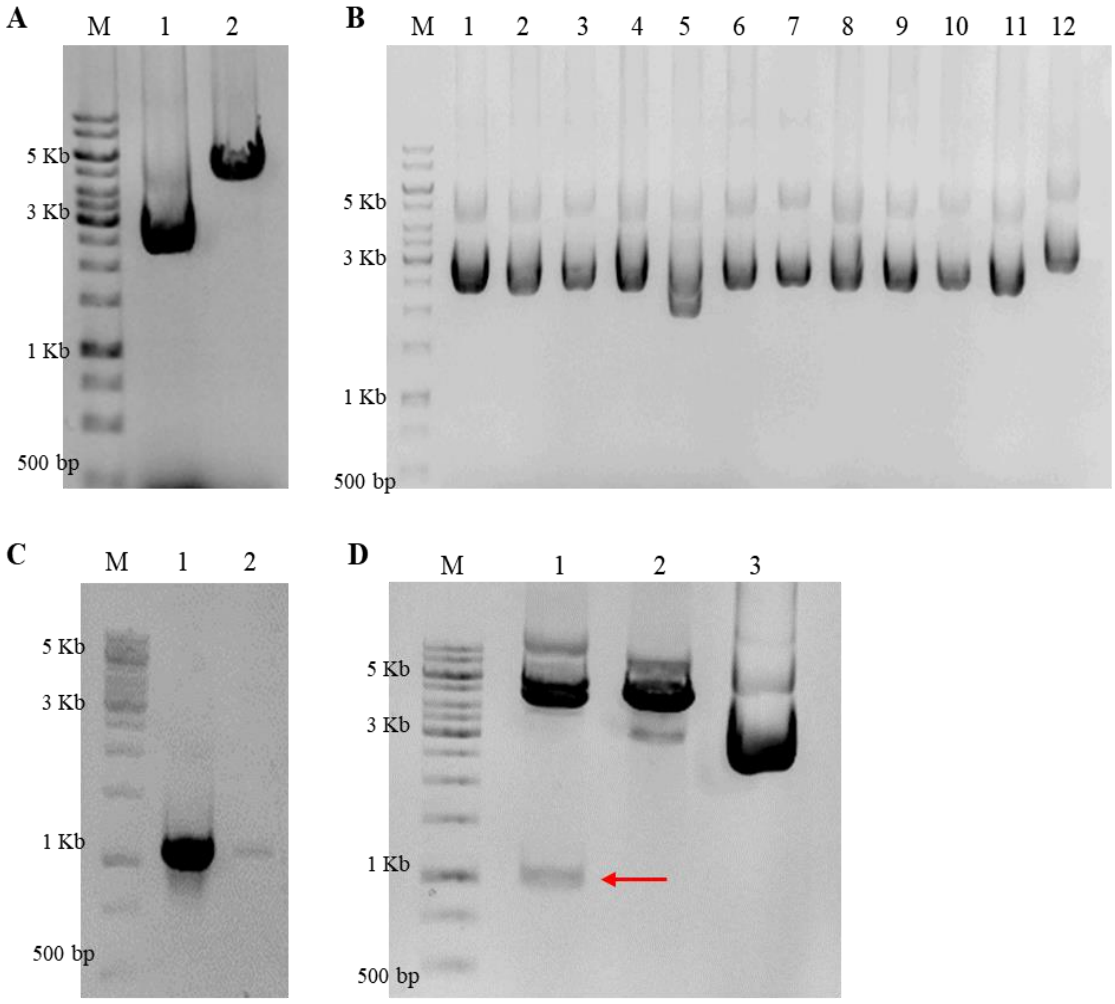


Plate 42 Cloning of BBrMV CP to pGEX-4T-2 and confirmation. A Sma1 digestion of pGEX-4T-2. M: 1 kb DNA ladder; Lane 1: undigested pGEX-4T-2; Lane 2: digested pGEX-4T-2. B Plasmid isolated form random colonies after transformation of vector ligated with BBrMV CP gene to DH5α. M: 1 kb DNA ladder; Lane 1: empty pGEX-4T-2. Lanes 2-11: negative clone; Lane 12: Plasmid with change in mobility due to increase in size. C Clone confirmation by PCR. M: 1 kb DNA ladder; Lane 1: positive clone; Lane 2: Negative control. D Clone confirmation by BamH1 digestion. M: 1 kb DNA ladder; Lane 1: positive clone. Red arrow indicated 1 kb insert released on digestion; Lane 2: negative; Lane 3: undigested pGEX-4T-2

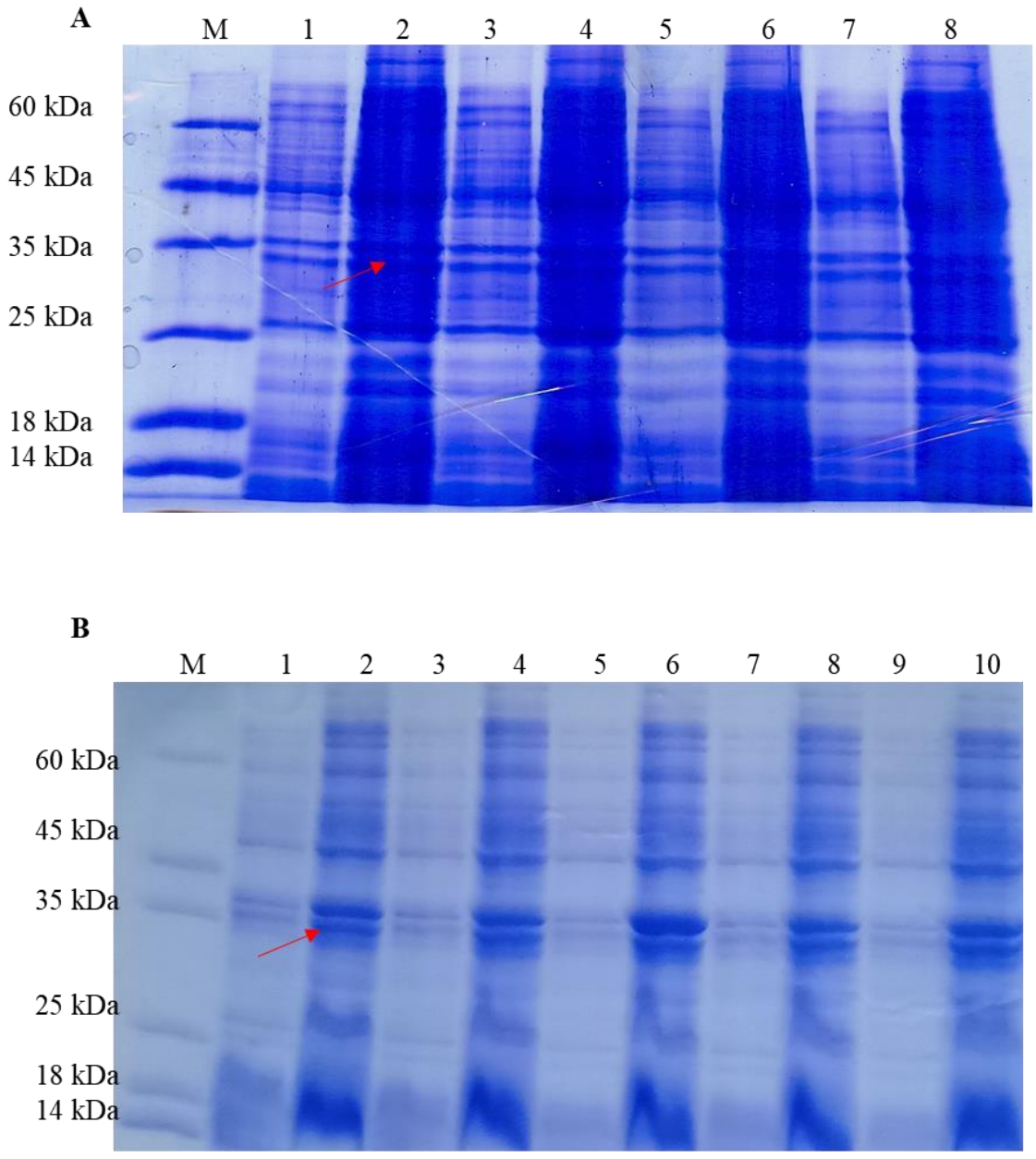


Plate 43. Overexpression of rBBrMV CP in A BL21pLysS M: Protein molecular weight marker; Lanes 1, 3, 5 and 7: uninduced cell lysate; Lane 2, 4, 6 and 8: induced cell lysate. **B** Overexpression of rBBrMV CP in Rosetta pLysS. M: Protein molecular weight marker; Lanes 1, 3, 5, 7 and 9: uninduced cell lysate; Lanes 2, 4, 6, 8 and 10: induced cell lysate. The red arrow indicates the doublet band corresponding to rBBrMV CP

also a potyvirus, CP specific antiserum was used (Plate 44A). Similarly, when anti-his monoclonal antibody was used as primary antibody for detection of rCP with N terminal his tag, only one band at 35 kDa was observed in the blot (Plate 44B). However, protein overexpressed in both the hosts were insoluble.

Further, overexpression of rBBrMV CP was standardised with the intention of enhancing solubility of the protein. The parameters influencing overexpression like, the media for culturing the cells harbouring the recombinant vector, the concentration of inducer, time and temperature of induction were standardised. Maximum expression was obtained when induced at 16 ºC for 12 h by 0.3 mM IPTG when cultured in YT broth (Plate 45).

Nevertheless, cell mass was observed maximum in TB. Surprisingly, more than the 75 per cent of the rCP in all the conditions was insoluble (Plate 45). Presence and absence of protease inhibitor cocktail did not affect the overexpression or solubility of rBBrMV CP.

As the rCP was insoluble in pRSET-C, pGEX/BBrMV CP clone was transformed to BL21 and Rosetta pLysS to asses overexpression and solubility. A 60 kDa fusion protein (34 kDa CP along the 25 kDa GST tag) was expected after induction. From the protein profile of the cell lysate before and after induction, evidently, a 60 kDa band was observed in the induced fraction which was absent in the uninduced fraction (Plate 46). Unlike the previous clone, no leaky expression was observed in this one. Western blotting with anti-PVBV CP polyclonal antiserum as primary antibody detected the 60 kDa fusion protein (Plate 47). A doublet band at 60 kDa as observed earlier was detected by the antiserum. In Western blot, doublet band was observed in the induced fraction as well as in the supernatant. From the blot it was affirmative that there was no leaky expression as well as the protein was completely soluble.

4.3.14 Purification of rBBrMV CP

4.3.14.1 Purification of pRSET/rBBrMV CP

For purification of rBBrMV CP, pRSET/BBrMV CP clone was transformed in to BL21 and Rosetta pLysS as both the hosts exhibited similar overexpression of the protein. The culture was grown in YT media and induced at 0.3 mM for 12 h at 16 ºC. The induced culture harvested by centrifugation was resuspended in lysis buffer (Tris buffer, pH 8.5) and sonicated. On sonication, it was observed that more than 75 per cent of the protein was

insoluble (Plate 48). On performing Ni^{2+}-NTA column chromatography, a high molecular weight *E. coli* protein as well as a low molecular weight membrane protein was eluted with the 35 kDa protein of interest. Single band was observed at 35 kDa. To confirm that this protein is rBBrMV CP, elution was blotted on to NCM membrane and detected using PVBV CP specific primary polyclonal antiserum (1:5000 v/v). Even though lysis was done in the presence of protease inhibitor cocktail, a doublet band at expected size confirmed presence of the protein (Plate 49).

However, to increase the solubility and yield of the protein, various parameters were standardised. The harvested induced cells were treated with lysozyme for 2 h prior to sonication. The time for sonication and amplitude was reduced to decrease the formation of inclusion bodies. When the cell lysates were observed on the gel, it was evident that rBBrMV CP was present in the insoluble fraction and many more non-specific proteins have bound to the Ni^{2+}-NTA beads and eluted with the protein of interest which was very meagre (Plate 50).

Lysis buffer for purification was standardised to obtain soluble protein. Sonication of induced cells was carried out in the presence of Tris buffer at pH 8.0, HEPES buffer at pH 7.5 and CAPS buffer at pH 9.2. Tris buffer at pH 8.0 and HEPES buffer at pH 7.5 generated very less amount of protein (Plate 51). Although, CAPS buffer performed the best out of all the lysis buffers, many *E. coli* proteins were present in the elution. The amount of protein eluted was also higher in case of CAPS buffer. Hence, for further standardisations, CAPS buffer was used.

4.3.14.2 Purification under denaturing conditions

Since pRSET/rBBrMV CP was insoluble, attempt was carried out to purify the protein under denaturing conditions. For the same, lysis buffer, wash buffer and elution buffer were supplemented with either urea (8 M) or guanidine hydrochloride (6 M). The Ni^{2+}-NTA purification was carried out in room temperature. Time taken for purification was much longer when compared to the previous methods.

The gel profile of urea denaturation depicted that the protein was denatured on lysis and that it has become soluble (Plate 52). The concentration of protein of interest in the elutions were higher than in the previous experiments. However, the affinity of the protein has

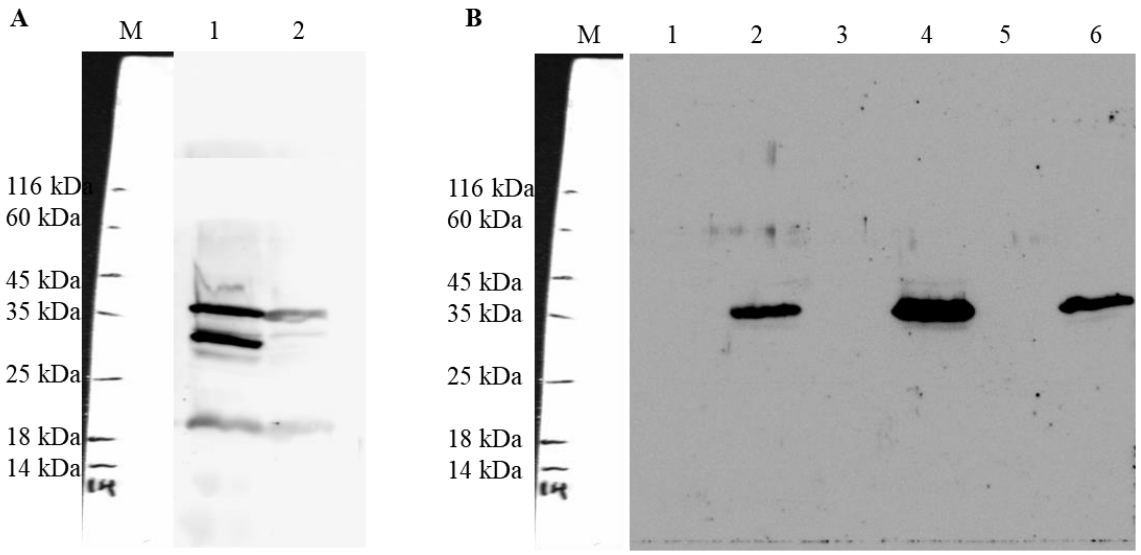


Plate 44 **Western blot analysis of overexpressed rBBrMV CP** detected using **A** PVBV polyclonal antiserum, M: Protein molecular weight marker; Lane 1: induced cell lysate; Lane 2: uninduced and **B** anti-his monoclonal antibody as primary antibody, M: Protein molecular weight marker; Lanes 1, 3 and 5: uninduced cell lysate; Lanes 2, 4 and 6: induced cell lysate.

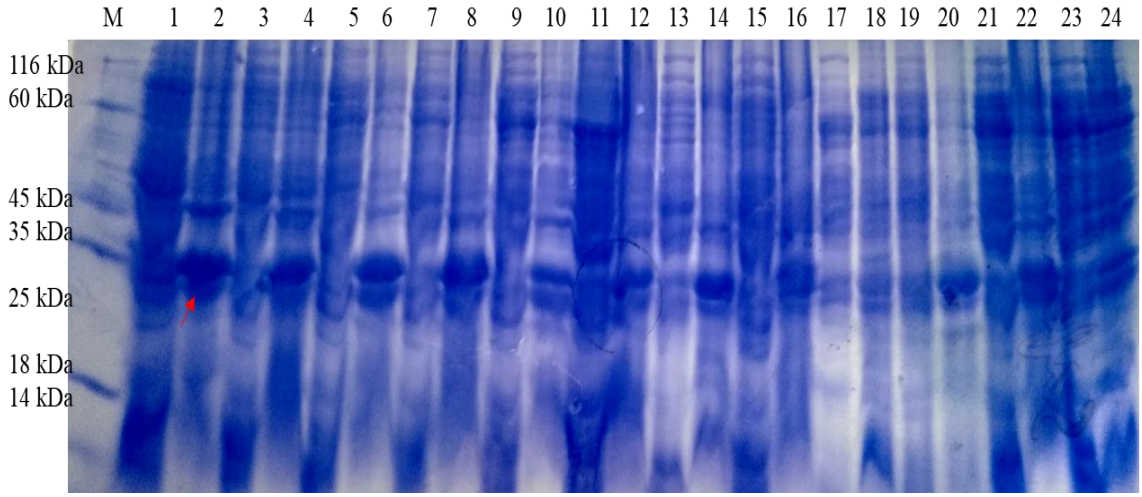


Plate 45 Standardisation of overexpression of rBBrMV CP M: Protein molecular weight marker; Lanes 1 and 2: supernatant (soluble fraction) and pellet (insoluble fraction) of Rosetta pLysS harbouring pRSET/BBrMV CP cultured in LB broth induced with 0.3 mM IPTG at 16 ºC for 12 h; Lanes 3 and 4: supernatant and pellet of cell lysate induced with 0.5 mM IPTG at 16 ºC for 12 h cultured in LB broth; Lanes 5 and 6: supernatant and pellet of cell lysate induced with 0.3 mM IPTG at 30 ºC for 6 h cultured in LB broth; Lanes 7 and 8: supernatant and pellet of cell lysate induced with 0.5 mM IPTG at 30 ºC for 6 h cultured in LB broth; Lanes 9 and 10: supernatant and pellet of cell lysate induced with 0.3 mM IPTG at 16 ºC for 12 h cultured in YT broth; Lanes 11 and 12: supernatant and pellet of cell lysate induced with 0.5 mM IPTG at 16 ºC for 6 h cultured in YT broth; Lanes 13 and 14: supernatant and pellet of cell lysate induced with 0.3 mM IPTG at 30 ºC for 6 h cultured in YT broth; Lanes 15 and 16: supernatant and pellet of cell lysate induced with 0.5 mM IPTG at 30 ºC for 6 h cultured in YT broth; Lanes 17 and 18: supernatant and pellet of cell lysate induced with 0.3 mM IPTG at 16 ºC for 12 h cultured in TB broth; Lanes 19 and 20: supernatant and pellet of cell lysate induced with 0.5 mM IPTG at 16 ºC for 12 h cultured in TB broth; Lanes 21 and 22: supernatant and pellet of cell lysate induced with 0.3 mM IPTG at 16 ºC for 12 h cultured in TB broth; Lanes 23 and 24: supernatant and pellet of cell lysate induced with 0.5 mM IPTG at 30 ºC for 6 h. The doublet band corresponding to rBBrMV CP in the insoluble fraction is indicated by red arrow.

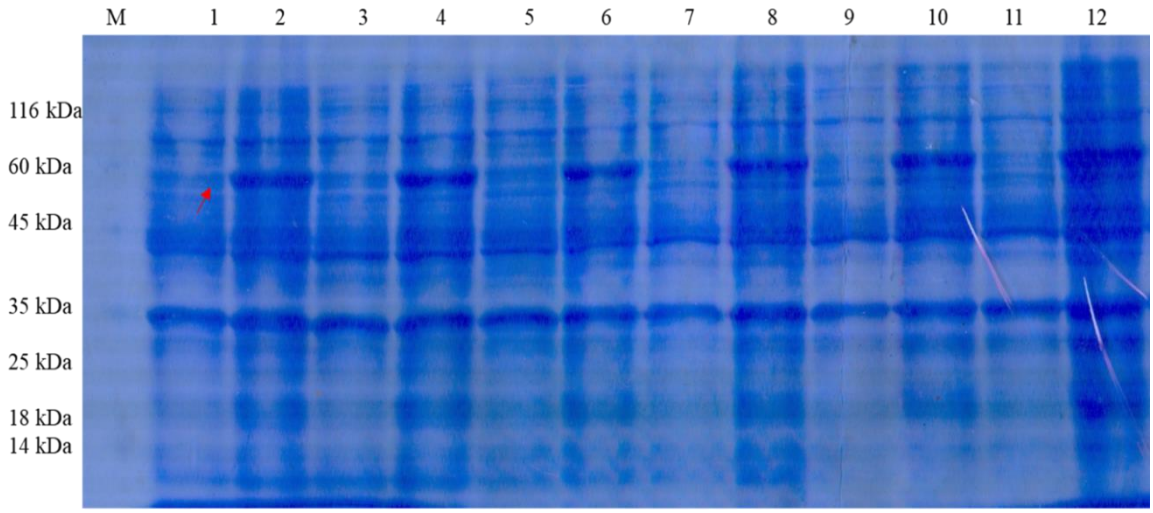


Plate 46 Standardisation of expression host for overexpression of rBBrMV CP M: Protein molecular weight marker, BBrMV CP overexpressed in Rosetta pLysS harbouring pGEX/BBrMVCP, The 60 kDa protein suspected to be rBBTV CP is marked using red arrow, Lanes 2, 4, 6: culture induced by 0.3, 0.5, 1 mM, Lanes 1, 3, 5 : uninduced, Lanes 8, 10, 12: overexpression of rBBrMV CP in BL21 pLysS induction by 0.3, 0.5, 1 mM, Lanes 7, 11, 13: uninduced, Lanes 1-4: Cultured in LB, Lanes 5-8: cultured in YT, Lanes 9-12: cultured in TB. All cultures induced at 16 ^{o}c for 14 h

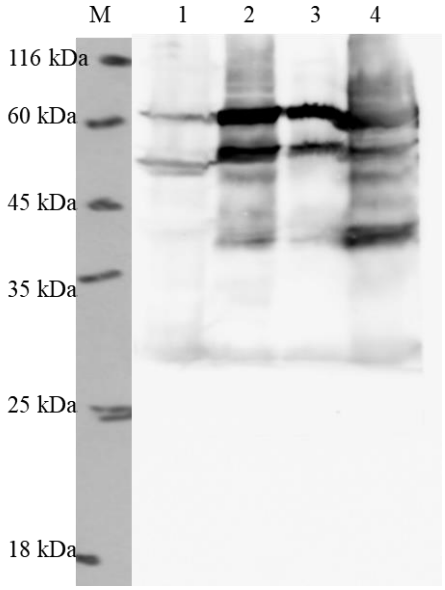


Plate 47 Confirmation of fusion protein by Western blot with polyclonal PVBV antiserum. M: Protein molecular weight ladder marked after Ponceau's staining of the blot. Lane 1: uninduced; Lane 2: induced, Lane 3: supernatant; Lane 4: induced pellet

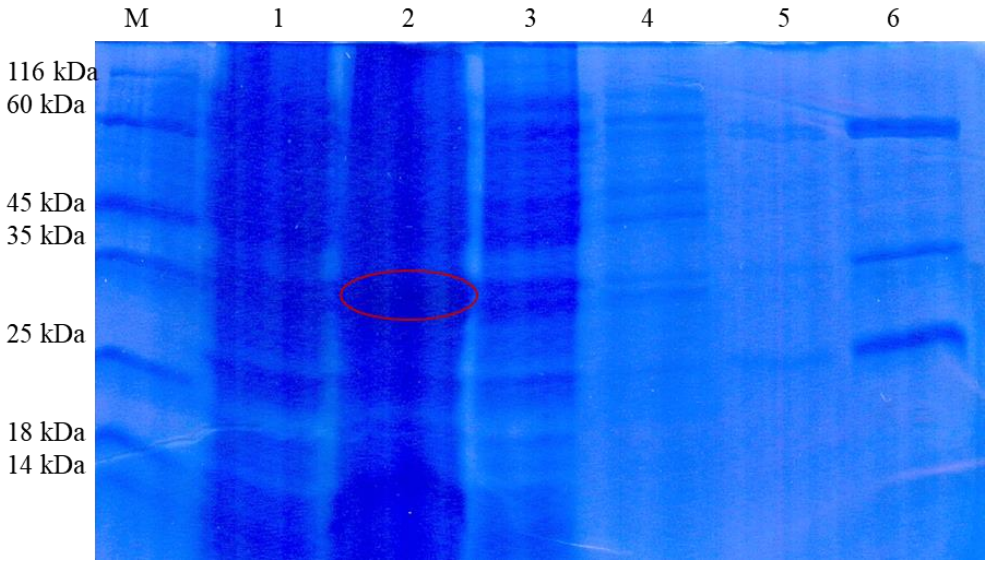


Plate 48 Purification of rBBrMV CP by Ni^{2+}-NTA affinity column chromatography M: Protein molecular weight marker, Lane 1: supernatant of induced cell culture after sonication; Lane 2: pellet (insoluble fraction); Lanes 3-4: wash; Lanes 5-6: elutions

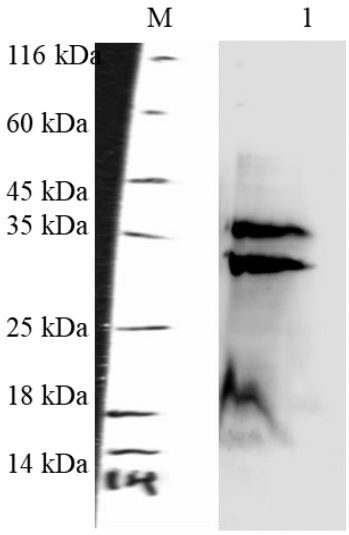


Plate 49 Western blot analysis of purified rBBrMV CP CP. M: Protein molecular weight marker; Lane 1: purified rBBrMV CP

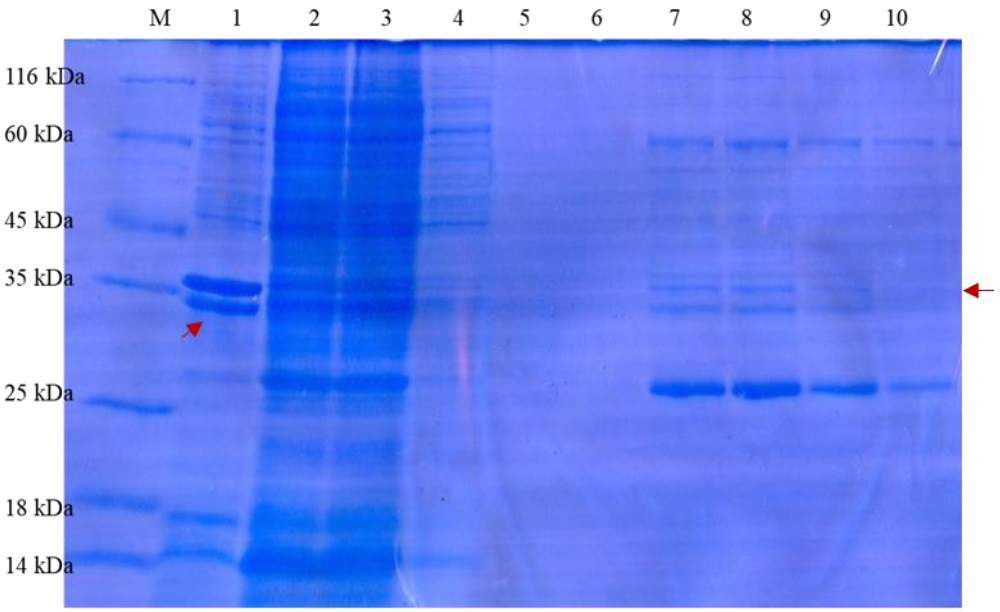


Plate 50 **Increasing solubility of protein by lysozyme treatment,** M: Protein molecular weight marker; Lane 1: pellet, Lane 2: supernatant, Lanes 3-4: wash; Lanes 5-10: elutions. Red arrow indicates the protein of interest in the pellet and elutions

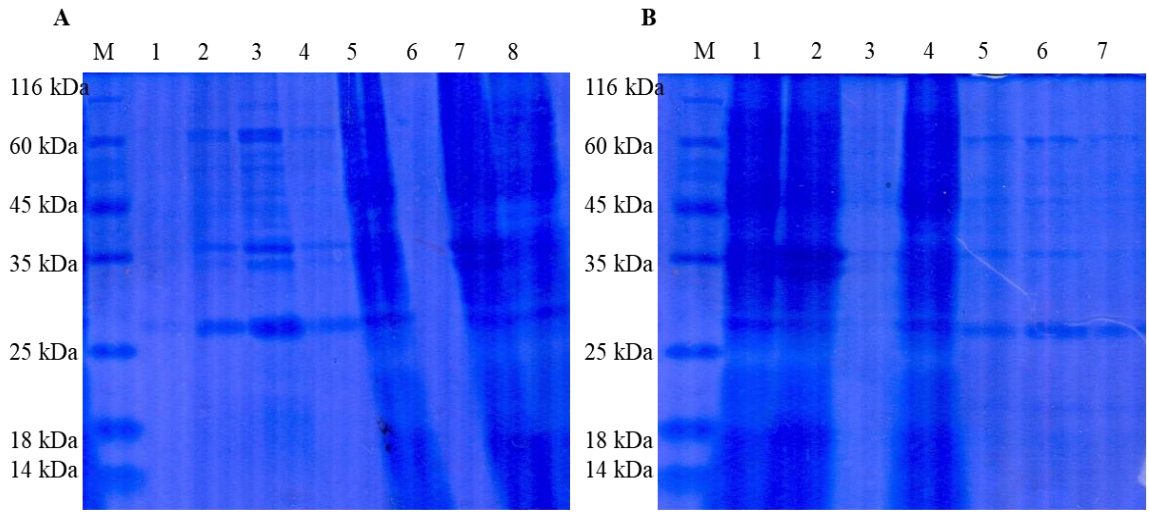


Plate 51 **Standardisation of lysis buffer for purification A** Purification of rBBrMV CP after sonication in CAPS buffer, pH 9.2; M: Protein molecular weight marker; Lanes 1-4: elutions; Lane 5: wash; Lane 6: flow through; Lane 7: pellet; Lane 8: supernatant B Purification of rBBrMV CP after sonication in HEPES buffer, pH 7.5; M: Protein molecular weight marker; Lane 1: supernatant; Lane 2: pellet; Lane 3: flow through; Lane 4: wash; Lanes 5-7: elutions

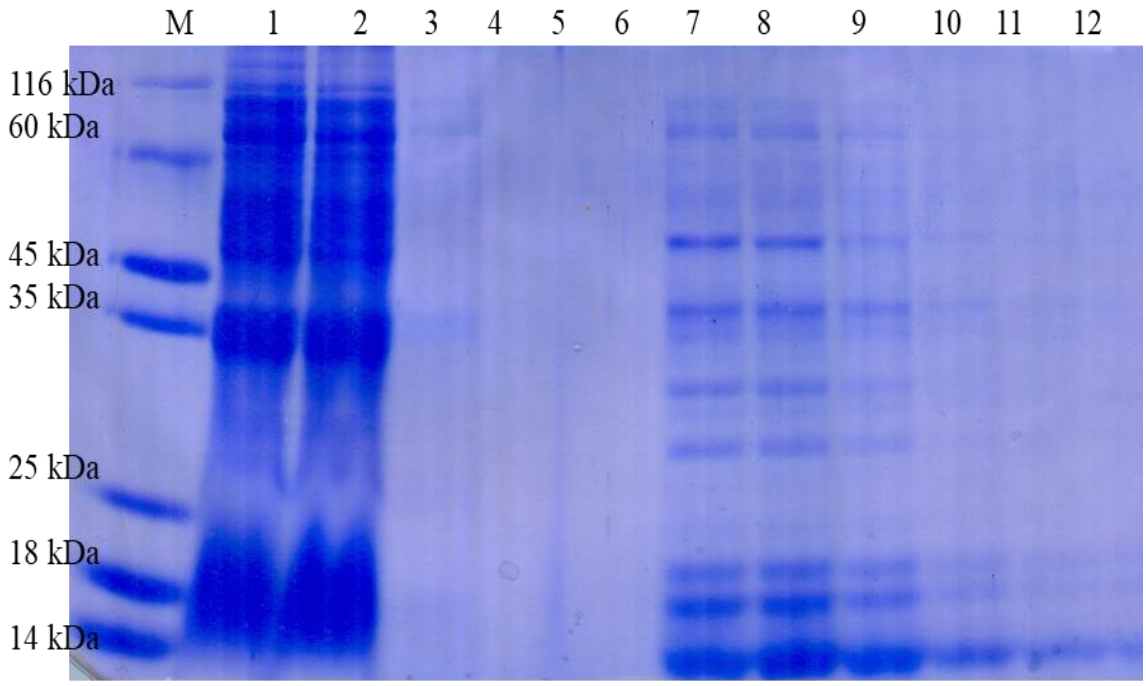


Plate 52 Standardisation of rBBrMV CP purification under denaturing conditions

M: protein molecular weight marker; Lane 1: supernatant; Lane 2: pellet; Lanes 3-5: wash; Lanes 6-12: elutions

lowered in the denaturing conditions and many more non-specific proteins were eluted with the recombinant protein.

When washes were given without the denaturing agent and elution of the protein was done in the absence of urea, the protein was insoluble and did not elute. Lysis with guanidine chloride also yielded similar results.

4.3.14.3 Purification of pGEX/rBBrMV CP

Since the rCP expressed in pRSET-C was insoluble, purification was attempted after cloning the gene to pGEX-4T-2 to the Sma1 site. The recombinant plasmid was transformed in to Rosetta pLysS and BL21 pLysS and overexpression was checked. The overexpression in both the hosts were similar so Rosetta pLysS was selected for purification as rare codons were present in the gene. Induction was done at 16 °C for 12 h using 0.5 mM of IPTG as standardised in the previous experiment (Plate 46).

The 60 kDa band in the protein profile of the induced fraction, which was absent in the uninduced fraction was confirmed to be the protein of interest by Western blot (Plate 47). The protein was highly soluble. For purification Rosetta pLysS harbouring pGEX/BBrMV CP was cultured in YT media. The cell was lysed and the supernatant containing the fusion protein was subjected to GSH sepharose affinity column chromatography as described in the methods section.

A single band at 60 kDa was observed in the elutions without host protein contaminations (Plate 53A). However, the bands were faint and on storage degradation was observed. Induction at 30 ºC reduced the yield of fusion protein in elution and proteolysis was seen to have increased (Plate 53B). The fusion protein was dialysed to remove glutathione and the protein was concentrated by addition of 10 per cent glycerol in the dialysis buffer.

4.3.15 Cleavage of affinity tag from the fusion protein

In order to further characterise the rBBrMV CP, the affinity tag in the fusion protein was cleaved off using thrombin. The conditions for cleavage were standardised and it was observed that on-column cleavage of the protein was much better than in-solution cleavage. Maximum efficiency of thrombin was observed when incubated at room temperature for 8 h, however, protein degradation was also observed. Even though lysis of the induced cells

were carried out in the presence of protease inhibitor cocktail, protein was subjected to partial proteolysis during the purification and tag removal processes (Plate 54A).

Western blot analysis to assess the immunogenicity of untagged rBBrMV CP as antigen was carried out using PVBV CP specific antiserum. The antigen was highly immunogenic. Degradation of tagged protein on storage was observed. However, all the degraded bands were observed to be that of rCP (Plate 54B). The on-column cleavage produced more intact bands of the protein than in solution cleavage.

4.2.16 Quantification of rBBrMV CP

The quantification of both rBBrMV CP fusion and untagged protein was carried out using Bradford's assay. The dialysed protein was used for the assay. Standard graph was plotted using various concentrations of BSA (Fig 40). The concentration of fusion protein was estimated to be 0.845 mg/ ml from 1 L culture and that of the cleaved protein was 0.342 mg/ ml.

4.2.17 *In silico* analysis of BBrMV CP

4.3.17.1 Fold analysis

The folding of rBBrMV CP was studied *in silico* using FoldIndex analysis. It was evident from the analysis that the protein contained two disordered region one at the N terminal and one at the C terminal apart from other minor disordered regions in the protein (Fig 41). The N terminal disordered region was 32 amino acids long and that in the C terminal was 47 amino acids long. Two DAG motifs present the protein were both observed in the disordered region. In case of the fusion protein the C terminal of the tag and the N terminal of the recombinant protein was disordered facilitating the cleavage of the affinity tag.

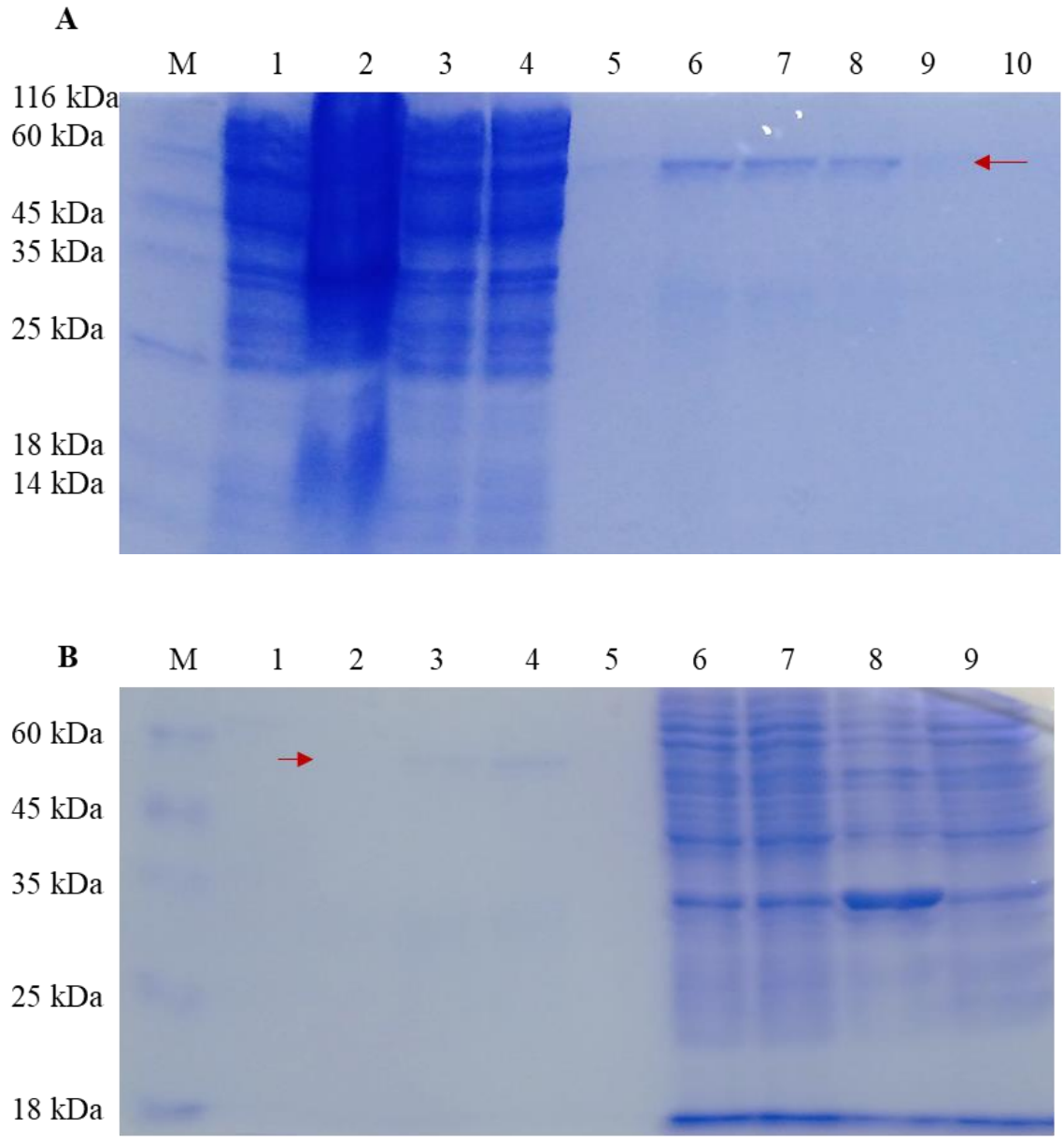


Plate 53 **Purification of pGEX/rBBrMV CP in Rosetta pLysS. A** induction at 16 ºC M: protein molecular weight marker; Lane 1: supernatant; Lane 2: pellet; Lane 3: flow through; Lane 4: wash; Lanes 5-10: elutions **B** induction at 30 ºC M: protein molecular weight marker; Lanes 1-5: elutions; Lane 6: wash; Lane 7: flow through; Lane 8: supernatant; Lane 9: pellet

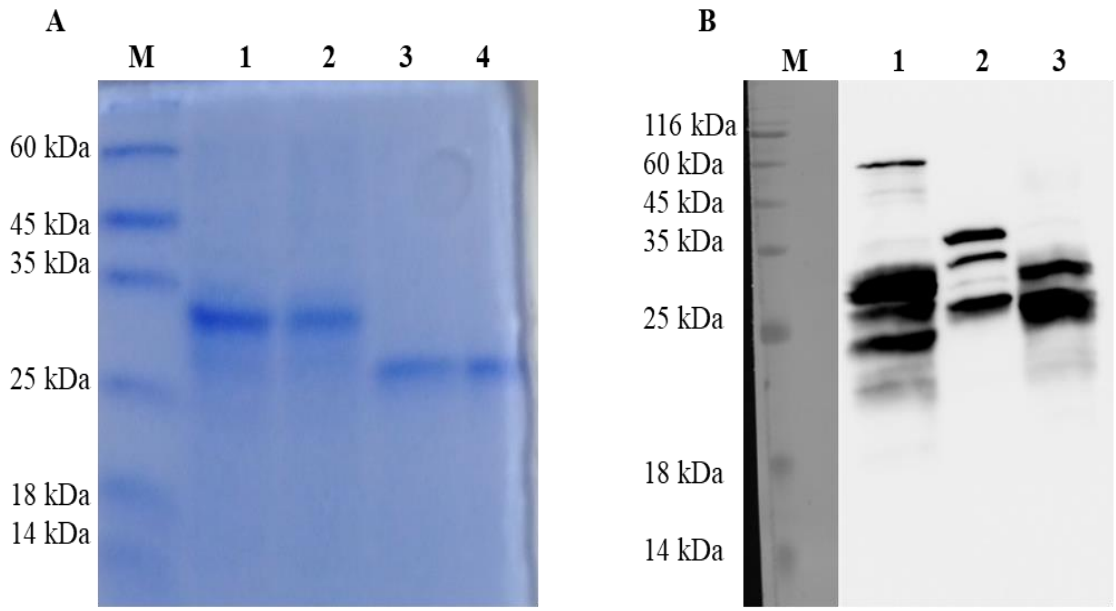


Plate 54 Cleavage of affinity tag from the fusion protein using thrombin. A SDS-PAGE gel profile of cleaved protein. M: Protein molecular weight marker; Lanes 1-2: on column cleavage of thrombin for 8 h, 12 h; Lanes 3-4: cleavage of fusion protein in solution for 8h, 12h. B Western blot analysis. M: Protein molecular weight marker; Lane 1: uncleaved protein; Lane 2: on-column cleaved protein; Lane 3: in-solution cleaved protein

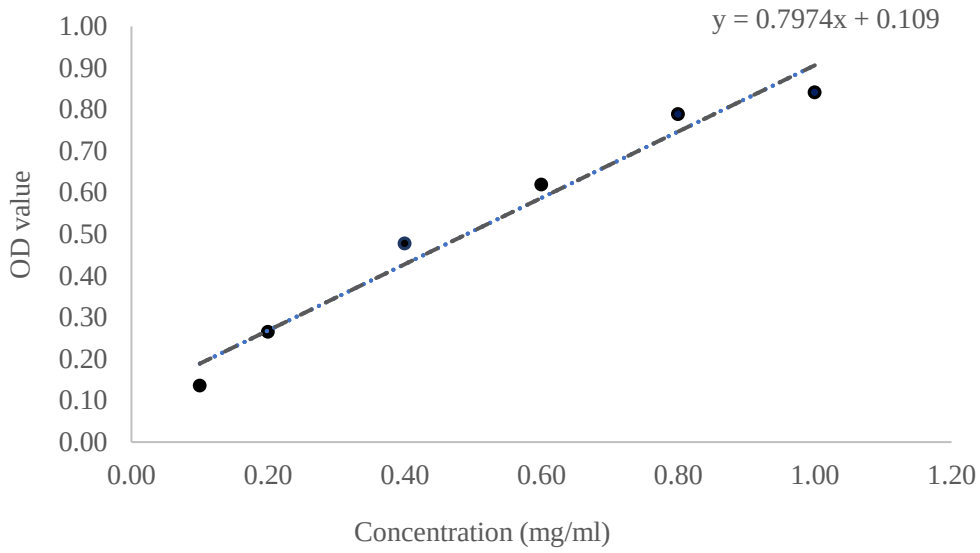


Fig 40. Standard graph plotted to determine the concentration of rBBrMV CP using Bradford's assay

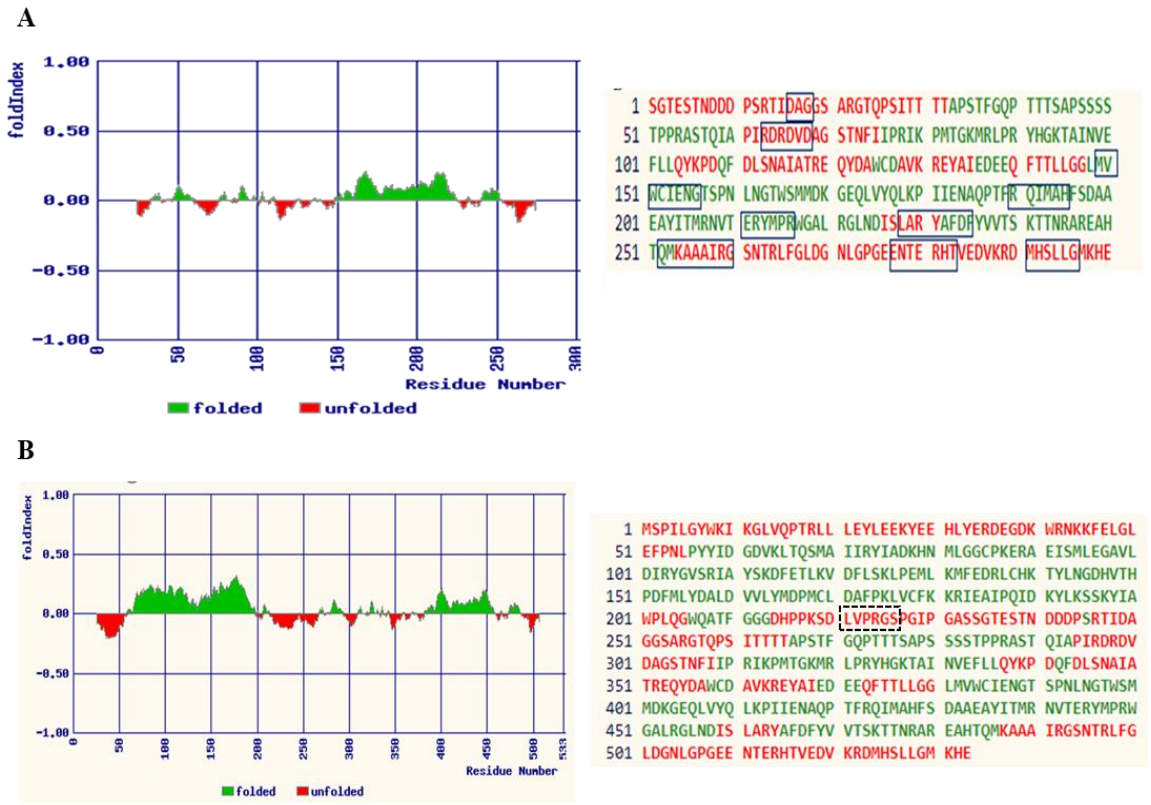


Fig 41. Assessment of protein folding by FoldIndex analysis of A rBBrMV CP **B** fusion protein. The blue squares are the conserved amino acids throughout all the isolates. The black dotted square represents the thrombin cleavage site in the fusion protein.

4.3.17.2 Secondary structure prediction

Secondary structure of rBBrMV CP was predicted *in silico* in PSIPRED workbench and is graphically represented in Fig 42. Secondary structure of BBrMV consisted of coiled coil domains, helical domains and extended strands. The secondary structure of all the isolates displayed a similar pattern. The essentially disordered N and C terminal of the CP is with coiled-coil domain. Four major helical domains were observed from 118-133 (H1), 138-154 (H2), 189-219 (H3) and 244-258 (H4) and two antiparallel β sheets were present between the H2 and H3. Two β strands in MT818185 (the most homologous isolate) is missing in MT818187 (most diverse isolate). The confidence level of helices was high. However, between the isolates most of the difference was detected in the N terminal.

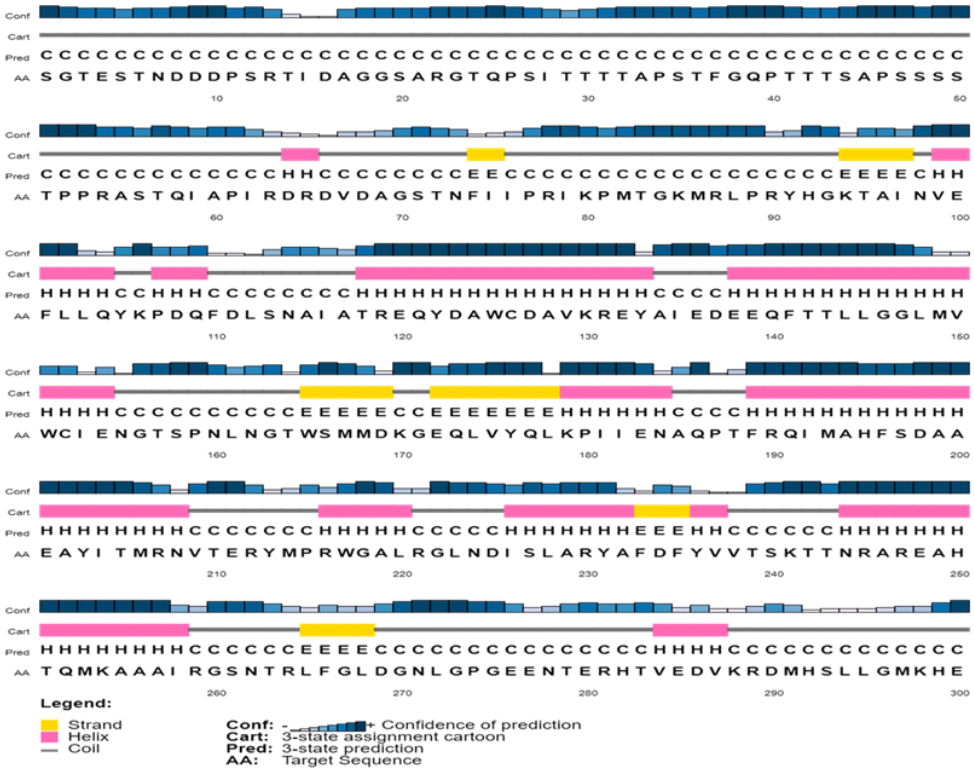


Fig 42. Secondary structure prediction of BBrMV CP

4.3.17.3 Structure prediction BBrMV CP

Structure of the CP of the most homologous and the most diverse isolates were predicted by i-TASSER suite by comparing with structure of *Watermelon mosaic potyvirus* CP (Fig 43). The overall quality factor of the homology model of MT818187 was higher than MT818185 (82.64 and 72.22 respectively). Nevertheless, most of the amino acids except

N terminal and C terminal residues were below 95 per cent cut off value in ERRAT plot. ProSA analysis showed that protein folding energy of the modelled structures were in agreement except for the N terminal (Fig 44). The z score of homology model of MT818185 and MT818187 were -2.3 and -0.96 respectively.

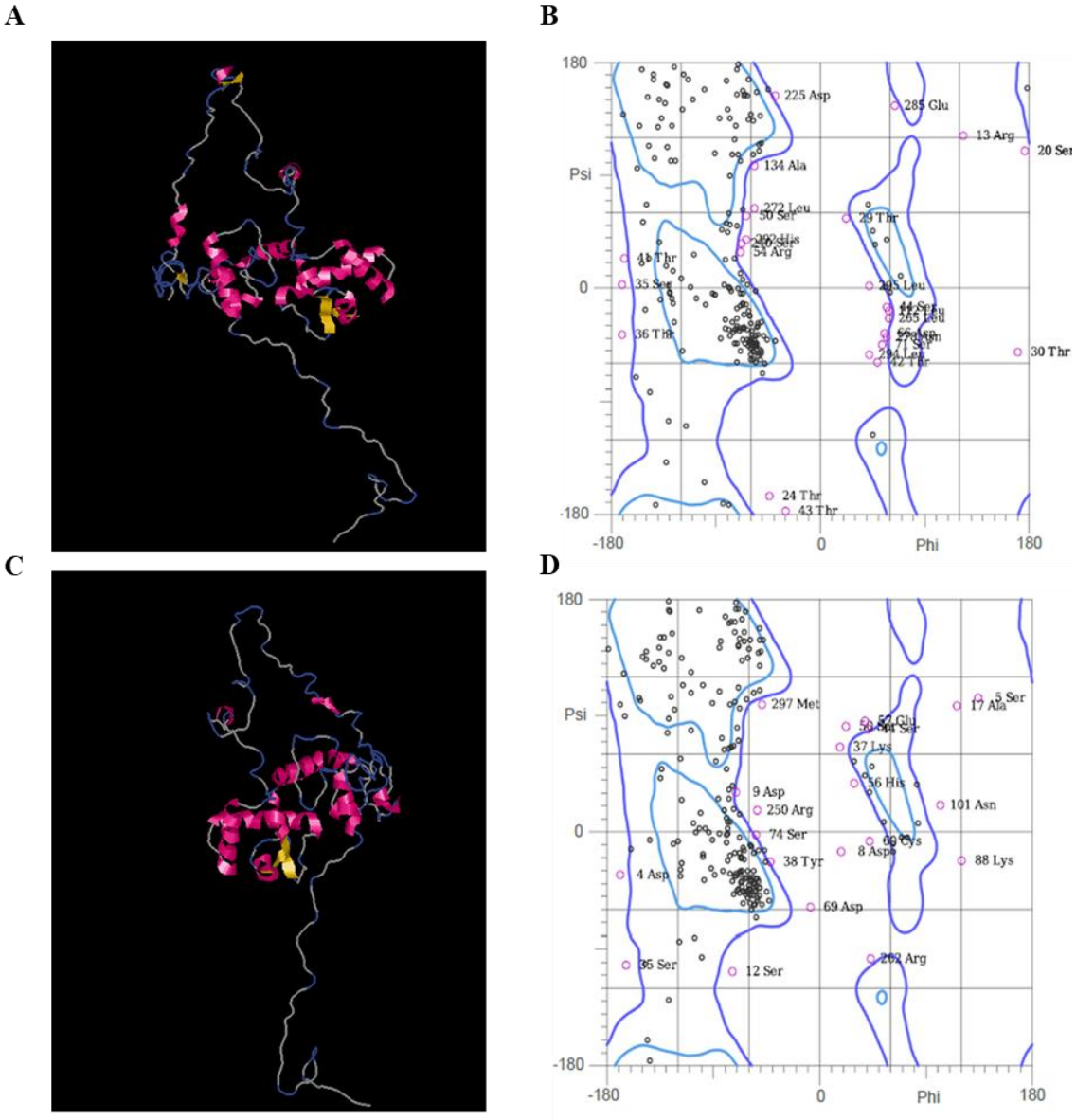


Fig 43. Homology model of MT818185 and MT818187 predicted by i-TASSER. **A** and **B** homology model of BBrMV CP isolate MT818185 and Ramachandran Plot. **C** and **D** homology model of BBrMV CP isolate MT818187 and Ramachandran Plot

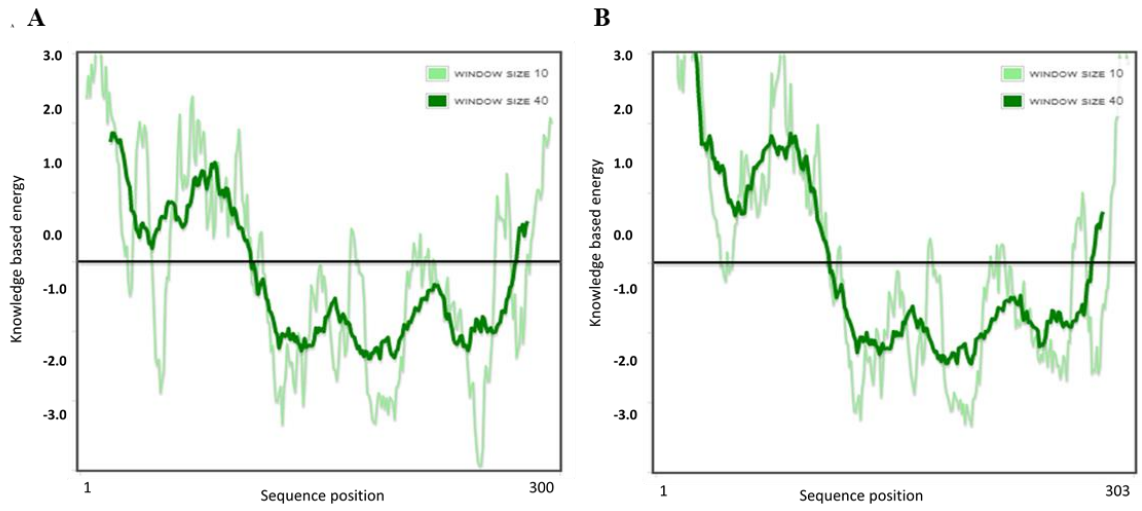


Fig 44. ProSA energy profile of CP of BBrMV isolates A MT818185 and **B** MT818187 modelled by iTASSER. Thin and thick lines indicate average energy over each 10 and 40 residue fragments respectively.

The homology models of most homologous and diverse isolates were aligned by FATCAT for comparison (Fig 45). Evidently amino acid changes had reflected in structure but only in the N and C terminal. However, the basic secondary structure was conserved throughout. Ramachandran plot for homology models showed that 84.6 per cent of the residues were in the favoured and allowed regions together. Surprisingly, 30 residues of MT818187 are in the outlier where as it was 36 in the case of MT818185 isolate (Fig 43). Interestingly, the amino acids that were associated with well established functions of the protein was conserved throughout the isolates. The nucleotide binding site predicted in the CP was in the armpit groove near the C terminal which was conserved in all the isolates (Fig 46). So was the coiled coil domain in the N terminal consisting of DAG motif responsible for aphid transmission.

Although nucleotide and amino acid differences were higher in few isolates of BBrMV, there was not many changes in the structure of the protein. Mutational analysis graph plotted in SusPect represented the predicted effect of mutations at a particular position in BBrMV CP amino acid sequence. The analysis suggested that the amino acid mutations at positions *viz.*, 89 (proline), 105 (tyrosine), 112 (leucine), 114 (asparagine), 121 (glycine), 190 and 225 (arginine), 234 (asparagine), 235 (phenylalanine), 236 (tyrosine) and 281 (arginine) drastic affected the function of the protein. Surprisingly, these residues were all

conserved in all the twelve isolates. Sequence profile graphs at these residues were also analysed to find out the most preferred amino acid residue in a particular position.

4.3.17.4 Epitope prediction

It was important to identify the major and minor epitopes in the CP that is proposed as a potential antigen for developing antiserum intended for immunodetection of the virus. Thus, prediction by Bepipred linear epitope prediction was carried out with the amino acid sequence of GST fusion protein and BBrMV CP untagged protein (Fig 47). Two major epitopes were identified one each in the N and C terminal. Four minor epitopes in the core region of the gene were also identified in the prediction. The major epitopes were also mapped on the structure of the protein by DiscoTope 2.0 (Fig 48). Remarkably, in the GST fusion protein, the major epitope constituted of the C terminal of the tag and N terminal of the CP. This might increase the immunogenicity of the protein while administering to an animal. However, since the fusion protein is also prone to proteolytic cleavage, high chances are that the polyclonal antibody contains more of antibodies specific to the GST protein and that targets the core region of the protein.

4.3.18 Biophysical characterisation of rBBrMV CP

4.3.18.1 Assessment of self-assembly of tagged and untagged virus

The fusion protein with a 25 kDa bulky tag, did not enter into the gradient. The 60 kDa band was seen in the 10 per cent fraction of the gradient. This indicated that the fusion protein did not form any virus like particles (VLPs) (Plate 55). However, the purified untagged protein was ultra-centrifuged, a pellet was obtained which was resuspended in Tris buffer. This was subjected to electron microscopy to understand whether VLPs are formed.

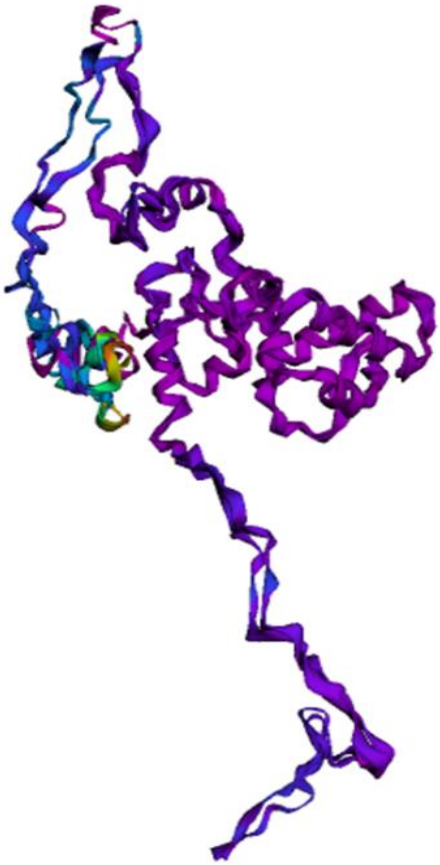


Fig 45. Homology model of MT818185 and MT818187 superimposed using FATCAT. Variable sites are given in different colours. Variability is observed in the N and C terminal

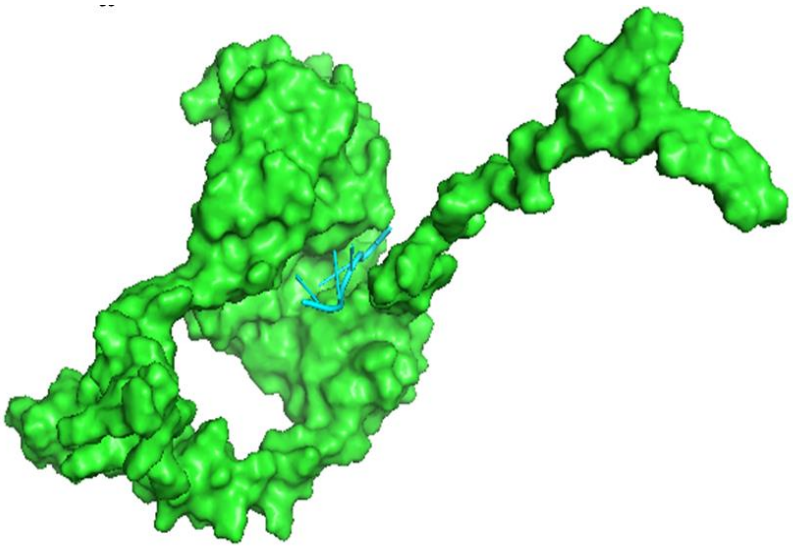


Fig 46. Predicted model of ssRNA- CP interaction in MT818187. Nucleotide binding occurs in the arm pit grove between the core region and C terminal of CP. Nucleic acid is represented as blue chain

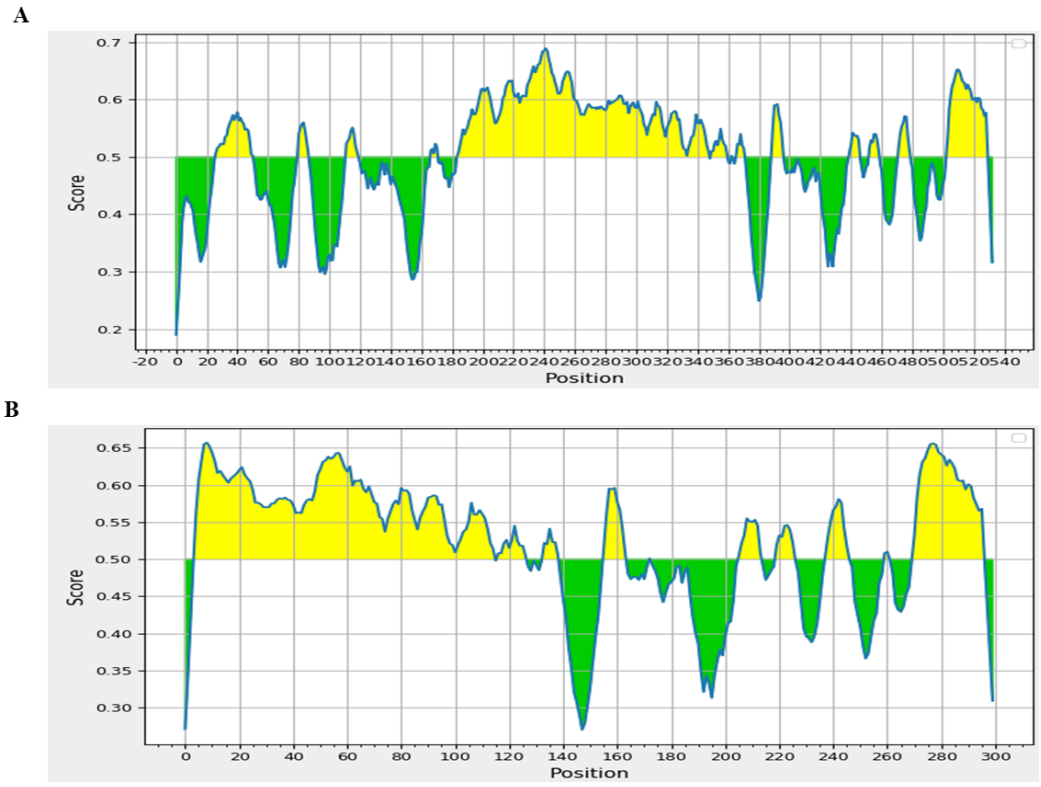


Fig 47. Epitope prediction of **A** BBrMV CP-GST fusion protein **B** untagged rBBrMV CP by Bepipred Linear Epitope Prediction

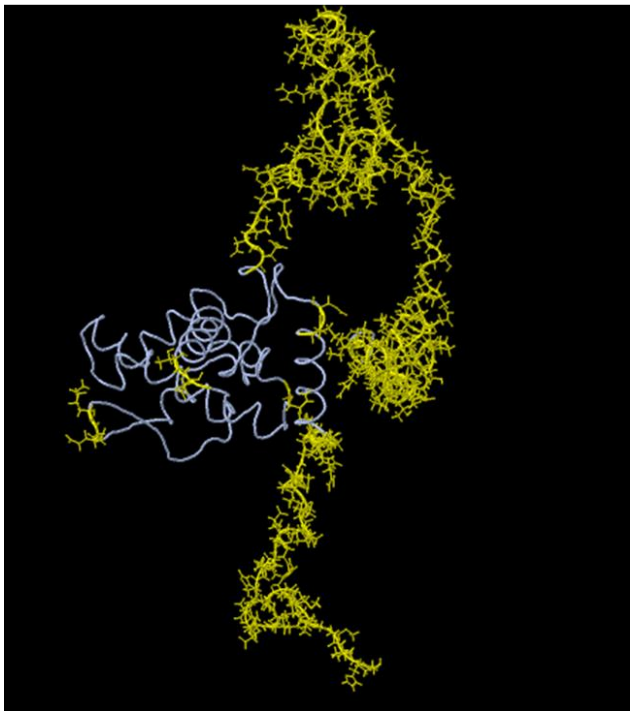


Fig 48. Prediction using DiscoTope 2.0 on the molecular structure of coat protein (MT818185). The major epitopes are marked on the molecular structure of the coat protein in yellow

4.3.18.2 Electron microscopy

Electron micrograph of rBBrMV CP ultra pellet suspected to have formed VLPs is shown in the Plate 56. Octameric ring like structures were observed in the micrograph. The initial step while assembly of virus with helical symmetry especially potyviruses are formation of the disc like structures. Since proteolysis of the rBBrMV CP was confirmed in the earlier experiments, the number of CP having intact N and C terminal would be less to form a complete VLP.

4.3.18.3 Fluorescence spectroscopy

The intrinsic fluorescence spectrum of untagged rBBrMV CP measured in the 300 to 400 nm region had maximum intensity (λ_{max}) at ~344 nm. The fluorescence of fusion protein was compared to the untagged protein. The intensity of latter was five fold higher than the former (Fig 49). However, λ_{max} of both the spectra were at approximately 340 nm. The fluorescence spectra of both rBBrMV CP and fusion protein denotes that the protein is folded with partial disordered region validating the FoldIndex prediction which depicted the N and C terminal of rBBrMV CP to be disordered.

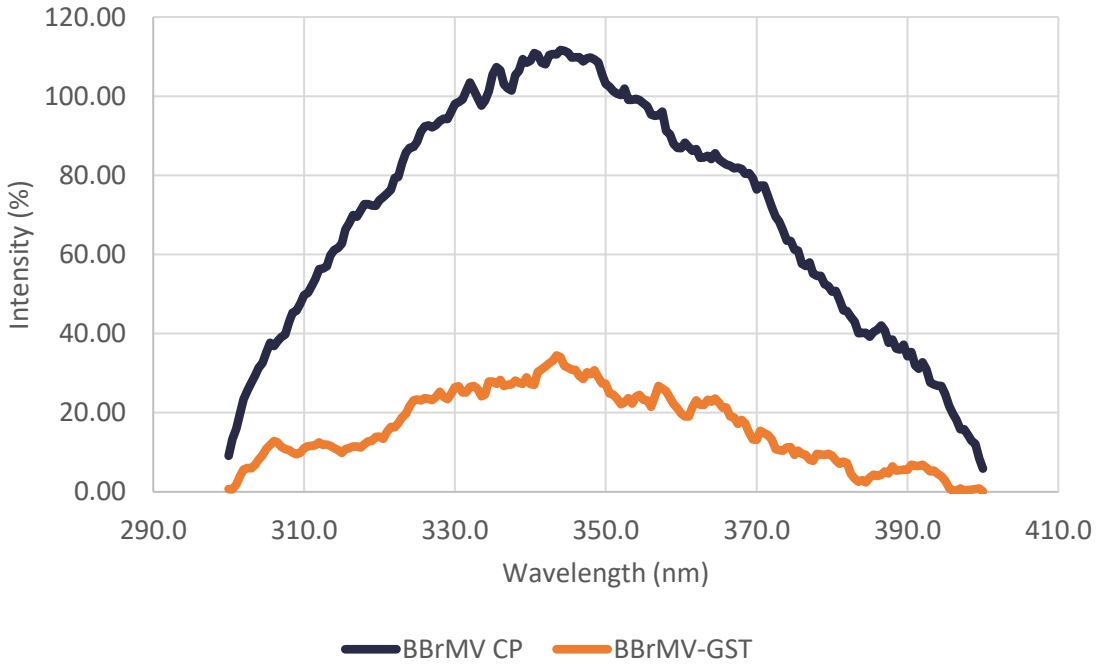


Fig 49. Fluorescence spectra of rBBrMV CP and GST-BBrMV CP

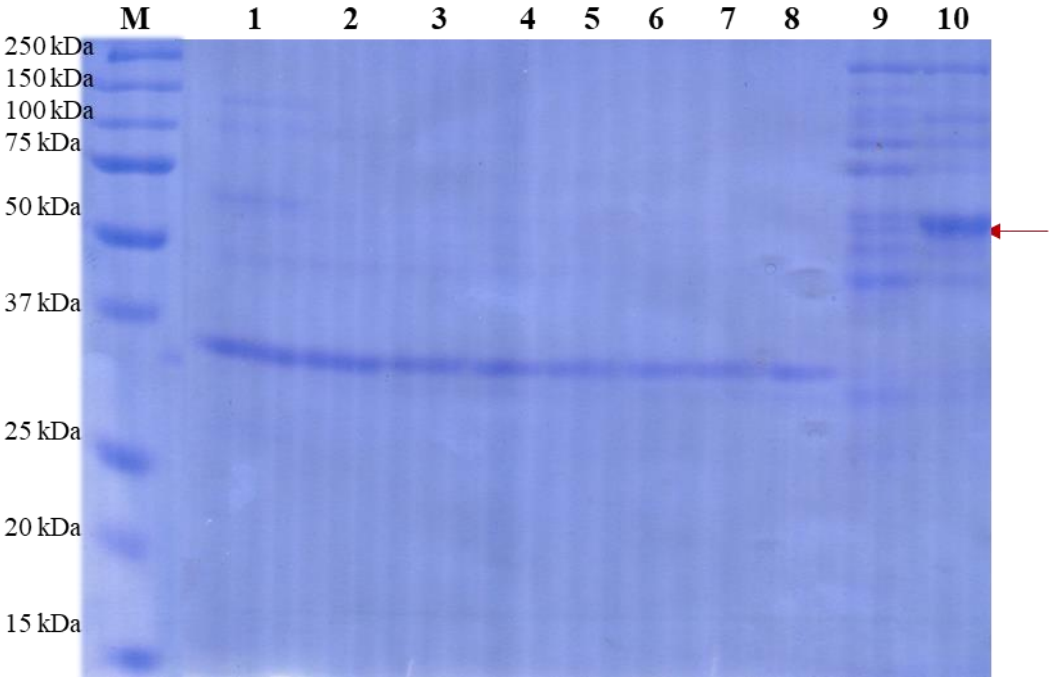


Plate 55 Sucrose density gradient ultracentrifugation of rBBrMV CP-GST fusion protein M: Protein molecular weight marker; Fractions collected from sucrose gradient Lanes 1-3: 40 %; Lanes 4-6: 30 %; Lanes 7-8: 20 %; Lanes 9-10: 10 %. The 60 kDa band corresponding to pGEX/rBBrMVCP is indicated by red arrow

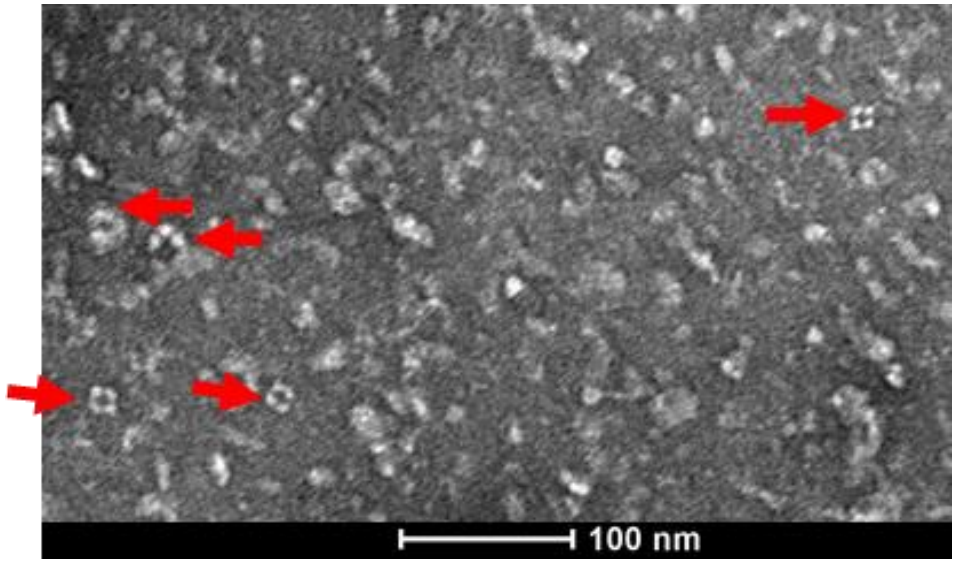


Plate 56 Electron micrograph of ultra centrifuged pure untagged rBBrMV CP

Total number of tryptophan, tyrosine and phenylalanine in the CP was 4, 9 and 10 respectively. However, in the GST fusion protein, 8, 23 and 19 of these amino acids were present. Apparently, few of the phenylalanine, tryptophan and tyrosine residues, responsible for fluorescence are present at the surface exposed N and C terminal which gives the fluorescence and the rest are buried in the well folded region of the protein.

As per the experiments, it was evident that N and C terminals of untagged rBBrMV CP was prone to proteolytic cleavage as other potyviral CP's. Partial degradation of the protein during purification and storage was detected. Consequently, strategies to obtain intact N and C terminal and standardisation of formation of VLP's *in vitro* have to be explored before using it as an antigen for developing immunodiagnostics.

Discussion

5. DISCUSSION

Viral diseases are detrimental to economically important crops especially those that are propagated through vegetative method. Diseases caused by virus in banana especially banana bunchy top and bract mosaic diseases have been reported to be damaging not only in the context of heavy economic losses incurred but also because of the impairment caused to germplasm exchange and export. These viruses remain latent in the infected plants for a long time. When the symptoms express which are confused with nutrient deficiencies, management becomes arduous. These viruses are also notorious as they are spread through plant parts used for propagation and insect vectors. Reports of new BBTV isolates being present in North Eastern part of India apparently resulted from recombination (Banerjee et al., 2014) and BBrMV widening its host range to flowering ginger and cardamom (Siljo et al., 2012; Zhang et al., 2016), calls for the immediate attention of research community, especially in Kerala where these diseases are endemic.

The thesis entitled "Development of recombinant coat protein for immunodiagnosis of banana bunchy top and bract mosaic virus" was formulated with the objective of tracing the evolution of these viruses in Kerala, study its molecular diversity and express coat protein in a bacterial system as a potential antigen for serodiagnostic assays. The results obtained during this study is discussed in this chapter of the thesis.

5.1 SURVEY AND SAMPLE COLLECTION

A total of 10 districts were surveyed during 2017-18 to study the symptoms and collect samples for further evaluations. One to two blocks per district were surveyed. The latitude and longitude of the locations surveyed were recorded and plotted on to the map of Kerala (Plate 57). Twenty-seven and thirty samples suspected to be infected by BBTV and BBrMV respectively were collected during the survey.

Hitherto, surveys for studying various aspects of BBTV and BBrMV have been conducted by various authors. In 2005, a survey was conducted to characterise the geographically different isolates of BBTV in India (Selvarajan et al., 2010b). Rodoni et al., (1997) had studied the geographic distribution of BBrMV isolated from all over the world and tried to explain the evolution and relatedness of these isolates. Similarly, BBrMV virus isolates collected from various parts of the country during 2006-2013 have been subjected to genetic diversity and recombination analysis (Balasubramanian and Selvarajan, 2014).

5.2 SYMPTOMATOLOGY AND DISEASE INCIDENCE

Ten districts divided into Northern, Southern and Central zones were surveyed to study varietal variation in the symptom and whether the molecular diversity of the virus is reflected in the phenotypic manifestation of the disease. However, no such varietal or spatial variations were observed during the present study. All the conspicuous symptoms of BBTD and BBrMD were observed during the study.

The symptoms observed in BBTV infected plants were characteristic discontinuous dark green flecks and streaks of variable length resembling Morse code on the midrib, sheath and veins of leaves and petioles (Plate 3). However, the disease is identified by the progressively produced short and brittle leaves with wavy margin which becomes yellow and later necrotic. These leaves that bunch together gives the disease its name (Thomas et al., 1994; Estelitta, 1998). Previously, Nelson, (2004) had discussed in detail about these symptoms caused by the infection of BBTV. Symptoms like stunting of the young plants and also the propagules emerging from the infected mother plant is readily identified by the farmers. Thomas et al., in 1994 reported that rarely, bracts of male flower buds become leaf like and display dark green dots and streaks. Estelitta (1998) reported constriction of newly formed bunches by the pseudostem leading to its splitting. The reduced size and quality of bunches were reported to considerably diminish the market value of the crop. However, this symptom was not observed during the present study. Although, the plants infected at a later stage do not display typical bunching of leaves nor form bunches, 100 per cent yield loss is recorded in such plants (Kumar et al., 2015a). It was also reported that BBTV infected banana was conducive for secondary infections by fungi and bacteria causing root decay and heart rot especially in humid conditions (Estelitta, 1998).

The virus has a latent period between the inoculation to symptom expression that ranges between 19-125 days reliant on the weather, stage of infection and cultivar as reported by various workers (Allen, 1978; Hooks et al., 2008). However, as discussed before the young suckers developing from the infected suckers show symptoms from the time of emergence.

The characteristic symptoms of BBrMV infecting banana identified during the survey, irrespective of variety and geography were, reddish to dark purple parallel or diamond shaped streaks on the bract, pseudostem and midrib of the leaves, seldom on fruits (Plate 34). Similar symptoms were observed and recorded by various authors (Rodoni et al., 1997; Thomas et al., 1997; Selvarajan and Jeyabaskaran 2006). Necrotic streaks on pseudostem

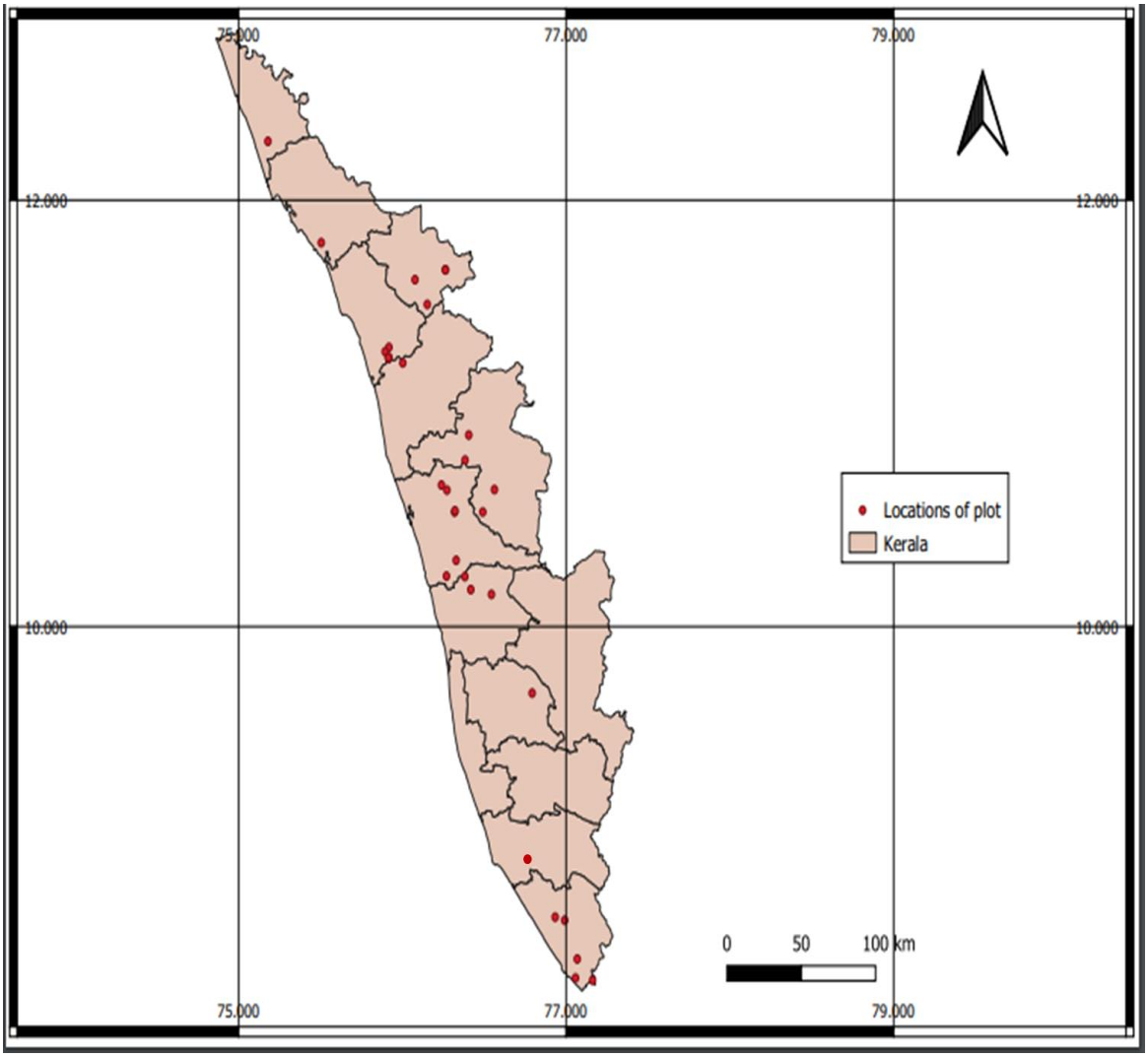


Plate 57 Locations surveyed plotted on Kerala map using qGIS

were seldom observed in the infected sample. However, previously, necrotic spots on fruits, leaves and midrib were also reported apart from the pseudostem (Selvarajan and Jeyabaskaran 2006). Traveller's palm like arrangement of the leaves were observed during the survey as reported by Balakrishnan et al., in 1996. There are reports of peduncle length varying significantly in the bunches along with reduction in size of fingers and weight of the bunches. Interestingly, reports say that the sweetness of the fruit increases in bunches of infected plants. Unlike BBTV infected plants, 30-70 per cent yield loss have been reported for BBrMV infected banana (Magnaye and Espino 1990; Thangavelu et al., 2000; Cherian et al., 2002). The bunch weight, length and girth of fingers in BBrMV infected banana were observed to reduce significantly irrespective of the variety (Estelitta et al., 1996).

The BBrMV isolates infecting cardamom was characterised by similar symptoms. Continuous or discontinuous spindle-shaped light green to yellow streaks along the veins and mid rib coalesce and veins appear yellow or light green in colour. The disease was also named as chlorotic streak due to formation of intraveinal chlorotic steaks. The spindle shaped mottling was also recorded on the pseudostem (Siljo et al., 2012). Similarly, BBrMV infecting flowering ginger, induced symptoms like streaking, mosaic and a severe cupping of leaves, besides browning of flowers and reduction in their size and shelf life (Zhang et al., 2016).

During the survey, disease incidence of BBTV and BBrMV were calculated. Per cent disease incidence was calculated district-wise, block-wise and also zone wise. The PDI throughout the state at the time of survey divided into North, Central and South zones was represented as box plot (Fig 50).

The least incidence of BBTD and BBrMD were recorded in the Southern part of Kerala (0.567 and 0.426 % respectively). Maximum BBrMD was recorded in the Central zone (8.327 %) and that of BBTD in Northern Kerala (1.011 %). BBTD incidence was observed highest in fields where, sanitation was not maintained. Moreover, from the details collected during the survey, it was evident that farmers adept rouging off of the infected plants all over Kerala as a primary management practice.

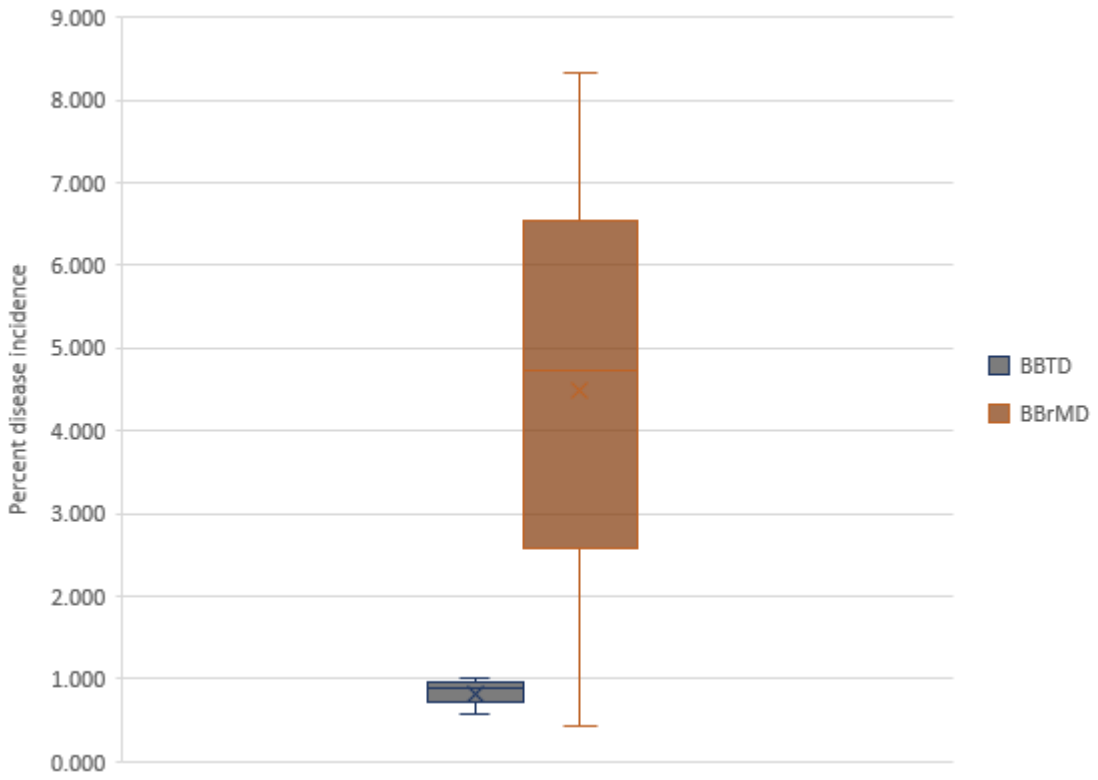


Fig 50. Per cent incidence of banana bunchy top and bract mosaic diseases in Kerala represented as boxplot

During the survey for characterisation of BBrMV infection in cardamom, zero to 15 per cent incidence was recorded in Kerala (Siljo et al., 2012). Increase in the host range of the virus especially in perennial crops, will result in presence of inoculum throughout the year.

5.3 DETECTION OF THE VIRUS FROM FIELD SAMPLES

Preliminary screening of field samples was conducted by DAC-ELISA. The young leaves of samples collected from apparently infected plants were screened in the laboratory (Plate 4 and 35). Out of 25 BBTV samples collected, 20 were tested positive and out of 30 samples 25 tested positive for BBrMV. Polyclonal antibody specific to the viruses were procured from NRC, Banana, Trichy and used for detection of the viruses.

Initially, diagnosis of BBTV was reported to be based on serology where in the antigen was detected based on polyclonal and monoclonal antibodies (Thomas and Dietzgen 1991; Wu and Su, 1990b). Several variants of ELISA have been used routinely for screening virus

free banana micropropagated plants, detection of the virus in field samples and aphid vectors (Wu and Su, 1990b; Thomas and Dietzgen, 1991; Geering and Thomas, 1996; Selvarajan et al., 2010a; Yasin et al., 2015). The DAC- ELISA assay is reported to detect up to 10 ng of antigen. According to Hook et al. (2008), in artificially challenged plants, BBTV was detected after 12–25 days of challenge inoculation by ELISA depending on the genotype and stage of infection. It was established that after this period of time virus inoculum was obtained in all parts of the plant especially from the mid-rib portion of the youngest leaf. This was also observed earlier by Estelitta, in 1998. Hence, for the present experiment, sample for ELISA was measured from the mid rib of the youngest leaf. Similarly, BBrMV detection is regularly carried out by ELISA (Espino et al., 1989; Espino et al., 1990; Thomas et al., 1997). Leaf lamina with mid rib was used for isolating antigen in coating buffer for the assay.

5.4 TOTAL DNA/RNA ISOLATION AND AMPLIFICATION OF CP GENE

Total DNA from 17 and total RNA from 12 representative samples detected to be infected with BBTV and BBrMV respectively by DAC-ELISA were isolated and CP genes were amplified by PCR (Plate 5B and 36B). Nucleic acid-based detection has been reported for detection of BBTV and BBrMV using various primer sets. It was experimentally proven that 15 days after inoculation, BBTV can be detected in PCR (Hooks et al., 2008). Primers specific to CP gene and replicase gene have been extensively used for detection of BBTV out of the six segments. In case of BBrMV, HC-Pro, CP and VPg are targeted for detection in the ssRNA genome.

Although nucleic acid based detection is highly specific and sensitive than serological assays, it is greatly affected by the nucleotide sequence variation. Especially in case of RNA virus like BBrMV, in which nucleotide variation is reported to be higher. This is explained based on the lack of proof reading mechanism in RNA dependent RNA polymerase (RdRp) replicating the genomic RNA (Domingo and Holland, 1997). It has been established before that the CP gene of potyviruses are hypervariable especially in the N and C terminal (Nigam et al., 2019). The reported primers used for the present study was complimentary to 5’ end of CP and 3’ UTR downstream to the gene (Sankaralingam et al., 2006). In few isolates that gave positive reaction in ELISA, CP gene amplification was unsuccessful apparently due to higher variability in the N terminal.

5.5 PHYLOGENETIC ANALYSIS

Phylogenetic analysis of BBTV isolates based on CP nucleotide and amino acid sequences displayed a delineation of isolates based on geography (Fig 5). Among the Kerala isolates, those collected from Northern and Southern parts of Kerala were strictly differentiated in the phylogenetic analysis. However, isolates collected from the Central zone of Kerala *viz.*, Ernakulam, Thrissur and Palakkad clustered together with either Northern or Southern isolates (Fig 5A).

It has been established earlier that BBTV is divided into two clusters based on its phylogeny and evolution to PIO group comprising of isolates from India, Myanmar, Australia, Pakistan, Egypt, Hawaii, Sri Lanka and Tonga and SEA group including the isolates from China, Japan, Taiwan, Indonesia, Philippines and Vietnam (Yu et al., 2012). This categorisation was initially based on DNA-R and was proposed by Karan et al. (1994). Subsequently, other genomic segments were also analysed and similar pattern was testified (Vishnoi et al., 2009; Selvarajan et al., 2010b; Yu et al., 2012; Banerjee et al., 2014). All the Kerala isolates clustered with the PIO isolates as expected. A monophyletic assembling of BBTV isolates from Tripura based on the CP gene was in agreement with the present study (Das and Banerjee, 2018).

In contrast to the result obtained for BBTV isolates, BBrMV isolates did not cluster together based on geography or host (Fig 27). It was evident from the phylogenetic analysis that they all diverged from a common ancestor at different point of time, newly evolved being the two Kerala isolates MT818177 and MT818187 indicated by its longer branch lengths.

Rodoni and co-workers (1999) pointed out that genetic diversity will be maximum at the place of origin. A common ancestor for BBrMV isolates from Philippines, India and other Asian countries, emphasised the chance of virus transmission through planting materials at different point of time (Rodoni et al., 1999). Even the clustering of all the South Indian BBrMV isolates evaluated in the present study indicate the virus transmission through planting material exchange between these regions due to negligent domestic quarantine laws and regulations.

Principle co-ordinate analysis (PCoA) also was in tandem with the phylogenetic studies of BBTV (Fig 6) and BBrMV (Fig 28). Among the BBTV isolates, one each from Thailand

and China clustered separately from other SEA isolates similar to the pattern of the most diverse Kerala isolates.

Evolutions and its complexities of *Nanovirus* have been explained based on the multipartite genome organization and the unusual association of several replication-competent components with viruses. More interestingly, nanoviruses have been suggested to have switched hosts from plants to a vertebrate and then recombined with a vertebrate-infecting virus, giving rise to circoviruses (Gibbs and Weiller, 1999). There are ample evidences for the same (Fu et al., 2009).

5.6 DIVERSITY ANALYSIS OF BBTV AND BBrMV BASED ON CP GENE

Sanger dideoxy sequencing helped to identify sequences of approximately 1.1 kb length of CP of both viruses. To obtain full length sequence of the PCR product sequencing was carried out using the forward as well as the reverse primers. Complete nucleotide sequences of all the six gene segments of BBTV Tamil Nadu isolate were determined by an ABI prism Big Dye Terminator Kit (Selvarajan et al., 2010b). Similarly BBTV isolate from Pakistan and India have also been characterised (Wanitchakorn et al., 2000b; Amin et al., 2008). Few other researches included characterisation of one or few segments of BBTV (Wanitchakorn et al., 1997).

Partial BBrMV genome was characterised earlier by many researchers (Balasubramanian and Selvarajan 2012; Balasubramanian et al., 2014) and complete characterisation of isolates from Philippines (Rodoni et al., 1997) and Trichy was also carried out (Balasubramanian and Selvarajan, 2012). Based on these studies, the CP gene obtained was annotated and complete coding region was identified for further studies.

There are many contradicting reports of variability of BBTV in India. The BBTV Indian isolates have been reported to be less diverse (Kumar et al., 2015a) until a new isolate from Umiam, Meghalaya was identified to be highly assorted from all other Indian isolates (Banerjee et al., 2014). Similarly in 2018, Das and Banerjee stated that some isolates from North Tripura was distinct from the reference population. Whereas, high variability and rapid evolution of favourable mutations were expected in the genome of BBrMV, being RNA virus. In this context, molecular diversity of BBTV and BBrMV isolates from Kerala was assessed based on the CP gene obtained after PCR amplification and sequencing. All the sequences of BBTV and BBrMV isolates were aligned with isolates retrieved from the

database using MUSCLE and further analyses were carried out. A total number of 65 and 56 isolates were considered for analysing diversity of BBTV and BBrMV respectively.

5.6.1 Sequence homology

Supposedly, the variability of BBTV is lower compared to RNA viruses infecting banana like BBrMV and CMV, being a DNA virus. However, significant variability of 1.77 per cent in the amino acid sequences of BBTV CP within the SEA group was reported and correlated with evolution during the period of time the virus has been present in this region (Wanitchakorn et al., 2000). Similarly, BBTV has been identified in India since 1943 and thus genetic variability among isolates of the PIO groups also exists which was reported earlier (Selvarajan et al., 2010b). Although, Kerala isolates have been included in such studies, no studies devoted exclusively to the Kerala isolates are reported.

The nucleotide and amino acid sequence homology between the CP genes of BBTV and BBrMV Kerala isolates were calculated and matrices were formed. Also, nucleotide homology between the isolates in the entire dataset were used for further comparisons. It was observed that BBTV sequence diversity was less compared to that of BBrMV. However, certain isolates belonging to South Eastern Part of Asia, were distinct from other isolates.

In case of BBTV, nucleotide sequence homology between the isolates ranged from 89-100 % (Fig 51A). According to ICTV, overall nucleotide sequence identity of less than 75 per cent or greater than 15 per cent difference in amino acid sequence identity is indicative of a distinct species in nanoviruses (Vetten et al., 2005). Thus, it is evident that all the isolates considered in this study belong to the same species. From the table it was evident that most of the sequence pairs had homology between 99-100 per cent which indicates the lesser sequence variability. This dismissed the hypothesis of the Kerala isolates, belonging to PIO group, being mutated rapidly. On the contrary, in *Faba bean necrotic stunt virus* (FBNSV), belonging to *Nanoviridae*, 7.52×10^{-4} and 5.07×10^{-4} mutations per nucleotide were determined for the field samples and laboratory-maintained population respectively (Grigoras et al., 2010). Rate of evolution was determined to 1.78×10^{-3} substitutions/site/year for FBNSV signifying that the molecular evolution rate of this virus is higher compared to that determined for many RNA and DNA viruses (Duffy et al., 2008; Jenkins et al., 2002).

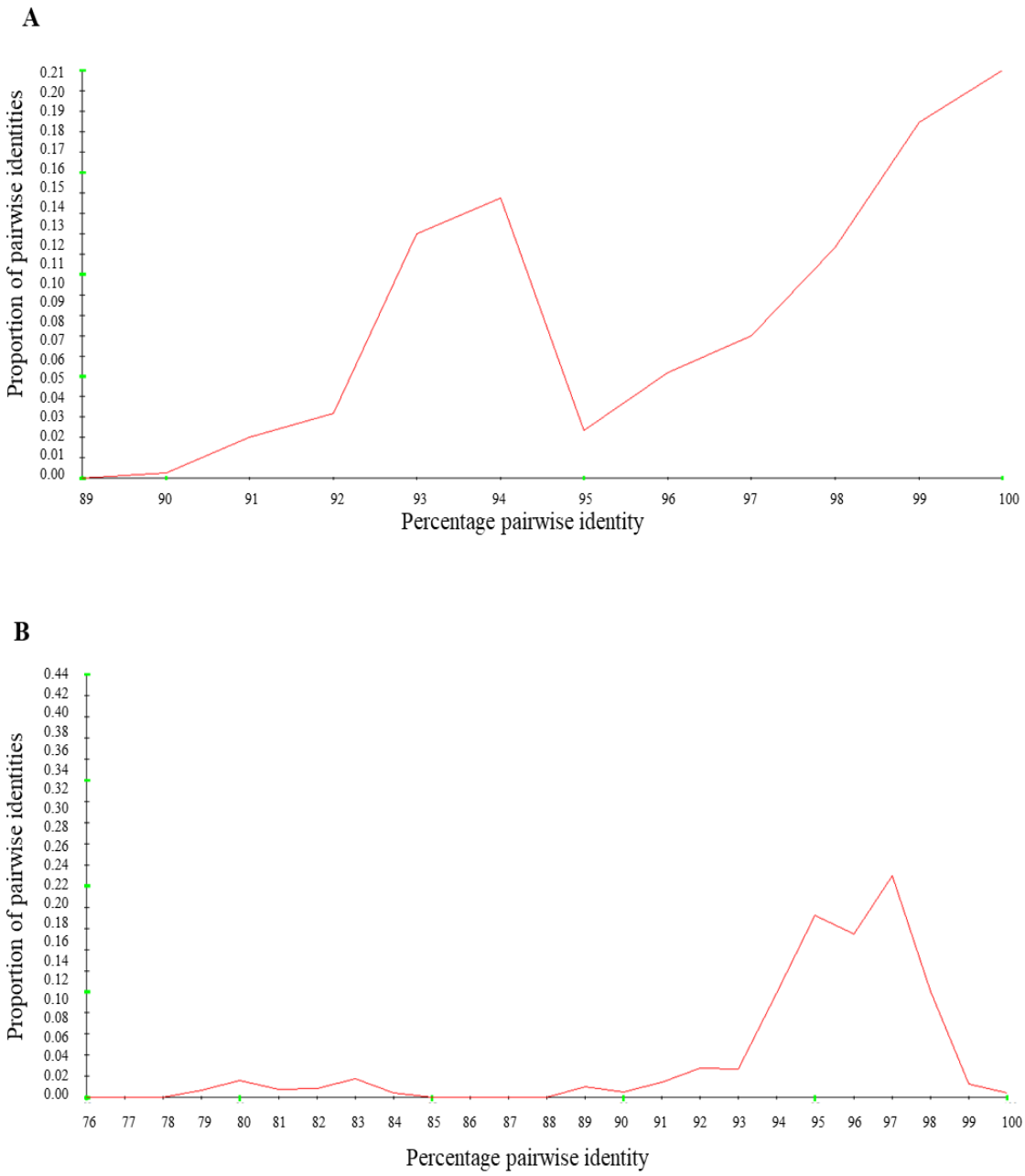


Fig 51. Sequence homology plot of **A** 65 isolates of BBTV and **B** 56 isolates of BBrMV. The percentage pairwise identity is plotted against proportion of pairwise identities.

Even specific studies on BBTV, revealed that the virus is constantly evolving using various mechanisms to cope with the plant defence mechanism and increase its survivability (Hu et al., 2007; Fu et al., 2009). Few BBTV isolates were collected and characterised from Tripura based on the DNA-R and -S segments to concluded that the isolates showed

variability ranging from 2 to 7 % (Das and Banerjee, 2018). Interestingly, even though the rate of mutation in two SEA isolates one each from China and Japan was high, they had negligible diversity in CP gene (He et al., 2000; Furuya et al., 2005). Likewise, on comparing isolates from different parts of the world, diversity was less than 10 % (Hu et al., (2007). The study also highlighted that the segments DNA-S, -R and -N of various other nanoviruses that followed similar trend.

In case of BBrMV, sequence identity varied from 76-100 % (Fig 51B). The proportion of isolate pairs that have sequence identity between 90-100 per cent are more. Interestingly, two Kerala isolates generated in the present study MT818177 and MT818187 were most diverse in the whole dataset with percentage sequence identity ranging from 76-92 %. A high degree of divergence of these isolates from KF385478 and KF385484 (Tamil Nadu isolates) were observed. According to Adams et al. (2005), the optimal species demarcation criterion for potyvirids is 76–77 per cent nucleotide identity of CP gene. They have established that same strains show a nucleotide identity of > 85 %. Thus, the result indicates that the isolates, MT818177 and MT818187 may be different strains of the same species. However, all the other isolates including the once infecting ginger and cardamom may belong to same strain irrespective of the country from which it was isolated. This also points out the spread of the virus from the point of origin to other parts of the world, most probably through planting materials.

5.6.2 Nucleotide diversity and mutation rate

Nucleotide diversity, Pi (π) signifies the genetic diversity in the sequence. In population genetics, Watterson estimator or mutation rate, theta (θ) is a measure of population mutation rate from the observed nucleotide diversity estimated by counting the number of polymorphic sites. Mathematically, it is the product of the effective population size and the neutral mutation rate. Nucleotide diversity and mutation rate in the coding region of BBTV CP and BBrMV CP were assessed. Observing from figures 9 and 31, BBrMV CP, π and θ were highest in the N and C terminal, however no such pattern was detected in BBTV CP. The hypervariability of N and C terminal of CP of other viruses in *Potyviridae* family has been discussed by various authors (Shukla et al., 1988; McKern et al., 1991; Nigam et al., 2019).

5.6.3 Ratio of non-synonymous and synonymous substitution

The ratio of rate of non-synonymous and synonymous substitutions in non-synonymous and synonymous sites respectively (Ka/Ks) of two genes, also denoted as d_N/d_S or ω, is indicative of not only the mutation rate but also the selection pressure acting upon the protein coding gene. Non-synonymous mutations if not lethal, are fixed in the population than synonymous mutations as they mostly have a fitness advantage. This positive selection is indicated by ω value > 1. Value of ω < 1 specifies negative (purifying) selection, whereas, a value equal to or close to 1 indicates neutral mutation (Yang and Nielsen, 2002). The Ka/Ks value was calculated for all the BBTV and BBrMV Kerala isolates by comparing with reference isolates selected from NCBI database.

In most of the BBTV CP sequence pairs, substitutions, both synonymous and non-synonymous were absent. Rate of synonymous substitutions were comparatively higher among the subpopulations collected from Northern and Southern Kerala. Although, all BBTV isolates investigated in a study by Banerjee et al. (2014) depicted more amino acid sequence identity than nucleotide identity in CP gene segment, Ka/Ks values were extremely low or zero indicating negative or neutral selection. Das and Banerjee (2018) assessed the pattern of selection constrain in the coding region of BBTV CP by calculating Ka/Ks ratio and reported negative selection in the isolates. However, they have also reported that in the replicase gene encoded by DNA-R segment, positive or diversifying selection is observed.

Similarly, in majority of sequence pairs of BBrMV CP, ω value were < 1 indicating purifying selection and ω of few sequence pairs were close to 1 indicative of neutral selection of amino acids in BBrMV CP. Three pairs that had ω >1 indicated Darwinian selection which is rare in plant viruses.

The DNA diversity between populations were calculated. The number of net nucleotide substitutions per site between populations, Da (Nei, 1987) was also represented in Fig 33. Apparently, N terminal of BBrMV CP was diverse between the population and among the population itself. Core region is more or less conserved except for Central zone isolates. Number of shared mutations in between isolates from Southern- Central, Central-Northern and Northern- Southern zones were 19, 55 and 32 respectively.

5.6.4 Gene flow

Gene flow between the BBTV and BBrMV populations in Kerala were calculated and very interesting observations were drawn. In BBTV, there is clustering based on phylogeography of the isolates which was also observed in Kerala. This can be attributed to the founder effect wherein, after genetic drift a new isolate establishes in a new geographical location and develops quasi species which in turn has lesser genetic variation from the original population (Templeton, 1980; Provine, 2004) (Fig 52). The statistically significant fixation index between Northern and Central isolates and Southern and Central isolates signifies frequent gene flow between these zones. Statistical significance of the diversity analysis was evaluated by Tajima D and Fu and Li's F and D tests. A negative Tajima value indicated recent population expansion whereas, positive value indicated high degree of polymorphism and balancing selection or decrease in population size. Demographic expansion of the BBTV population after a bottleneck was inferred from the negative and significant values of Fu and Li's D and F tests (-2.147 and -2.104 respectively). However, Tajima D test was statistically non-significant. Ka/Ks values between most of the isolates have also indicated purifying selection earlier. During such demographic expansions, usually, beneficial mutations increase in frequency and are fixed in the population.

In the case of BBrMV CP, gene flow was evident between the subpopulations, so much so that there was no delineation observed in phylogeny and genetic divergence between the populations. All the South Indian isolates considered in the dataset depicted frequent gene flow, substantiating their clustering in phylogeographic analysis. Tajima D test gave a statistically significant negative value (-2.33) which signifies abundance of rare alleles in the population post a recent selective sweep. Some codons with multiple evolutionary paths were detected during the analysis. Tajima D value at synonymous and non-synonymous sites were negative but non-significant in former and highly significant in latter. Fu and Li statistics (Fu and Li' s D and F test) also substantiated this result as the values were also negative and highly significant. All the three tests were performed using total number of mutations.

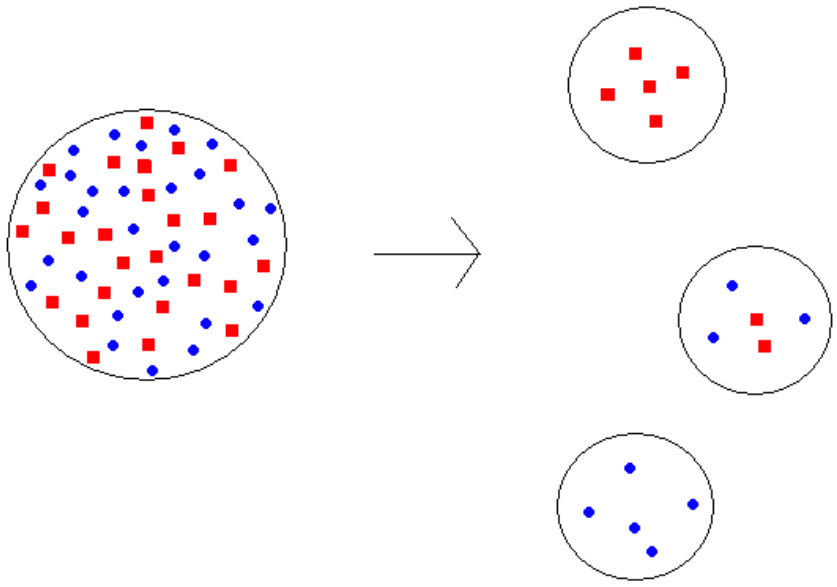


Fig 52. Pictorial representation of genetic drift and founder effect

5.7 CODON USAGE BIAS

Mutational bias is major determinant of codon usage variation midst plant viruses (Adams and Antoniw, 2004). According to them the analysis of CP gene more or less represents the whole genome and gives an insight of the codon preferences of the same. In this context, CUB in CP gene of BBTV and BBrMV were evaluated using various parameters like CAI, Nc and RSCU and its implications are discussed below.

The compositional patterns of the BBTV CP gene were more complex than the commonly observed GC- and/or AU-rich virus genes. For instance, a GC- or AU-rich genome tends to contain codons preferentially ending with either G/C or A/U. BBrMV CP gene depicted a similar trend which support the influence of mutational pressure than selection pressure. Previously, an argument presented by Kumar et al. (2016) the preference of A/U-ended codons over G/C-ended codons in an AU-rich genome of *Equine influenza virus* (EIV) was that mutational pressure determined the CUB. However, BBTV CP despite being AU rich gene, G or C was preferred in the wobble position. This kind of complexity was also observed in CTV CP genes expressed in three different citrus species (Biswas et al., 2019). This unequal use of nucleotides suggests the stimulation of both natural selection and mutational pressure on the codon preferences substantiated by Nc and neutrality plots (Fig 11 and 12). Comparable results were obtained in Zika virus (ZIKV) wherein, overlapping

influences of mutational pressure and natural selection were demonstrated (Butt et al., 2016).

The ratio of observed frequency to expected frequency of a codon, provided, all the synonymous codons for a particular amino acid are used equally is termed as relative synonymous codon usage (RSCU) (Sharp and Li, 1986). Therefore, RSCU values <1.0, 1.0, and >1.0 represent negative, no bias, and positive CUB, respectively. In the present study, a synonymous codon with RSCU values >1.30 was referred to as a high frequency codon (Hfc). Similarly, the effective number of codons (Nc) is used as a measure of CUB of a gene. The value ranges between 20-61, determining the degree of CUB. A very strong CUB is indicated by Nc value equal to or lesser than 35, whereas the gene having Nc value of 61 indicates equal usage of all synonymous codons (Wright, 1990). CUB was evident from the RSCU and Nc values in the CP gene of both BBTV and BBrMV. The high Nc value was indicative of stability of the gene.

The CAI of BBrMV isolated infecting banana, flowering ginger and cardamom implied that they are equally adapted to all the three hosts (Table 18). This in turn substantiated the phylogeny of the isolates that they have all diverged from a common ancestor. The Nc plot and neutrality plot communicated the role of mutation and selection pressure over CUB of BBrMV CP gene. The two most diverse isolates of BBrMV depicted a different pattern in RSCU as well as in the neutrality plot. Few rare codons in the most homologous isolate were optimal or preferred in these isolates and vice versa. However, more studies in depth are required to confirm if these will be selected and fixed in the population. Further clarification can be obtained if Hfc preferred by host is also compared with that of the viruses.

The CUB of CP gene of PRSV and CTV was investigated to obtain similar results (Chakraborty et al., 2015; Biswas et al., 2019). In both the studies, similar Nc plots were obtained and it was substantiated that not only total nucleotide constraints (mutational bias) but also transitional selection and other associated factors were responsible for shaping CUB. However, the results were in contrast to that obtained in *E. coli* and yeasts wherein, large variation in Nc at very similar GC3 values were observed (Wright, 1990; Sharp et al., 1988).

5.8 RECOMBINATION

Recombination has been stated as one of the main driving forces hastening plant virus evolution (Roossinck, 2003; García-Arenal et al., 2003). The evolutionary advantage of genome segmentation in viruses have been hypothesised based on efficiency in replication due to the shorter length of the genome and stability of virion with segmented genomes (Nee, 1987; Chao, 1988; Szathmary, 1992; Ojosnegros et al., 2011). Apart from these advantages, apparently an increase in genetic diversity of the virus population forming quasi species have much more fitness gain. This consequently eliminates harmful mutations and maintain favourable ones through various methods like genomic reassortment (component exchange) among different isolates and or inter or by intra-molecular recombination (Nelson et al., 2008; Hyder et al., 2011).

The Robinson-Fould's matrix (Plate 14A) implies the degree to which phylogenetic trees differ from each other which are constructed from different parts of the alignment. Like RF matrix Shimodaira-Hasagawa (SH) matrix is also a useful tool to visualise the effect of recombination on phylogeny. However, it utilises degree of statistical support for differences between trees rather than the numbers of differences in topological features as in RF matrix. Although, from the the orange/red regions in matrices it is evident that CP region of BBrMV mostly have low degrees of phylogenetic compatibility, recombination hot spots cannot be defined. However, two recombinantion events were identified from recombinant break point matrix between 400-600 nucleotide positions. Hot spots were identified at 400 nucleotide from region count matrx where recombinant breat point matrix also displayed significant recombinant events. Two rombination cold spots, between ~600-800 nucleotides were distinguished from the matrix (Fig 39). However, no such recombination hot spots were found in recombinant region count matrix of BBTV isolates.

The McVean's LD matrix detected pairs of sites amidst analysed gene with unusual linkage disequilibrium pattern. Low marginal likelihood values (>-4.0) were indicative of recombinant cold spots. In the McVean's LD matrix of BBTV, the extremes of marginal likelihood values were very low implying that the recombinant cold and hot spots projected were statistically non-significant (Fig 14). On the contrary, the encircled region in the matrix of BBrMV indicated a low rate of recombination than expected in recombinantion hot spots within CP gene (corresponding to the region inidctaed as hot spots from breakpoint pair matrix).

Hudson and Kaplan's RMin matrix estimated the minimum number of recombination events (Rmin) separating every pair of nucleotide position in an alignment (Hudson and Kaplan, 1985). The Rmin/Distance matrix was a distance normalised version of the RMin matrix *ie.*, it helped to visualise large changes in Rmin that occur over short genetic distances (such might occur across recombination hot-spots). The RMin and RMin/D matrices of BBTV CP gene displayed a dark red region corresponding to recombinant hot spots between 200-400 nucleotide and 510 nucleotide positions.

MAXCHI and LARD matrices of recombination event in BBTV dataset was drawn. These matrices were useful for identifying the statistically optimal positions of breakpoint. The Chi squared values for different breakpoint pairs were colour coded. In LARD matrix, the nucleotide substitution models were also accounted. The dark red colour in the matrices represented statistically significant recombinant breakpoint pairs. Two of the most diverse BBrMV isolates, MT818177 and MT818187 from Kerala were recombinants whereas none of the BBTV isolates were identified as recombinants.

However, Banerjee et al., (2014) had reported a new PIO isolate which was allegedly formed by recombination. The CP gene length was predicted to be 528 bp translating to 176 amino acid long polypeptide of 20.1 kDa in contrast to earlier reports of 19 kDa CP encoded by 513 bp ORF. Nevertheless, most variant among the six genomic segments were identified as DNA-U3 with sequence identity less than 92 per cent on comparison with other PIO isolates. Two significant inter-component recombination signals were identified in DNA-S segment of BBTV isolate from Meghalaya. Curiously, five intragenomic recombination signals were observed in the DNA-R segment of the isolate all in the non-coding region.

Recombination events have been detected in HC-Pro, VPg as well as CP gene of BBrMV earlier (Balasubramanian and Selvarajan 2014; Balasubramanian et al., 2014; Anuradha and Selvarajan 2018). Eleven, one and twenty four recombinants were identified when CP, VPg and HC-Pro respectively were considered to study evolution of BBrMV Tamil Nadu isolates. However, the recombination breakpoints were scattered and hot spots were not recognised. In previous investigations, potential recombinants were detected in HC-Pro gene of *Potato virus Y* (PVY), *Sugarcane streak mosaic virus* (SCSMV) and *Turnip mosaic virus* (TuMV) (Tian et al., 2011; Bagyalakshmi et al., 2012; Nguyen et al., 2013). A noteworthy information put forth after investigation conducted on recombination of HC-

Pro of BBrMV was that a potential recombinant was major parent for other recombination events detected in this study (Balasubramanian et al., 2014). The authors, explained the phenomenon on the basis of banana being clonally propagated and the co-existence of these parental isolates in the same banana clone through aphid transmission resulting in recombinants. Mutation and recombination accelerate evolution of viruses and the population in a particular locality consists of many such strains of the virus. Any diagnostic assay developed, whether serological or nucleic acid based, should be inclusive of these variants that are produced from time to time. Thus, the above analyses were inevitable before proceeding to further experiments.

5.9 PRIMER DESIGNING AND VALIDATION OF NUCLEIC-ACID BASED ASSAY

Primers were designed considering the variability of the isolates prevalent in Kerala for detection of BBTV and BBrMV by amplifying the CP gene which is most abundant in all the stages of infection. In case of BBTV, the coding region did not show much variability and so was the flanking region of the ORF. Thus, primers were designed to amplify 850 bp comprising of 513 bp of coding region and the conserved upstream and downstream region (Plate 7). The BBTV CP specific primers have been designed for nucleic acid-based detection of the virus in field samples as well as in tissue culture plants (Mansoor et al., 2005; Mahadev et al., 2013). Molecular characterisation of complete CP of BBTV from various parts of the world have been carried out through PCR and sequencing (Wanitchakorn et al., 1997; Furuya et al., 2005; Mansoor et al., 2005; Amin et al., 2008; Selvarajan et al., 2010). The CP gene along with 3' untranslated region (UTR) was characterised to infer that the genetic variation among BBTV isolates belonging to PIO and SEA groups are less.

However, more than 90 per cent of the variability in CP gene region in BBrMV is concentrated in the N and C terminals as in other viruses belonging to *Potyviridae* family (Nigam et al., 2019). In the initial experiment, many isolates that gave a positive result in ELISA did not give CP gene amplification using reported primers. This might be due to the fact that these isolates had a highly variable N terminal which was targeted by the primer. The potyvirus ssRNA codes for a polyprotein which is post translationally modified by its own proteases. NIb protein upstream to CP codes for RNA-dependent-RNA polymerase responsible for RNA replication (Revers and Garcia, 2015). Nigam and co-workers (2019)

established that this region is considerably stable and less variable. However, variation is present at the C terminal of the NIb at the junction of NIb- CP which is also the site for NIa cleavage. Hence the conserved NIb region was chosen for designing forward primer and reverse primer was designed form the highly conserved 3'UTR region to give an amplicon of length over 1300 bp (Plate 38). The primer pairs amplified certain samples which were not amplified earlier using primers reported by Sankaralingam et al., (2006) confirming the hypothesis. Degenerate primers have been designed for potyviruses specific to various regions like NIa, NIb, CI etc for detection and complete genome characterisation (Ha et al., 2008). The authors proceeded by amplifying the 5' and central regions of the potyvirus genome using degenerate primers, subsequently amplifying intervening and terminal sequences by virus-specific primers. This strategy was used to obtain the entire genomic sequences of a partially characterized BBrMV isolate from the Philippines. However, no such characterisation has been carried out in the Kerala isolates.

5.10 CLONING OF CP GENE TO VARIOUS EXPRESSION VECTORS

The BBTV CP gene, amplified using primer sets designed for cloning, was cloned into various expression vectors like pRSET-C at Nhe1 and BamH1 sites, pGEX-4T-2 in Sma1 site and pET32a (+) vector at EcoRV and BamH1 site (Plate 11 and 12). The BBTV CP was also cloned to pGEX-4T-2 by RF cloning method (Plate 14). The BBrMV CP gene amplified using reported primers were cloned to pGEM-T easy TA cloning vector (Plate 39). This vector was used previously for cloning and sequencing of BBrMV CP (Rodoni et al., 1999). Later, CP gene amplified using primer set with Nhe1 and BamH1 sites in the 5' and 3' terminals were cloned in to corresponding sites in pRSET-C and to the Sma1 site of pGEX-4T-2 (Plate 40-42). The ligation reaction was set up at molar ratio 1:10 of vector and insert. The transformed colonies were randomly picked to screen for recombinants.

In case of blunt cloning, to Sma1 site of pGEX-4T-2, recombinants were selected by PCR and sequencing. However, restriction digestion to confirm the orientation was also carried out. On digestion with BamH1, the insert which is ligated in the correct orientation was released from the recombinant vector. The BamH1 is present upstream to Sma1 site where the gene is cloned and the 3' end of the gene consist of the recognition site incorporated while amplifying with the designed reverse primer.

For expression, BBTV CP was cloned to pMAL-c2 (Wanitchakorn et al., 1997), pQE-30 (Abdelkader et al., 2004; Shilpa et al., 2016) and pET 28a (Arumugam et al., 2017) expression vectors and BBrMV CP to pProEX-1 (Rodoni et al., 1999) and pET28a (Selvarajan et al., 2020) expression vectors.

Thrombin is a specific protease commonly used to cleave off the affinity tag from many fusion proteins in a wide range of pH and temperatures. The recognition site of thrombin is LVPRGS (leucine-valine-proline-arginine-glycine-serine) which is incorporated in the vectors between the affinity tag and the multiple cloning sites where the gene is cloned. Thrombin cleaves between the arginine and glycine residues. However, the specificity is not absolute (Jenny et al., 2003) and the commercially available protease is expensive. On assessing the protease cleavage sites in BBTV CP, thrombin cleavage site was predicted inside the protein (Fig 17). Moreover, The N terminal of the protein with arginine residues responsible for nucleic acid binding during assembly was also prone to trypsin cleavage. Thus, restriction free (RF) cloning of BBTV CP gene to pGEX-4T-2 was attempted to remove thrombin cleavage site between GST tag and the CP to replace it with TEV protease recognition site. The RF cloning also known as ligation independent cloning (LIC) has been widely used for cloning of any target gene to any vector irrespective of restriction sites without addition of unnecessary sequences and moreover facilitating easy molecular manipulations (Aslanidis and Dejong, 1990; Tillett and Neilan, 1999; Unger et al., 2010; Wang et al., 2014). Since the entire plasmid is replicated in this reaction, mega-primer binding to a daughter molecule fails to expose a free 3′-OH for polymerase elongation, as a result, accumulation of new product is not geometric but linear. The principle of RF cloning is illustrated in Fig 53.

The 550 bp PCR product obtained was used as megaprimer for RF reaction and amplified BBTV CP gene cloned to pGEX plasmid (pGEX/BBTV CP) to obtain 5.6 kb product. In the present investigation, TEV protease site was inserted between the GST tag and CP gene replacing thrombin site. The highly specific 27 kDa catalytic domain of nuclear inclusion protein (NIa) of TEV, recognises the amino acid sequence ENLYFQS/G and cleaves between glutamine (Q) and serine (S) or glycine (G) residues (Daros et al., 1999; Nallamsetty et al., 2004). The TEV protease was overexpressed and purified in-house for the experiment (Plate 24).

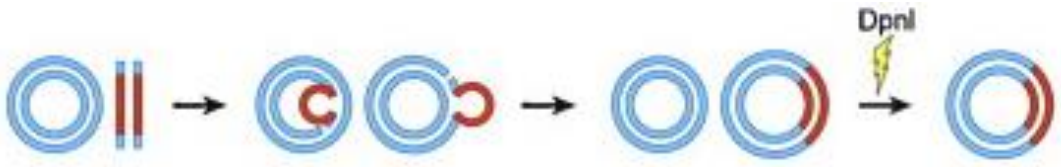


Fig 53. Schematic representation of a typical RF cloning protocol. Hybrid primers are designed with complementarity to the CP gene (red) and the destination expression vector (blue). First round of PCR created a 'mega-primer' comprising the complete coding sequence of CP flanked by sequences complementary to the pGEX-4T-2. During a second round of PCR, the mega-primer initiates replication of the destination plasmid. The plasmid is purified from a DAM+ bacterial stain, DpnI is used to selectively degrade parental DNA after the second PCR reaction, leaving the unmethylated daughter products intact.

5.11 OVEREXPRESSION AND PURIFICATION

The CPs of BBTV and BBrMV cloned in to various vectors were overexpressed in different hosts and later purified by affinity column chromatography. Hosts like BL21 (DE3) pLysS, Rosetta (DE3) pLysS and C41 were used for overexpression of the CP and BL21 (DE3) for overexpression of TEV protease. The cell strains denoted as DE3, lack lon and ompT proteases, also indicated that the cell includes T7 RNA polymerase gene in it's chromosome under control of the lacUV- 5 promoter. Such strains produce protein on induction with IPTG only. The advantage of cell strains with pLysS plasmid encoding for T7 lysozyme is that it suppresses the leaky expression of T7 RNA polymerase prior to induction. Hence the proteins governed by a T7 promoter or any inducible promoter only is translated on induction (Sørensen and Mortensen, 2005a; 2005b).

Viral proteins are expressed in eukaryotic system and there are certain synonymous codons which are preferred in this system that are rare in the prokaryotes. Many errors during translation are incurred from rare codon bias and include premature translational termination, mistranslational amino acid substitutions and frameshift events (Kurland and Gallant, 1996; Sørensen et al., 2003). For instance, codons like AGA coding for arginine are rare in prokaryote system and it is sometime replaced by lysine causing mistranslation (Seetharam et al., 1988; Calderone et al., 1996). The codon bias problem in eukaryote and prokaryote can be bypassed by using hosts like Rosetta which contains plasmid coding for rare tRNAs for AGG, AGA, AUA, CUA, CCC, GGA or by substituting amino acids like arginine in the media, while induction and overexpression of target protein. On analysing rare codons present in BBrMV-GST fusion protein using Rare codon search online tool, 5

rare codons in *E. coli* expression system coding for arginine was found (Fig 54). The CAI was 0.56 only. Thus, Rosetta strain of *E. coli* was used for overexpression and purification of the fusion protein.

```
  1  mspilgywki kglvqptrll leyleekyee hlyerdegdk wrnkkfelgl
 51  efpnlpyyid gdvkltqsma iiryiadkhn mlggcpkera eismlegavl
101  dirygvsria yskdfetlkv dflsklpeml kmfedrlchk tylngdhvth
151  pdfmlydald vvlymdpmcl dafpklvcfk krieaipqid kylksskyia
201  wplqgwqatf gggdhppksd lvprgspgip gassgtestn ddpsrtida
251  ggsargtqps ittttapstf gqptttsaps ssstpprast qiapirdrdv
301  dagstnfiip rikpmtgkmr lpRyhgktai nveflqykp dqfdlsnaia
351  tReqydawcd avkreyaied eeqfttllgg lmvwciengt spnlngtwsm
401  mdkgeqlvyq lkpiienaqp tfrqimahfs daaeayitmr nvterympRw
451  galrglndis laryafdfyv vtskttnRar eahtqmkaaa irgsntrlfg
501  ldgnlgpgee nteRhtvedv krdmhsllgm khe
```

Fig 54. Rare codon search. The one letter code of amino acid in red indicated those coded by rare codon in *E. coli*

The C41 *E. coli* strain is highly efficient in expressing toxic proteins from all classes of organisms, including eubacteria, yeasts, plants, viruses, and mammals which has been validated through rigorous experimentation. These strains are genetically mutated to tolerate and stabilise while expressing toxic proteins. As cell biomass on expressing BBTV CP was less, to rule out the possibility of the rBBTV CP being toxic, C41 strain was used for overexpression and purification. Surprisingly, there was a significant difference in level of expression and recovery of pure protein in C41 strain than the other two strains of *E. coli* evaluated (Plate 18 -20).

The attempts to overexpress BBTV CP in pRSET-C and pGEX-4T-2 was unsuccessful. The protein expressed without bulky affinity tag in the former was unstable (Plate 16) where as in the latter, protein was separated from the GST tag during the process of overexpression and purification (Plate 17). This was predicted to be due to the trypsin cleavage sites present at the junction of C terminal of the tag and N terminal of the protein. The 25 kDa GST band was visible in the SDS-PAGE profile but, 20 kDa band was faint. However, post purification the 20 kDa band was invisible the gel.

Overexpression of ΔpGEX/BBTV CP and pET/BBTV CP was successful and fusion proteins of 45 kDa and 37 kDa respectively were obtained. On purification by GSH or Ni^{2+}-NTA column chromatography, pure protein without any host contamination were obtained

(Plate 22 & 23). Interestingly, earlier successful attempt to purify BBTV CP cloned in to pMAL-c2 vector and overexpressed in *E. coli* DH5α was as maltose binding protein (MBP)-BBTV CP fusion protein (Wanitchakorn et al., 1997). Fairly pure preparations of fusion protein were obtained and was characterised by Western blotting. Other attempts include expression of BBTV CP cloned in to pQE30 and pET28a vectors and transformed to M15 and BL21 cells respectively (Shilpa et al., 2016; Arumugam et al., 2017). The 21 kDa recombinant protein was overexpressed and purified using Ni^{2+}-NTA column chromatography in the former. However, purification of pET28a/BBTV CP in BL21 was not attempted.

The purified ΔpGEX/BBTV CP was subjected to cleavage with the tetrameric TEV protease but the affinity tag could not be removed. Since, BBTV CP was unstable above 4 ºC and the activity of TEV protease below 20 ºC is reported to be less, cleavage of fusion protein might not have happened. On the contrary, Nallemsetty et al., (2004) reported that the activity of certain variant of TEV protease at 4 ºC is only three fold lesser than that at 20 ºC. However, it is recognised if the fusion protein exists in form of soluble aggregates or if the cleavage site is close to ordered structure of target protein causing steric hinderance, the cleavage site may become unavailable for the protease making the fusion proteins an intrinsically poor substrate for TEV protease (Kapust and Waugh, 2000).

From the FoldIndex analysis it is clear that both GST protein and BBTV CP are ordered except the N terminus of CP (Fig 16). This region is rich in positively charged lysine (K) and arginine (R) residues which are responsible for nucleotide binding (Kauffman and Karypis, 2008). Hence, it can be hypothesised that the N terminus is crucial in binding to the DNA for assembly of the virus which needs experimental validation. In the present study, higher concentration of TEV protease also did not aid in resolving the issue. Addition of extra residues between TEV protease cleavage site and N terminus of BBTV CP might make the site available for proteolysis.

Purified pET/BBTV CP was dialysed and subjected to cleavage with thrombin. The cleavage was only partial at 4 °C. Cleavage at 25 °C was successful, with higher incubation period. Nevertheless, after cleavage the 20 kDa band was not visible in the gel. To ascertain that the low molecular weight protein bands have not coalesced (17 kDa of Trx-S-His tag and 20 kDa BBTV CP) to seem like a single band on SDS-PAGE, Tricine SDS-PAGE was

performed. On running the cleaved samples on Tricine SDS-PAGE also the bands could not be distinguished. This can be corroborated to the low stability of BBTV CP at higher temperatures. Hence the fusion protein was used for immunisation of animal.

The BBrMV CP cloned into pRSET-C and pGEX-4T-2 were overexpressed and purified in *E. coli* Rosetta strain. A doublet band at 34 kDa was observed on induction of *E. coli* harbouring pRSET/BBrMV CP (Plate 43). In a previous investigation conducted by Dhanya (2004), on comparing the gel profile of protein extracted from BBrMV infected and healthy banana, three new bands were observed in the former. Out of which, doublet band at approximately 32 kDa, was observed corresponding to the CP gene similar to the present study. The difference in size in the recombinant protein may be due to the addition of few amino acids during cloning. Formerly, CP of other potyvirids have been observed to undergo partial degradation resulting in doublet bands in SDS-PAGE (Hiebert and McDonald, 1973; Huttinga and Mosch, 1974; Moghal and Francki, 1976). Interestingly, TuMV CP was seen as triplet in SDS-PAGE profile and a single band was observed in the case of SCMV CP (Choi and Wakimoto, 1979; Gough and Shukla, 1981; Shukla et al., 1987). The explanation for this phenomenon was put forth by various authors. The surface exposed N and C terminal of potyvirus CP is prone to proteolytic cleavage during the process of purification or while storage by proteases of microbial origin depending upon the host in which they are purified (Moghal and Francki, 1976; Hiebert et al., 1984; Shukla et al., 1988). However, the 60 kDa pGEX/rBBrMV CP fusion protein appeared as single band in the gel profile (Plate 46). Nevertheless, post cleavage of the tag, doublet band was observed both in the gel and Western blot.

Even though all the parameters determining solubility and yield of pRSET/rBBrMV CP were optimised, the protein remained insoluble. Due to non-availability of soluble protein to bind to the Ni^{2+}-NTA beads, many non-specific *E. coli* proteins were eluted with the target protein (Plate 48). The GST-BBrMV CP fusion protein (pGEX/rBBrMV CP) was highly soluble (Plate 53). However, the yield was less. The protein on storage even in the presence of glycerol was seen to be subjected to proteolysis. Multiple bands of the protein were observed in the SDS gel profile. Confirmation that these bands are indeed that of the fusion protein was done by Western blotting using anti-GST antibody.

As discussed earlier, the surface exposed N and C terminal of potyvirus CP is prone to proteolytic cleavage during purification or storage by proteases of host origin (Moghal and Francki, 1976; Hiebert et al., 1984; Shukla et al., 1988). This cleavage was observed even though protease inhibitor cocktail was added to the sonication buffer for purification. The GST tag also contributes to degradation due to unknown reasons. The multiple bands of the purified rBBrMV CP could also be due to this proteolytic cleavage. Proteolysis was observed in the cleaved as well as uncleaved protein (Plate 54).

Previously, BBrMV CP cloned in pProEX-1 was transformed into *E. coli* M15 cells containing the pREP4 repressor plasmid for overexpression. Fusion protein was purified using QIA expressionist. It was reported that low levels of 39 kDa protein was obtained on induction which was absent in the uninduced fraction (Rodoni et al., 1997). Recently, urea denaturation of insoluble BBrMV CP was carried out by Selvarajan et al. (2020b). The cells after induction were harvested and sonicated. The insoluble protein from the pellet was released by resuspending in lysis buffer (100 mM NaH_2PO_4, 10 mM Tris HCl, pH 8.0) with 8 M urea and incubated for 1 h. After another round of centrifugation, the expressed 34 kDa CP was purified using Ni^{2+}- NTA column chromatography. The purified protein was quantified as 1.2 mg/ml from 1 L culture. Similar experiment was conducted in the present study. Major difference in the methodology was that 8 M urea was added in the lysis buffer for sonication and the same was used for washing and elution. However, contradicting results were obtained. Many non-specific bands that of *E. coli* protein were obtained on purification, hence this method was not pursued further.

Hitherto, rCP of many plant viruses have been overexpressed and purified to be used for developing diagnostics. The SCSMV CP cloned in to pRSET- C vector was highly soluble (Hema et al., 2003). The solubility of PVBV CP cloned to pRSET-C, also a potyvirus, was also high when expressed in *E. coli* BL21(DE3) pLysS (Sabharwal, 2017). Similarly, CP of CMV infecting cucumber, was cloned to pRSET- B and pET21a to obtain soluble protein (Pandey, 2015; Kim et al., 2016). Gulati et al. (2016a) had expressed Tobacco streak mosaic virus (TSMV) CP in pRSET-C vector. From the experiments performed during the present study and the evaluating previous works it is evident that each protein behave differently and that the overexpression levels and yield is highly dependent on the nature of the protein as well as the vector and host used for the same.

5.12 PROTEIN FOLDING

The fold index analysis of pET/rBBTV CP, ΔpGEX/rBBTV CP with Trx and GST tags respectively, indicated that the protein was partially disordered. The N terminal of the protein without tag was apparently disordered as well. The untagged rBBrMV CP as expected was partially disordered with ordered core regions. This was substantiated by fluorescence spectroscopy results.

5.13 SECONDARY STRUCTURE PREDICTION AND HOMOLOGY MODELLING

The protein fold, secondary and tertiary structure at certain conserved motifs are maintained in the population which reflects virus adaptability. Usually, the protein structure is studied by X-ray crystallography, NMR spectroscopy or cryo-electron microscopy. However, soluble, pure protein in large quantities is required for any of these studies. Through the present study, success have been attained in producing high quality and quantity of soluble CPs of BBTV and BBrMV *in vitro*. This can pave way for future researches to solve the structure of the virus and elucidate the multifunctionality based on the structure.

Proteins are composed of amino acids, linked to each other by peptide bonds. Within each protein these amino acids interact with each other to form secondary structures known as α-helices and β-sheets. Tertiary structures are result of the three-dimensional folding of α-helices and β-sheets, which produces a variety of interactions to form the globular protein. These interactions cause the protein to attain its final, quaternary, structure and this 3D structure is of immense biological importance. Protein folding refers to folding of the protein to form a unique shape. The way in which each protein folds is critical to its structure and function.

The secondary struture of protein can be highly corroborated to its function and hence it was predicted using PSIPRED workbench. Secondary structure of BBTV isolates were predicted and all the seventeen isolates depicted similar secondary struture (Fig 18). For homology modelling of BBTV CP, *Ageratum yellow vein virus* CP was used as template. However, structure of none of the nanoviruses have been solved to use as a better template. There were many gaps in the alignment of the amino acid sequences between the query and template.

Recently, there has been an attempt to solve the structure of BBTV CP. However, stable soluble rCP could not be produced in the *E. coli* system (Sairam et al., 2020). BBTV CP cloned into pET28a produced protein in the insoluble fraction and thus the authors attempted *ab initio* structure prediction and homology modelling using iTASSER, Robetta, Quark, Schrodinger, SwissModeler, RaptorX, Phyre and MMM. The structure quality was assessed using QMean server and SAVES. According to the best model among the lot, the fold was predicted to be analogous to the β-sandwich fold of the CPs of other plant viruses consisting of two β -sheets packed against each other. Two short α helices between β sheets were predicted. Moreover, the aphid binding motif EAG was predicted to be located at the highly accessible loop between βF and βG. According to the authors the N terminal consisted of basic amino acid residues responsible for nucleic acid binding. The same inference have been drawn from various of the previous experiments in the present study.

The secondary structure of all the BBrMV Kerala isolates displayed a similar pattern (Fig 42). The disordered N and C terminals were composed of coiled-coil domain. Four major helical domains were also identified. Geourjon and Deléage (1995) predicted the secondary structure of PVBV CP and inferred that the 20 and 23 residues in the N and C terminal respectively do not possess any secondary structure just as observed in the present study. Despite the amino acid sequence at the N terminal being highly variable among the species of *Potyviridae* family, the structure is apparently highly conserved.

Threading of BBrMV CP was done using i-TASSER suite by comparing with CP of WMV and was in agreement with the predicted secondary structure of the protein (Fig 43). A noteworthy observation was, even if the amino acid variations were observed between the most diverse and most homologous isolates, it did not change the secondary structure of the CP. Thus, the function must be conserved.

The predicted structure and nucleotide binding of BBrMV CP was in tandem with the molecular structures of WMV, TuMV and PVY solved by cryo -electron microscopy (Zamora et al., 2017; Cuesta et al., 2019; Kežar et al., 2019).

The CP of plant viruses is multifunctional. They take part in assembly of the genomic DNA/ RNA and form infectious particles. Potyviral CP facilitates not only infection but also systemic movement and transmissibility. The highly conserved DAG motif (Asp-Ala-Gly),

in the CP of BBrMV like any other potyvirus acts together with HC-Pro and arbitrates aphid transmission. The DAG motif together with adjacent amino acids of N terminal is believed to aid in aphid transmissibility of the virus (Nigam et al., 2019). Except MT818183 and MT818184 all isolates contain two DAG motifs, one at proximal (16) and one at distal (67) position of N terminal. However, deviations from DAG were also observed in two isolates (MT818187 and MT818177). The position of motif was conserved throughout the dataset. Although, DAG motif is believed to be highly conserved throughout *Potyviridae* family it is not universal as it varies in position and sequence (Nigam et al., 2019). Despite this, potyviruses without a DAG motif is also transmitted by aphids (Dombrovsky et al., 2005). This is attributed to the structural complementarity of the partially disordered N terminal of the CP with hinge domain of HC-Pro (Kehoe et al., 2014).

The extended N and C terminal of BBrMV contributes to its flexibility as suggested in case of other flexuous plant viruses (Agirrezabala et al., 2015; DiMaio et al., 2015; Zamora et al., 2017). The highly proteolytic flexible surface exposed N terminal is believed to be significant in viral transmission. It has been proposed by Harrison and Robinson (1988) that proteolysis occurring in the region between highly conserved DAG motif in the N terminal and core region will help release the infectious potyviral particles from vector stylet during inoculation. Unlike the DNA binding site at the N terminal of BBTV CP, the predicted RNA binding site in BBrMV CP was in the “armpit-shaped groove” located between the core and the C-terminal region of CP as observed in case of PVY (Kežar et al., 2019).

On superimposing the 3-D structure of BBrMV CP of MT818185 (most homologous isolate) and MT818187 (most variable isolate) it was clear that the N terminal differed whereas the core region remained the same. The RNA binding sites were predicted to be located in the armpit grove of the CP in both isolates, were the amino acid residues D, H, K, I (aspartic acid, histidine, lysine and isoleucine) were highly conserved, however, the positions varied. Thus, it is evident that the multifunctionality of BBrMV CP is conserved irrespective of the sequence variation through its molecular structure.

Ramachandran plot, ERRAT plot and ProSA energy profile clearly explained the goodness of the model. However, while analysing the quality of the models using ProSA, z score of

BBTV CP was -4.34. The protein folding energy plot was drawn with energies as a function of amino acid sequence position. The knowledge-based energy of BBTV CP was also high for the predicted model (Fig 20). In general, positive values correspond to problematic or erroneous parts of the input structure (Wiederstein and Sippl 2007). This may be due to some misinterpretations in the modelling based on the template used. It was observed that the calculated z score of BBrMV isolate, MT818187 was higher than that of MT818187. The former was close to zero indicating a better model. Thus, it can be inferred that the mutations have some advantage and can be fixed in the population. From the graph (Fig 55), it can be inferred that the z scores of BBTV CP as well as BBrMV CP structures were within the range of scores typically found for native proteins of similar size.

The overall quality factor of the homology models of MT818187 was higher than MT818185 (82.64 and 72.22 respectively) and that of BBTV CP was 75.92. Nevertheless, in both the models, most of the amino acids except N terminal and C terminal residues were below 95 per cent cut off value in ERRAT plot (Fig 56). In BBTV residues approximately between 25-35, 45-50, 95-100, 145 and 165 were having a very high error value. All these amino acid residues formed loop structures on the model.

In SuSPect analysis, the 20 possible amino acid types were labelled along the x-axis with their one-letter code. The coloured bars indicated the probability that a mutation to the corresponding residue will affect the function of the protein or on the phenotype of the organism. The residue with least bar height indicated the wild type. From the graph, (Fig 57) residues *viz.,* tyrosine (105), arginine residues at positions 225 and 234 were the residues mostly determining the function. From molecular model, the armpit grove near the C terminal binds to the RNA and the basic amino acid like arginine are responsible for binding and initiation of virion assembly. The residues and residue-pairs with charged and aromatic side chains are important for protein-protein interactions. Tyrosine and phenylalanine residues which were detected in the program to be important for the function of the protein might be involved in this kind of interactions.

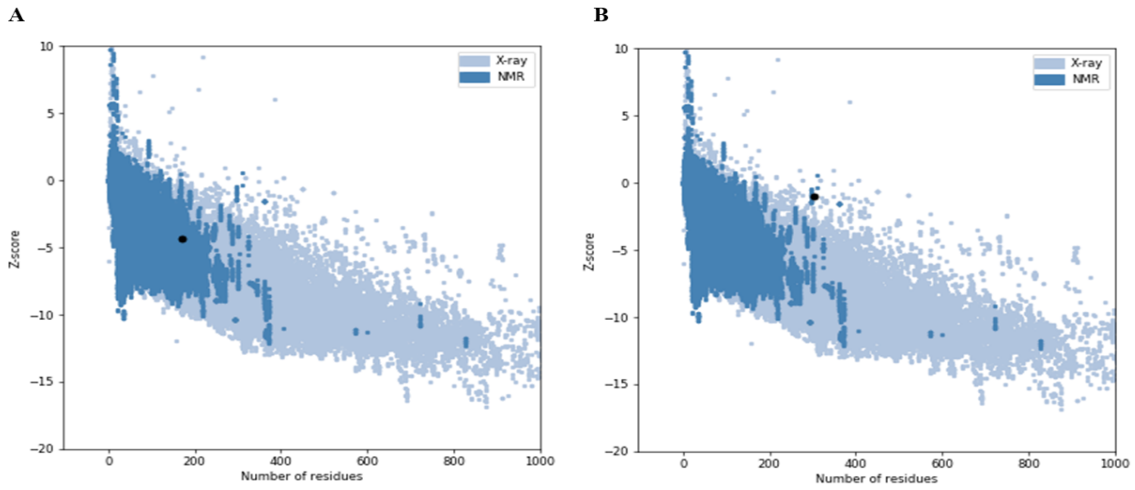


Fig 55. Assessment of overall quality of homology model by ProSA A BBTV CP **B** BBrMV CP (MT818187). In this plot, groups of structures from different sources (X-ray, NMR) are distinguished by different colours (light and dark blue respectively). The z score of input structure is denoted in red colour

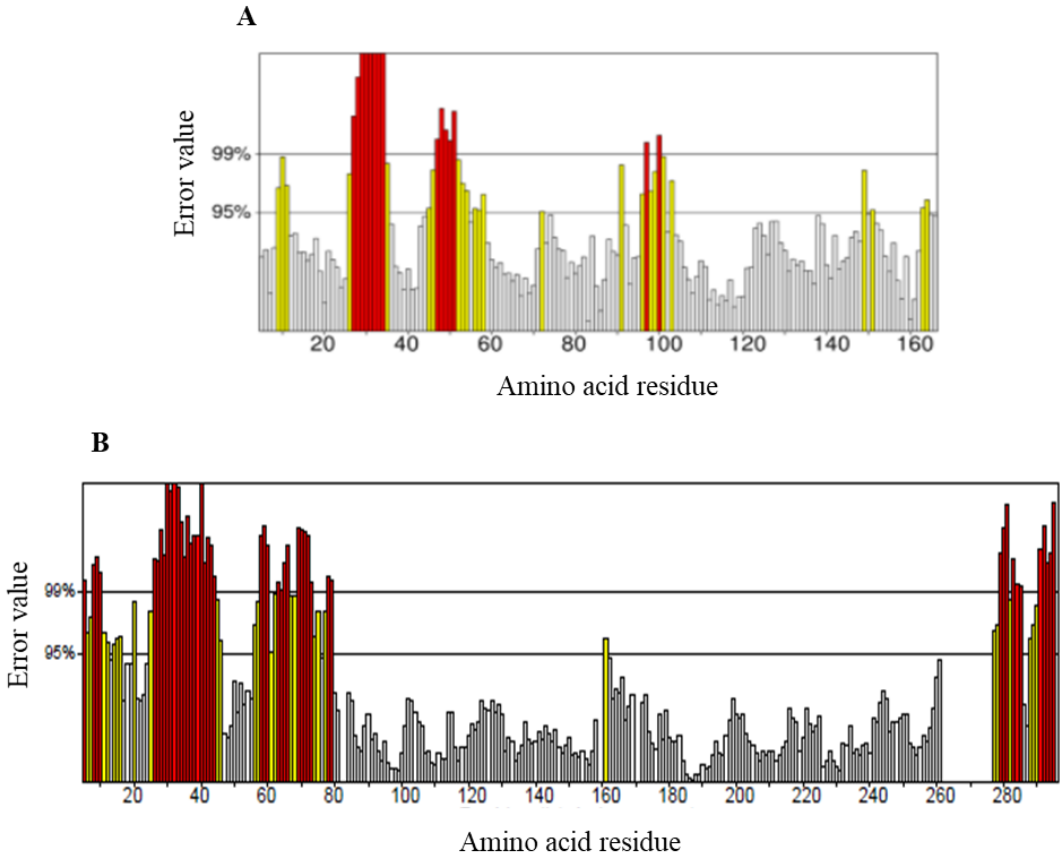


Fig 56. ERRAT plot for **A** BBTV CP and **B** BBrMV CP. The error value higher that the 95 % threshold is indicated as yellow and that above 99 % as red bars.

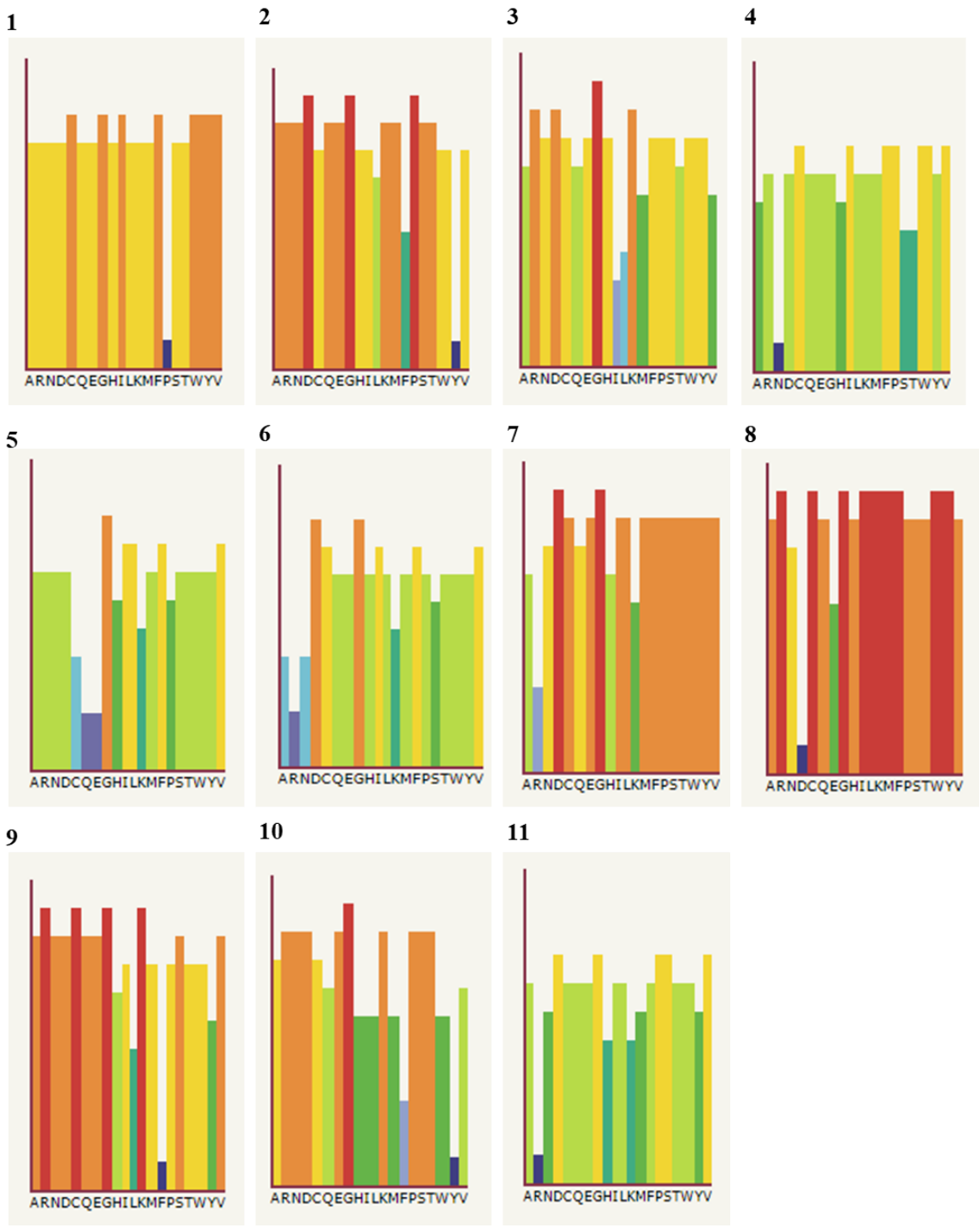


Fig 57. Mutational analysis graph generated in SuSPect. 1-11. Proline- 89 position, tyrosine-105, leucine-112, asparagine- 114, glycine-121, arginine- 190 and 225, asparagine-234, phenylalanine- 235, tyrosine- 236 and arginine- 281

5.14 BIOPHYSICAL CHARACTERIZATION OF PROTEIN

5.14.1 Western blot analysis

The antigenicity of the rCP of both BBTV and BBrMV was assessed by Western blotting (Plate 19B and 54B). It was evident that the rCP was highly antigenic. For identification of the fusion protein however, antibody targeting the affinity tag was also used. Further, the epitopes were predicted using Bepipred linear epitope prediction and DiscoTope 2.0. From the predictions it was unambiguous that the N terminal of both the CP consisted the major virus specific epitopes (Fig 21A and 47B).

It was recognized in the case of potyviruses that the surface exposed N terminal accounts for most of the variation at the same time, the specific epitopes of the virus are present in this region. The core region however, consists of amino acid sequences that are conserved throughout the family and these epitopes produces antibodies that can detect all the viruses in *Potyviridae* family (Shukla et al., 1988; Jagadish et al., 1991; McKern et al., 1991). Therefore, PVBV CP antiserum detected BBrMV CP in Western blot and did not bind non-specifically to any other *E. coli* protein. It has been advocated and experimentally proven that the potyvirids can be identified and classified based on the CP amino acid sequences (Shukla et al., 1988).

5.14.2 Assembly of virus like particles *in vitro*

The virus like particles (VLPs) are the assembly of capsomeres to capsids without the association of the genomic RNA making it non-infectious. The VLP's can assemble *in vitro* either in the presence or absence of RNA (McDonald and Bancroft, 1977). Formation of VLPs by assembly of BBTV CP subunits were assessed by sucrose density gradient centrifugation. In the gel profile of pellet after first ultracentrifugation of BBTV-Trx fusion protein (pET/rBBTV CP), a prominent band at 37 kDa was observed indicating that the monomeric capsid proteins were forming oligomers or even polymers to form VLPs. After sucrose gradient ultracentrifugation, the 37 kDa band was observed between 20-30 per cent sucrose gradients and were pooled. However, due to presence of many *E. coli* proteins the purified preparation was heterogenous (Plate 28B). The electron micrograph depicted isometric particles of size close to the expected size of BBTV virions (Plate 29). The expected size of VLP was 20 nm and the size of isometric particles ranged between 21-23

nm may be due to the presence of 17 kDa tag. However, BBTV-GST fusion protein did not form VLPs and was found in the 10 per cent fraction of the sucrose gradient.

Previous attempts to purify the virions from infected banana was conducted by Thomas and Dietzgen in 1991. They encountered similar problem of co-sedimentation of various plant protein in sucrose density gradient. The authors opinionated that it was essential to use Cs_2SO_4 gradients to separate BBTV particles from these impurities. It was also pointed out that under certain storage conditions like presence of in 95 per cent ethanol or buffers with pH < 8.5 made the purified BBTV particles unstable. They also reported freezing at -20 °C without protectants lead to similar consequences. The BBTV VLPs after final ultracentrifugation was stored in Tris buffer of pH 8.0 at -20 °C before electron microscopy. This might have led to the disassembly of the protein and in identification of less VLPs in electron micrograph.

A total of 296 octapeptide corresponding to amino acid sequences of the entire *Johnson grass mosaic potyvirus* (JGMV) CP was screened for immunological properties and N-terminal was found to be most immunodominant when intact virions are used as an immunogen (Shukla et al., 1989). Undoubtedly, VLPs formed *in vitro* by self-assembly of the CP subunits would be best candidate for raising antiserum than monomeric CP as it is very hard to obtain the intact N and C terminal of BBrMV CP in the purified protein preparations. Hence *in vitro* self-assembly to VLPs were assessed. In the electron micrograph, ring like structures were observed. Jagadish et al. (1993) stated that during the assembly of virions, potyviral capsomeres assembles as stacked rings after binding to the RNA at the origin of assembly. This then grows bidirectionally towards 5’ and 3’ end to pack the entire ssRNA. In 2003, Anindya and Savithri succeeded in establishing the role of the surface exposed N and C terminal of PVBV CP subunits for oligomerisation and assembly of PVBV CP *in vitro*. According to them, octameric rings are formed initially by the head to tail interactions of the N and C terminals which is later assembled to helical aggregates. The study also substantiated that truncated capsomeres cannot form VLPs. Proteolysis of rBBrMV CP post purification was suspected, as multiple bands were observed in the SDS-PAGE gel profile. The failure of purified CP to assemble to complete VLPs and only forming the intermediate disk- like structures may be attributed to partial proteolysis of N and C terminal residues of few subunits.

Hitherto, filamentous VLPs of heterogenous length of rJGMV CP overexpressed in *E. coli*, yeast, insect, and mammalian systems have been reported (Jagadish et al., 1991; Hammond et al., 1998). Similarly pRSET/rPVBV CP could self-assemble to form VLPs in sucrose density gradient centrifugation (Joseph and Savithri, 1999; Sabharwal et al., 2019). On the contrary, pRSET/rBBrMV CP overexpressed in various hosts in this study was highly insoluble hence could not be analysed for self-assembly.

Although, free N and C terminal is prerequisite for self-assembly of capsid, presence of few extra amino acids in the N terminal either due to cloning strategy or in a chimeric VLP did not hamper the assembly (Sabharwal et al., 2020). Curiously, both BBTV and BBrMV fusion proteins with GST tag did not assemble into VLPs possibly because the tag is a well folded bulky peptide of 25 kDa.

5.14.3 Fluorescence spectroscopy

Fluorescence spectroscopy is a very important tool to help understand the tertiary structure of the protein which is not or cannot be characterised by crystallography. Previously, CP of few viruses have been characterised by florescence spectroscopy. The λmax of 30 kDa rBCMV CP, was 345 nm similar to the spectrum of rBBrMV CP, indicating slight disorder in the protein (Kumar et al., 2019). It is evident from previous studies that N and C terminus of CP of potyvirids are surface exposed and disordered facilitating in oligomerisation and assembly (Anindya and Savithri, 2003). Similar spectra were exhibited by *Potato virus A* with two maxima at 314 nm and ~330 nm (Ksenofontov et al., 2013).

On comparing the fluorescence spectra of rBBrMV CP and pET/rBBTV CP, the latter seemed to be depicting high intensity emission spectra of a globular protein. Fluorescence spectra of ΔpGEX/rBBTV CP fusion protein measured between 300-400 nm displayed two peaks one at 325 nm and 345 nm (Fig. 22). However, the intensity was less compared to that of pET/rBBTV CP fusion protein (Fig 23). Among BBrMV CP-GST fusion protein and untagged BBrMV CP, the latter displayed higher intensity than the former (Fig 49). The spectra of BBrMV untagged protein pointed out that most of the protein is monomeric. The electron micrograph also substantiated this result.

Globular proteins usually depicts fluorescence emission maximum (λmax) at 330 nm in native condition (Kumar et al., 2019) as seen in case of Alfalfa mosaic virus (AMV) CP (Kan et al., 1986). Therefore, the spectra of BBTV fusion protein and untagged BBrMV

CP indicated that protein is partially folded with some disordered region. This corroborates the FoldIndex analysis wherein except N terminus the protein is predicted to be ordered. These results also reinstate that the reason for absence of cleavage of the GST tag from BBTV-GST fusion protein at TEV protease site was due to the steric occlusion created by this folded structure.

5.14.4 Mass spectroscopy

The pET/rBBTV CP monomer was digested in-gel and subjected to MALDI-TOF analysis. The peptides identified were analysed in MASCOT server and the results were interpreted. Maximum score was obtained with CaMV virion-associated protein (VAP) and capsid assembly scaffolding protein of Bacillus phage phi29. The misinterpretation of result may be due to the fact that there are around 11 continuous trypsin cleavage sites in 25 amino acid residues in the N terminal of BBTV CP giving rise to short peptide of 2 amino acid length or even single amino acids which might not have been detected in the analysis (Fig 17). On analysing amino acid sequence homology of these protein with the rBBTV CP, 21-24.68 per cent identity was observed.

Previously, identification of unknown proteins associated with PVBV CP while internalisation into HeLa and HepG2 cells were carried out by in-gel digestion followed by LC MS/MS analysis. The proteins were identified as vimentin protein and heat shock proteins from *Homo sapiens* respectively (Sabharwal, 2017).

5.15 ANTISERUM PRODUCTION

Quality antiserum is prerequisite for any serology based detection methods. To produce quality antiserum the antigen used for immunisation is also equally important (Hema et al., 2003). Partial purification of BBrMV was attempted by Thomas et al., (1997) for antiserum production. The infected tissue was powdered in the presence of liquid nitrogen homogenised in extraction buffer (EB) containing 0.5 M borate, pH 6.8 and 0.2 per cent v/v β-mercaptoethanol. After removing the debris, the filtrate was clarified by centrifugation and treated with Triton X-100. The mixture was the centrifuged at ultra high speed on 15 ml cushion of 30 per cent (w/v) sucrose in EB. Pellets resuspended in 0.25 M borate buffer, pH 6.8 and further purified by equilibrium centrifugation in caesium chloride; $CsCl_2$ (0.44 g/ml) at 10°C for 16 h at 32,000 rpm. The fractions were analysed and the ones with the virions were pooled and ultra pelleted. This was again resuspended

in borate buffer and used for immunisation by series of three intramuscular injections into rabbit. Antiserum titre of 1:10000 was obtained after the experiment.

Yet another method for purification of VLPs of BBTV from BBTD infected banana plants were reported by Iskra-Caruana and co-workers (1989) and Wu and Su (1990a). Monoclonal antibodies (MAbs) raised against Taiwanese isolate purified from infected sample by Su and Wu were used to develop ELISA and immunofluorescence assays to detect BBTV and provide for the first time a diagnostic tool for the disease in 1990. A variant of ELISA, plate-trapped antigen (PTA) -ELISA was also developed using the MAbs produced against the viral CP.

Thomas and Dietzgen (1991) have previously proposed another method for virion purification from infected samples. This method was utilised for purifying BBTV virions from infected plants with the intention of using it as antigen for raising antiserum (Estelitta, 1998; Selvarajan et al., 2010a). The average yield of BBTV was reported to be 1.62 mg/kg of plant tissue from artificially inoculated tissue culture samples and 1.10 mg/kg tissue when purified from infected field sample (Estelitta, 1998). Partially purified virus was emulsified with an equal volume of Freund's incomplete adjuvant and three intramuscular injections were administered into a ten-week-old New Zealand white rabbit. The titre of the antiserum was not specified however, the detection limit of DAC-ELISA with the antiserum was up to 50 times diluted antigen preparation.

Partial purification of virus is cumbersome and demand expertise and expensive equipments. An argument made by many researchers is that purification of the virus from banana for antibody production is unsuitable because, banana contains high levels of polyphenols and polysaccharides which interferes with the process of purification (Selvarajan et al., 2016; Selvarajan et al., 2015; Yasmin et al., 2001). Low concentration of the virus and inconsistency in the yield are other problems encountered. This notably challenges the specificity of the antisera as well as considerably reduced sensitivity of the assay. Development of rCP and its use in immunising the animal, greatly solves the above problem.

During the present study, pET/rBBTV CP Trx fusion protein was administered into rabbit to produce polyclonal antiserum that could detect purified antigen at 1:1000000 dilution of the antiserum in DAC-ELISA. The sensitivity of the antiserum was also determined by DIBA. The antiserum dilution of commercially available antiserum ranges between 1:500

to 1:1000. The dilution of the antiserum produced in the present study was 100 times more than that of the commercially available antiserum. This would indicate 100 more tests can be carried out using 1 ml of the sera.

However, considering that the field samples will have different inoculum load at different parts, age and stage of infection, the titre was increased to 1:10000 so that it can detect even trace amount of antigen present in the sample. The titre value was validated using field and tissue culture samples collected from Banana Research Station, Kannara. The antiserum produced against the pET/rBBTV CP successfully detected the latent infection in the asymptomatic tissue culture plantlets.

The titre value of antiserum previously produced against purified virus preparations were reported to be 1:128 in double diffusion test and 1:512 in ring interface precipitin test (Estelitta, 1998). Antiserum produced against BBTV-MBP fusion protein had a titre of 1:1000 and was proven to have high affinity to the 20 kDa protein. However, the antiserum also bound to certain plant proteins present in both healthy and infected sample. MBP was experimentally proven to be immunogenic contributing to few epitopes against which antibodies might have been produced. The probable explanation put forth was MBP tag in the fusion protein might be analogous to some plant protein thus giving rise to non-specific bands in the Western blot (Wanitchakorn et al., 1997). However, in antiserum produced against CTV-MBP fusion protein, and SCMV MBP-fusion protein such cross reactions were not observed (Nikolaeva et al., 1995; Smith et al., 1995). The titre value of both the antisera were around 10^{-5}.

In another instance BBTV CP cloned into pQE30 vector and purified was used to raise polyclonal antisera that could detect BBTV and FBNYV in infected tissues by dot blot - immunoassays at 1:1000 dilution (Abdelkader et al., 2004). However, positive reaction in sap dilution up to only 1:15 was distinguishable from the negative sample.

5.16 SEROLOGICAL ASSAY FOR DETECTION OF BBTV

There are several nucleic-acid based detection method for BBTV from infected samples. Apart from PCR, Loop mediated Isothermal Amplification (LAMPS) and Rolling circle amplification (RCA), nucleic acid spot hybridization (NASH) using DNA probes have been applied for the sensitive detection of BBTV since the last couple of years (Harding et al., 1991; Xie and Hu 1995; Hafner et al., 1997; Selvarajan and Balasubramanian 2008). However, serological detection methods are much more amicable and comparatively

inexpensive. Also, molecular detection, does not take to account the nucleotide or amino acid variation happening in the gene from time to time. Antigen epitope constitutes many amino acids that are highly conserved. Antigen antibody interactions are greatly dependent on the structure and hence serological detection is consistent, sensitive and can be used for a longer period.

In the present study, serological assays like DAC-ELISA, TAS-ELISA and DIBA were standardised in the laboratory with the antisera produced against rCP and the performance was compared with the commercially available antiserum. The sensitivity of the three serological assays is depicted in the table (Table 18). From the table it is evident that the sensitivity of TAS-ELISA is higher than DAC-ELISA or DIBA. However, with high quality antiserum up to 1:40 dilutions of antigen could be detected using DAC-ELISA and DIBA which previously could only detect up to 1:20 and 1:10 respectively.

Table 22. Comparison of the sensitivity limits of DAC-ELISA, TAS-ELISA and DIBA for the detection of BBTV

Test	Healthy	Undiluted	Dilutions			
			1:10	1:20	1:40	1:80
DAC-ELISA	-	+++	++	++	+	-
TAS-ELISA	-	+++	++	++	++	+
DIBA	-	+++	++	++	++	-

Serial dilutions of antigen were prepared by mixing BBTV infected sap with buffer

-Negative; + Positive; ++/ +++ strong positive reaction.

Combination of both nucleic- acid and serology based assay, IC-PCR was also standardised in the laboratory (Plate 31). However, high concentration of antibody had to be used for capturing the antigen. Antigen diluted up to 1:10 only amplified to give 850 bp of product using the newly designed primers (BBTV DNA-S FP and RP). The assay has to be further perfected in order to attain maximum sensitivity.

To carry out large scale screening for episomal BSV in banana, IC-PCR was developed (Harper et al., 1999). Twenty five µl of purified IgG (2 ng/ml) in carbonate buffer, was used to coat the PCR tubes and incubated at 37 °C for 2 h. Antigen (25 µl) extracted in 1:10 w/v PBS was loaded and subjected to PCR amplification. Only the virion captured by the antibody releases the genome and act as template for amplification.

A combination of Immuno capture and reverse transcriptase PCR have been reported to be highly sensitive in detecting many RNA viruses like SCSMV (Hema et al., 2003), *Grapevine leafroll-associated virus* 3 (GLAV-3) (Kumar et al., 2015b), potyviruses infecting yam (Mumford and Seal, 1997), *Plum pox virus* (PPV) (Wetzel et al., 1992), *Bean yellow mosaic virus* (BYMV), *Cherry leafroll virus* (CLRV), CMV, CTV, *Grapevine fanleaf virus* (GFLV), *Potato leafroll virus* (PLRV), *Pepper mild mottle virus* (PMMoV) and *Tomato spotted wilt virus* (TSWV) (Nolasco et al., 1993). Also, multiplex IC-RT-PCR was developed to detect three banana viruses simultaneously (Sharman et al., 2000).

From the present study, it was evident that the BBTV and BBrMV is fast evolving. The BBTV Kerala isolates belonging to PIO group are less diverse compared to BBrMV, however, chances of recombination in this multipartite virus cannot be ignored. Thus, it is essential to evaluate nucleic-acid based detection methods periodically.

The amino acid sequences of BBTV CP and that in BBrMV CP core region is highly conserved and is a very good target for serodiagnostics. However, partial purification of the virus from banana is tedious and each batch will be different. Recombinant CP overexpressed in suitable bacterial host, yield highly pure protein. Recalling the results of the experiments conducted, BBTV CP gene cloned into pET32a (+) and modified pGEX-4T-2 vector (ΔpGEX/BBTV CP), yielded high concentration of pure protein. Whereas expression in BL21 pLysS, Rosetta pLysS and C41 strains harbouring pRSET/BBTV CP and pGEX/BBTV CP were unsuccessful.

Nevertheless, rBBrMV CP overexpressed in both BL21 pLysS and Rosetta pLysS transformed with pRSET/BBrMV CP recombinant clone were insoluble whereas, that overexpressed in the *E. coli* strains harbouring pGEX/BBrMV CP was highly soluble. These experiments pointed out that the quality and quantity of the rCP is greatly determined by the vector and host used for production of the recombinant protein apart from the properties of the protein itself.

Both the rCPs were highly antigenic and from the predictions, fusion proteins consisting of complex epitopes, were expected to be more immunogenic. This was proven right in the case of pET/rBBTV CP with 17 kDa Trx-S-His tag at its N terminal. The antiserum thus raised against the fusion protein was validated for its specificity and sensitivity in both field and TC samples and standardisation of various serodiagnostic assays were carried out in the laboratory.

Summary

6. SUMMARY

The present study entitled “Development of recombinant coat protein for immunodiagnosis of banana bunchy top and bract mosaic diseases” aimed at developing recombinant coat protein as a potential antigen for immunodiagnostic assays for detection of BBTV and BBrMV isolates in Kerala considering their evolution. The study was conducted during the period of 2016-2021 at Banana Research Station, Kannara, College of Agriculture, Vellanikkara, and Indian Institute of Science Bengaluru.

Virus samples were collected from twenty representative blocks in ten districts. The districts surveyed were Thiruvananthapuram, Kollam, Kottayam, Ernakulam, Thrissur, Palakkad, Kozhikode, Wayanad, Kannur and Kasaragod. These locations were sub divided into Northern, Central and Southern zones. A total number of 30 and 35 apparently BBTV and BBrMV infected samples respectively were collected. Preliminary assay was conducted by DAC-ELISA in the laboratory. Positive samples from ELISA, were selected for further studies

Seventeen and twelve representative samples from BBTV and BBrMV infected samples collected from all over Kerala that gave positive reaction in DAC-ELISA were chosen for total DNA/RNA isolation. The quality of the DNA and RNA was checked in agarose gel and quantified by nanodrop.

The coat protein gene of BBTV was amplified from the total DNA isolated from BBTV infected sample using reported primer pairs to obtain an amplicon of 1058 bp. The complete coding region of BBTV was obtained from the amplification. The total RNA isolated from BBrMV infected samples were converted into cDNA using Revert Aid First strand cDNA synthesis kit. The cDNA was comparatively stable and stored at -20 °C until it was used as template for PCR to amplify CP gene of BBrMV. Reported primers were used to amplify the gene to obtain an amplicon of 1062 bp with the complete coding region of the virus along the 3’ UTR.

The BBTV and BBrMV CP gene of representative samples were sequenced in both directions using gene specific forward and reverse primers. The sequences observed were processed and contigs with full length were obtained. The sequences were blasted against the NCBI database to ascertain the identity of the gene. These sequences were submitted in

GenBank and accession numbers were obtained. BBTV isolates were designated as MT174314-MT174330 and BBrMV isolates as MT818176-MT818187.

Phylogeographic and diversity analysis of BBTV and BBrMV were carried out based on its coat protein gene. Apart from the Kerala isolates generated in the present study 65 and 56 sequences of BBTV CP and BBrMV CP respectively were retrieved from NCBI database for meaningful comparisons. In BBrMV CP dataset, isolates infecting cardamom and flowering ginger were also included apart from that infecting *Musa* sp.

Phylogenetic tree based on nucleotide and amino acids were constructed after aligning the sequences using MUSCLE algorithm. Clustering of BBTV isolates were greatly influenced by the geography. The major clusters included those isolates collected from India, Sri Lanka, Myanmar, Egypt, Tonga and Fiji forming Pacific Indian Oceans (PIO) group and those collected from China, Taiwan, Indonesia, Philippines, Japan and Vietnam forming South East Asian (SEA) group. From the tree constructed with only Kerala isolate an inclination to geographic demarcation was observed. The isolates collected from Northern and Southern zones were distinct. However, central zone isolates seem to be clustering alone or with either northern or southern zone isolates. Principle co-ordinate analysis substantiated that the isolates were grouped based on geography. However, two SEA isolates one each from China and Thailand were apparently very distinct from the larger group.

The phylogeographic analysis of BBrMV CP revealed that all the isolates were derivatives of a common ancestor. Due to this, the isolates clustered in no particular order. It was noted that isolates from Philippines closely related to that from India, indicative of the spread of the virus from point or origin to various parts of the world through planting materials years ago. Kerala isolates clustered together with other South Indian isolates and closely clustered with isolates infecting cardamom. Out of all the isolates, two Kerala isolates collected from Ernakulam and Kasaragod respectively, MT818177 and MT818187 had the maximum branch lengths indicating that they have evolved recently from the common ancestor. The principal co-ordinate analysis also differentiated these two isolates as they did not cluster together or with other BBrMV isolates considered in this study.

Parameters like nucleotide and amino acid sequence homology and diversity, gene flow between the subpopulations and gene differentiation, nucleotide diversity and mutation rate in the coding region of the gene, ka/ks ratio, codon usage bias were calculated and analysed

for both BBTV and BBrMV CP as measure of genetic diversity. The nucleotide diversity among the BBTV Kerala isolates ranged from 99.2-100 per cent and that in BBrMV isolates ranged between 88.2-100 %. The amino acid diversity of BBTV and BBrMV were 98-100 and 65-100 per cent respectively. According to the analysis carried out, it was evident that in both the virus populations, gene flow was frequent between the northern, central and southern subsets. The fixation index was statistically significant indicative of this fact.

The nucleotide diversity and mutation rate were calculated for both CP genes. Interestingly, the N and C terminal of the BBrMV CP was highly mutated and diverse compared to the conserved core region. However, no such distinction was observed on BBTV CP. All the isolates had more or less conserved amino acid sequence in BBTV CP. The statistical indices for nucleotide diversity, Tajima D test, Fu and Li D and F tests also indicated that the above results were significant. The ka/ks ratios of sequence pairs indicated purifying selection in both BBTV CP and BBrMV CP. Surprisingly, sequence pairs with Kerala isolates and GU085261.1, GU085260.1, GU085262.1, JX171699.1 had Ka/Ks >1 indicating Darwinian selection which is rare event considering the results of plant viruses analysed previously. Both CP gene exhibited strong codon usage bias which was calculated on the basis of effective number of codons (Nc), relative synonymous codon (RSCU) and Codon adaptation index (CAI). A strong correlation between the composition of the gene and Nc value was detected. BBrMV isolates infecting *Musa* sp, cardamom and ginger were apparently adapted to all the three host equally. The Nc plot and neutrality plot indicated that the mutational pressure and selection pressure contributed to CUB.

From recombination analysis, MF039868.1, BBTV isolate from Thailand was detected to be a recombinant. This isolate was distinct from others considered for the analysis and clustered separately in principle co-ordinate analysis. Similarly, MT818177 and MT818187, the most diverse isolates in the dataset collected from Kerala, were also identified as recombinants. However, the major parent was not identified but minor parents were HM348782.1 and KF385489.1 respectively.

Considering the molecular diversity primers were designed for nucleic-acid based detection of the virus. It was however evident that in BBTV CP diversity was less among the Kerala isolates. However, primers were so designed complimentary to most conserved region on the upstream and downstream of the ORF to amplify the complete gene. In case of BBrMV, the N terminal of CP was hypervariable and so the primer was designed complimentary to

the conserved NIb region to the 5' of the CP gene in ssRNA. This primer set could amplify gene from few samples that was previously undetected by the reported primers.

Cloning of BBTV CP and BBrMV CP was carried out in to multiple vectors. New sets of primers were designed for cloning so that restriction recognition sites could be incorporated to the gene for cloning to the vectors. BBTV CP gene amplified using the designed primers were cloned into pUC19 at EcoR1 and BamH1 sites for maintaining the gene. It was then cloned into expression vectors like pRSET-C at Nhe1, BamH1 sites, pGEX-4T-2 at Sma1 site and pET32a (+) at EcoRV BamH1 sites. The clones were designated as pRSET/BBTV CP, pGEX/BBTV CP and pET/BBTV CP Apart from conventional cloning, BBTV CP was also cloned into pGEX-4T-2 by restriction free (ligation independent) cloning method in order to manipulate thrombin cleavage site of the vector and designated as ΔpGEX/BBTV CP. The ligated products were transformed to DH5α and selection of recombination was carried out by PCR, restriction digestion and sequencing.

The BBrMV CP gene was cloned into pGEM-T easy vector by TA cloning method for multiplying and maintaining the gene. The gene was also cloned into various expression vectors for production of recombinant coat protein. The gene was then amplified using designed primers and restricted by Nhe1 and BamH1. The product was ligated to double digested pRSET-C. BBrMV CP was also ligated into Sma1 site of pGEX-4T-2. As performed for BBTV CP the selection of clone was carried out by PCR, restriction digestion and sequencing. The orientation of BBTV CP and BBrMV CP cloned into Sma1 site of pGEX-4T-2 was confirmed by restriction digestion and PCR with universal primer as forward and gene specific primer as reverse primers. The BBrMV CP clones were designated as pRSET/BBrMV CP and pGEX/BBrMV CP.

Recombinant coat protein of BBTV CP was produced by inducing Rosetta (DE3) pLysS, BL21 (DE3) pLysS and C41 strains of *E. coli* harbouring clones pRSET/BBTV CP, pGEX/BBTV CP, ΔpGEX/BBTV CP and pET/BBTV CP. While analysing the overexpression pattern, it was clear that the protein was highly soluble in all the vectors. However, pRSET/rBBTV CP was unstable above 4 °C and was subjected to proteolysis on storage. Although, pGEX/rBBTV CP was expressed at 16 °C, the GST tag dissociated from the protein and the untagged protein was apparently unstable. Nevertheless, it was impossible to purify the untagged protein and hence this clone was not used for further studies. ΔpGEX/rBBTV CP when overexpressed produced a highly stable 45 kDa fusion

protein which was identified by Western blot. Similarly, overexpression of pET/rBBTV CP was also highly satisfactory in C41 strain of *E. coli*. A 37 kDa fusion protein was obtained after induction with 17 kDa Trx-S-His tag. Recombinant BBTV CP was purified as GST and Trx fusion protein by GSH and Ni^{2+}-NTA affinity column chromatography. The antigenicity of the fusion protein was assessed by Western blotting. As cleavage of both the affinity tags were unsuccessful and GST tag being bulkier than Trx tag, rBBTV CP-Trx fusion protein was considered for immunisation. Moreover, the epitope prediction of the rBBTV CP-Trx fusion protein, predicted very few epitopes in the affinity tag and a complex major epitope at the junction of the tag and the rBBTV CP including the N terminal of BBTV CP and C terminal of Trx tag, this fusion protein was used for immunisation of animal to raise polyclonal antiserum.

Overexpression and purification of rBBrMV CP was carried out in Rosetta (DE3) pLysS, BL21 (DE3) pLysS harbouring clones pRSET/BBrMV CP, pGEX/BBrMV CP. In the former, the protein was insoluble. Various parameters influencing solubility of the recombinant protein like, temperature and time of induction, concentration of inducer, culture medium, buffer and pH for purification were standardised. The best result was obtained when Rosetta (DE3) pLysS harbouring pRSET/BBrMV CP was cultured in YT medium, and induced by 0.3 mM IPTG for 12 h at 16 °C and CAPS buffer of 9.2 pH was used for resuspending and lysis of induced cell in the presence of protease inhibitor cocktail. In order to increase the solubility, protein purification at denaturing conditions were also attempted but failed to get pure protein after affinity column chromatography. Later, Rosetta (DE3) pLysS harbouring pGEX/BBrMV CP was induced to obtain GST fusion protein of 60 kDa. The soluble protein was assessed for its antigenicity by Western blotting after on-column cleavage of the tag with thrombin. However, the untagged protein was subjected to proteolysis even when purification was done in the presence of protease inhibitor cocktail due to the proteolysis prone surface exposed N and C terminal. Since on storage the protein was degraded, it was not carried forward for antiserum production.

The recombinant coat proteins were characterised experimentally by sucrose gradient ultracentrifugation, electron microscopy and fluorescence spectroscopy apart from Western blot analysis. rBBTV CP with Trx tag apparently formed VLPs. The fusion protein with GST tag did not participate in VLP formation. However, in rBBrMV CP without the tag, octameric ring like structures which are apparently the intermediate of assembly were seen in electron micrograph. Since the N and C terminal of these untagged protein were

subjected to proteolysis, the complete VLP formation could not be accomplished *in vitro*. Fluorescent spectra indicated partial disordered region in the otherwise ordered protein and substantiated the FoldIndex analysis.

Purified tagged rBBTV CP was used for antiserum production. After periodic administration of the antigen into white New Zealand rabbit antiserum was separated from the blood of the animal. The tire was determined as 1:128000 against the antigen used for immunisation. Field sample with high antigen titre was detected by this dilution of antiserum. Inorder to develop a more inclusive assay, the antiserum dilution was standardised as 1:10000. Over 250 field and tissue culture samples were screened for presence of the virus using DAC-ELISA with recombinant antiserum. The antiserum produced against the recombinant coat protein was successful in detecting infection in asymptomatic TC plantlets. On comparing, the titre of commercial antiserum was 1:1000 only and many non-specific reactions could be predicted from the high absorbance value in negative samples. The recombinant antiserum was also used for standardising serological assays like TAS-ELISA, DIBA and IC-PCR. The sensitivity of each assay was determined. DIBA was perfected to be used for in-field detection of the virus.

Future scope of Research

7. FUTURE SCOPE

- Increase stability of rBBrMV CP and use VLPs produced *in vitro* for production of antibody
- Amalgamate nanotechnology to serodiagnostic assays in order to increase its sensitivity
- Perfect DIBA for infield diagnosis of the disease
- Solve the structure of BBTV CP and BBrMV CP

Bibliography

BIBLIOGRAPHY

Abdelkader, H.S., Abdel-Salam, A.M., Saghir, S.M., and El-Hussein, M.H. 2004. Molecular cloning and expression of recombinant coat protein gene of banana bunchy top virus in E. coli and its use in the production of diagnostic antibodies. *Arab J. Biotechnol*. 7: 173–187.

Ach, R. A., Durfee, T., Miller, A. B., Taranto, P., Hanley-Bowdoin, L., Zambryski, P. C., and Gruissem, W. 1997. RRB1 and RRB2 encode maize retinoblastoma-related proteins that interact with a plant D-type cyclin and geminivirus replication protein. *Mol. Cell. Biol*. 17: 5077-5086

Acotto, G.P., Mullineaux, P. M., Brown, S. C., and Marie, D. 1993. Digitaria streak geminivirus replicative forms are abundant in S-phase nuclei of infected cells. *Virology*, 195: 257-259.

Adams, M.J. and Antoniw, J.F. 2004. Codon usage bias amongst plant viruses. *Arch. Virol.* 149: 113–135.

Adams, M.J., Antoniw, J.F., and Fauquet, C.M. 2005. Molecular criteria for genus and species discrimination within the family Potyviridae. *Arch. Virol.* 150: 459–479.

Adegbola, R.O., Ayodeji, O., Awosusi, O.O., Atiri, G.I., and Kumar, P.L. 2013. First report of Banana bunchy top virus in banana and plantain (Musa spp.) in Nigeria. *Plant Dis.* 97: 290p

Agirrezabala, X., Méndez-López, E., Lasso, G., Sánchez-Pina, M.A., Aranda, M. and Valle, M. The near-atomic cryoEM structure of a flexible filamentous plant virus shows homology of its coat protein with nucleoproteins of animal viruses. *Elife.* 4:e11795.

Alexander, D.C., 1987. An efficient vector-primer cDNA cloning system. *Methods Enzymol.* 154, 41–64.

Allen, R. N. 1978. Epidemiological factors influencing the success of roguing for the control of bunchy top disease of bananas in New South Wales. *Aust. J. Agric. Res.* 29(189):535-544.

Allison, R., Johnston, R.E., and Dougherty, W.G. 1986. The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: Evidence for the synthesis of a single polyprotein. *Viroogy*,. 154: 9–20.

Almarza, J., Rincon, L., Bahsas, A., and Brito, F. 2009. Molecular mechanism for the denaturation of proteins by urea. *Biochemistry,* 48: 7608–7613.

Amin, I., Qazi, J., Mansoor, S., Ilyas, M., Briddon, R.W., 2008. Molecular characterisation of Banana bunchy top virus (BBTV) from Pakistan. *Virus Genes* 36, 191–198.

Anhalt, M.D. and Almeida, R.P.P. 2008. Effect of temperature, vector life stage, and plant access period on transmission of Banana bunchy top virus to banana. *Phytopathology,* 98: 743–748.

Anindya, R. and Savithri, H.S. 2003. Surface-exposed amino- and carboxy-terminal residues are crucial for the initiation of assembly in Pepper vein banding virus: A flexuous rod-shaped virus. *Virology,* 316: 325–336.

Anuradha, C. and Selvarajan, R. 2018. Molecular characterization of banana bract mosaic virus from India reveals recombination and positive selection in the VPg gene. *J. Plant Pathol.* 100: 523–531.

Arguello-Astorga, G., Herrera-Estrella, L., and Rivera-Bustamante, R. 1994. Experimental and theoretical definition of geminivirus origin of replication. *Plant Mol. Biol.* 26: 553–556.

Aronson, M.N., Meyer, A.D., Gyorgyey, J., Katul, L., Vetten, H.J., Gronenborn, B. and Timchenko, T. 2000. Clink, a Nanovirus-Encoded Protein, Binds both pRB and SKP. *J. Virol.* 74(7): 2967-2972.

Arumugam, C., Kalaimughilan, K., and Kathithachalam, A. 2017. Banana Bunchy Top Viral Coat Protein (CP) Gene Expression Studies at Molecular Level in Hill Banana cv. Sirumalai (AAB). *Int. J. Curr. Microbiol. Appl. Sci.* 6: 398–411.

Aslanidis, C. and Dejong, P J. 1990. Ligation-independent cloning of PCR products (Lic-PCR). *Nucleic Acids Res*. 18: 6069– 74.

Bagyalakshmi, K., Chinnaraja, C., Karuppaiah, R., Kumar G.V., and Viswanathan, R. 2012. Genetic variability and potential recombination events in the HC-Pro gene of Sugarcane streak mosaic virus (SCSMV). *Arch Virol.* 157(7): 1371–1375.

Balakrishnan, S., Gokulapalan, C., and Paul. S. 1996. A widespread banana malady in Kerala, india. *Infomusa.* 5(1): 28-29.

Balasubramanian, V. and Selvarajan, R. 2012. Complete genome sequence of a banana bract mosaic virus isolate infecting the French plantain cv. Nendran in India. *Arch. Virol.* 157: 397–400.

Balasubramanian, V. and Selvarajan, R. 2014. Genetic diversity and recombination analysis in the coat protein gene of Banana bract mosaic virus. *Virus Genes*, 48: 509–517.

Balasubramanian, V., Sukanya, R.S., Anuradha, C., and Selvarajan, R. 2014. Population structure of Banana bract mosaic virus reveals recombination and negative selection in the helper component protease (HC-Pro) gene. *Virus Dis.* 25: 460–466.

Banerjee, A., Roy, S., Behere, G.T., Roy, S.S., Dutta, S.K., and Ngachan, S.V. 2014. Identification and characterization of a distinct banana bunchy top virus isolate of Pacific-Indian Oceans group from North-East India. *Virus Res.* 183: 41–49.

Bateson, M.F. and Dale, J.L. 1995. Banana bract mosaic virus: Characterization using potyvirus specific degenerate PCR primers. *Arch.Virol.*140: 515–527.

Beetham, P.G.J. Hafner, H.R.M., and Dale, J.L. 1997. Two mRNAs are transcribed from Banana bunchy top virus DNA-1. *J. Gen. Virol.* 78: 229-236.

Behura, S.K. and Severson, D.W. 2013. Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol. Rev. Cambridge Philos. Soc.* 88(1):49-61.

Bennetzen, J.L. and Hall, B.D. 1982. Codon selection in yeasts. *J. Biol. Chem*. 257: 3026-3031.

Bennion, B.J. and Daggett, V. 2003. The molecular basis for the chemical denaturation of proteins by urea. *Proc. Natl. Acad. Sci. USA,* 100: 5142–5147.

Berger, P.H, Adams, M.J., Barnett, O.W., Burnt, A.A., Hammond, J., Hill, J.H., Jordan, R.L., Kashiwazaki, S., Rybicki, E., Spence, N. 2005. Potyviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., and Ball, L.A. (eds), *Virus Taxonomy: Eighth Report of the International Committee on the Taxonomy of Viruses*. Elsevier Academic, San Diego, pp 819–841.

Biswas, K.K., Palchoudhury, S., Chakraborty, P., Bhattacharyya, U.K., Ghosh, D.K., Debnath, P., Ramadugu, C., Keremane, M.L., Khetarpal, R.K., and Lee, R.F. 2019. Codon usage bias analysis of citrus tristeza virus: Higher codon adaptation to citrus reticulata host. *Viruses*, 11: 1–17.

Blackman, R.L. and Eastop, V.F. 1984. *Aphids on the world's crops: an identification and information guide*. John Wiley and Sons, Chichester.

Bressan, A. and Watanabe, S. 2011. Immunofluorescence localization of banana bunchy top virus (family Nanoviridae) within the aphid vector, Pentalonia nigronervosa, suggests a virus tropism distinct from aphid-transmitted luteoviruses. *Virus Res*. 155: 520-525.

Bulmer, M. 1988. Are codon usage patterns in unicellular organisms determined by selection-mutation balance?. *J. Evol. Biol.* 1: 15–26.

Burns, T.M., Harding, R.M. and Dale, J.L. 1995. The genome organization of banana bunchy top virus: analysis of six ssDNA components. *J. Gen. Virol*. 76: 1471–1482.

Butt, A.M., Nasrullah, I., Qamar, R. and Tong, Y. 2016. Evolution of codon usage in Zika virus genomes is host and vector specific. *Emerg. Microbes Infect.* 5: e107.

Calderone, T.L., Stevens, R.D., and Oas,T.G. 1996. High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in Escherichia coli. *J. Mol. Biol.* 262: 407–412.

Camilloni, C., Guerini Rocco, A., Eberini, I., Gianazza, E., Broglia, R.A., and Tiana, G. 2008. Urea and guanidinium chloride denature protein L in different ways in molecular dynamics simulations. *Biophys. J.* 94: 4654–4661.

Cerovska, N., Moravec T., and Veleminsky. 2002. Polyclonal antibodies to a recombinant coat protein of potato virus A. *Acta Virol*. 46: 147–151.

Cerovska, N., Moravec, T., Plchova, H., Hoffmeisterova, H., Folwarczna, J., and Dedic, P. 2010. Production of polyclonal antibodies to potato virus X using recombinant coat protein. *J. Phytopathol*. 158: 66–68.

Cerovska, N., Moravec, T., Rosecka, P., Dedic, P., and Filigarova, M. 2003. Production of polyclonal antibodies to a recombinant coat protein of potato mop-top virus. *J. Phytophatol.* 151: 1–6.

Chakraborty, P., Das, S., Saha, B., Sarkar, P., Karmakar, A., Saha, A., Saha, D., and Saha, A. 2015. Phylogeny and synonymous codon usage pattern of Papaya ringspot virus coat protein gene in the sub-Himalayan region of north-east India. *Canadian J. Microbiol.* 61: 555–564.

Chao, L. 1988. Evolution of sex in RNA viruses. *J. Theor. Biol.* 133: 99–112.

Chen, Q. and Lai, H. 2013. Plant-derived virus-like particles as vaccines. *Hum. Vaccines Immunother*. 9(1):26-49.

Chen, Y. and Hu, X. 2013. High-throughput detection of banana bunchy top virus in banana plants and aphids using real-time TaqMan®PCR. *J. Virol. Methods*. 193: 177–183.

Cherian, A.K., Menon, R., Suma, A., Nair, S., and Sudheesh, M.V. 2002. Impact of banana bract mosaic diseases on the yield of commercial banana varieties of Kerala [abstract]. In: *Global Conference on Banana and Plantain*; 28–31, October, 2002, Bangalore, Abstract pp. 155.

Chiaki, Y., Nasir, N., Herwina, H., Sonoda J. A., Fukumoto, T., Nakamura, M., and Iwai, M. 2015. Genetic structure and diversity of the Banana bunchy top virus population on Sumatra Island, Indonesia. *Eur. Plant Pathol*. 143: 113–122

Choi, J.K. and Wakimoto, S. 1979. Characterization of the protein components of turnip mosaic virus. *Ann. Phytopathol. Soc. Japan*. 45: 32-39.

Chung, B.Y.W., Miller, W.A., Atkins, J.F., and Firth, A.E. 2008. An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci. U. S. A.* 105: 5897–5902.

Clark, M.F. and Adams, A.N. 1977. Characteristics of the microplate method of enzyme linked immunosorbent assay for the detection of plant viruses. *J. Gen. Virol*. 34: 475–483.

Colovos, C. and Yeates, T.O. 1993. Verification of protein structures: patterns of non bonded atomic interactions. *Protein Sci.* (9):1511-1519.

Costa, S., Almeida, A., Castro, A., and Domingues, L. 2014. Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: The novel Fh8 system. *Front. Microbiol*. 5: 1–20.

Cuesta, R., Yuste-calvo, C., Gil-cartón, D., Sánchez, F., Ponz, F., and Valle, M. 2019. Structure of turnip mosaic virus and its viral-like particles. *Sci. Rep.* 9:1-6

Dale, J.L. 1987. Banana bunchy top: an economically important tropical plant virus disease. *Adv. Virus. Res*. 33: 301-325.

Das, T. and Banerjee, A. 2018. Distribution, molecular characterization and diversity of banana bunchy top virus in Tripura, India. *Virus Dis*. 29: 157–166.

Desbiez, C., Chandeysson, C., and Lecoq, H. 2014. A short motif in the N-terminal part of the coat protein is a host-specific determinant of systemic infectivity for two potyviruses. *Mol. Plant Pathol*. 15: 217–221.

Dhanya, M.K. 2004. Purification and immunodetection of Banana bract mosaic virus. Ph.D thesis, Kerala Agricultural University, Thrissur, 120p.

Diekmann, M and C. A. J. Putter. 1996. FAO/IPGRI. Technical guidelines for the safe movement of germplasm No 15. Musa. 2nd edition. Food and Agriculture Organization of the United Nations, Rome/International Plant Genetic Resources Institute: Rome.

DiMaio, F., Song, Y., Li, X., Brunner, M.J., Xu, C., Conticello, V., Egelman, E., Marlovits, T., Cheng, Y. and Baker, D. 2015. Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat Methods*, 12(4):361-365.

Dombrovsky, A., Huet, H., Chejanovsky, N. and Raccah, B. 2005. Aphid transmission of a potyvirus depends on suitability of the helper component and the N terminus of the coat protein. *Arch Virol.* 150(2):287-98.

Domingo, E. and Holland, J.J. 1997. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51: 151–178.

Domingo-Espín, J., Unzueta, U., Saccardo, P., Rodríguez-Carmona, E., Corchero, J.L., Vazquez, E., and Ferrer-Miralles, N. 2011. Engineered biological entities for drug delivery and gene therapy: Protein nanoparticles, in: Progress in Molecular Biology and Translational Science. *Elsevier B.V.* 104: 247–298.

Dougherty, W.G. and Carrington, J.C. 1988. Expression and Function of Potyviral Gene Products. *Annu. Rev. Phytopathol.* 26: 123–143.

Drew, R.A., Moisander, J.A., and Smith, M. K. 1989. The transmission of banana bunchy top virus in micropropagated bananas. *Plant Cell Tissue and Organ Cult*. 16: 187–193.

Dubendorff, J.W. and Studier, F.W. 1991. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* 219: 45–59.

Duffy, S., Shackelton, L.A., and Holmes, E.C. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet*. 9**:** 267-276.

Dugdale, B., Beetham, R., Becker, K., Harding, R.M., and Dale, J.L. 1998. Promoter activity associated with the intergenic regions of banana bunchy top virus DNA-1 to 6 in transgenic tobacco and banana cells. *J. Gen. Virol.* 79: 2301–2311.

Eastwood, H.W. 1946. A successful campaign. Bunchy top disease of bananas. Controlled by co-operative effort. Agricultural Gazette Nov, Dec, 1946 and Jan, 1947.

Edgar, R.C., 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinforma*. 5: 113.

Eigen, M., McCaskill, J., and Schuster, P. 1988. Molecular quasi-species. *J. Phys. Chem.* 92: 6881–6891.

Elayabalan, S., Subramaniam, S., and Selvarajan, R. 2015. Banana bunchy top disease (BBTD) symptom expression in banana and strategies for transgenic resistance: A review. *Emirates J. Food Agric.* 27: 55–74.

Espino, R.C., Magnaye, L.V., Johns, A.P., and Juanillo, C. 1993. Evaluation of Philippine banana cultivars for resistance to bunchy-top and Fusarium wilt. In: Valmayor, R.C., Hwang, S.W., Ploetz, R., Lee, S.C., Roa, V.N. (eds), *Recent Developments*

in Banana Cultivation Technology. Proceedings of the international symposium, Taiwan. Banana Research Institute, Taiwan and Philippines, pp. 89–102.

Espino, T.M., Exconde, S.B., Zipagan, F.B., and Espino, R.R.C. 1990. Banana bract mosaic, a new disease of banana II. Isolation and purification for monoclonal antibody production. *Philipp. Agric*. 73: 61–68.

Espino, T.M., Exconde, S.B., Zipagan, F.B., Maroon, M.C.M., and Espino, R.R.C. 1989. Production of monoclonal antibodies for diagnosis of banana bunchy top virus. *Philipp. Agric*. 72: 473–485.

Esposito, D. and Chatterjee, D.K. 2006. Enhancement of soluble expression through the use of fusion tags. *Curr. Opi. Biotechnol*. 17: 353–358.

Estelitta, S. 1998. Purification and serology of Banana bunchy top virus. Ph.D thesis, Kerala Agricultural University, Thrissur, 116p.

Estelitta, S., Suma, A., and Radhakrishnan, T. C. 1996. Field screening of banana germplasm against kokkan disease, *J. Trop. Agric.* 34 (1): 68-69.

FAOSTAT [Food and Agricutural Organisation Statistics] 2018. FAO Production statistics for Bananas. http://www.fao.org/faostat/en/.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol*. 17: 368–376.

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

Folwarczna, J., Plchová, H., Moravec T., Hoffmeisterova H., Dedic P., and Cerovska N. 2008. Production of Polyclonal Antibodies to a Recombinant Coat Protein of Potato Virus Y. *Folia Microbiol*. 53 (5): 438–442.

Fu, H.C., Hu, J.M., Hung, T.H., Su, H.J., and Yeh, H.H. 2009. Unusual events involved in Banana bunchy top virus strain evolution. *Phytopathol.* 99(7):812-22.

Fu, Y.X. and Li, W.H. 1993. Statistical tests of neutrality of mutations. *Genetics*, 133: 693–709.

Furuya, N., Bahet, N.B., Kawano, S., and Natsuki, K. 2003. Detection of viruses and genetic analysis of Banana bunchy top virus associated with Abaca. *Japan J. Phytopathol.* 69: 326–327.

Furuya, N., Kawano, S., and Natsuaki, K.T. 2005. Characterization and genetic status of Banana bunchy top virus isolated from Okinawa. *Japan J. Gen. Plant Pathol*. 71: 68–73.

Galal, A.M. 2007. Use of polymerase chain reaction for detecting banana bunchy top nano virus. *Biotechnology*. 6: 53–56.

Garcia-Arenal, F., Fraile, A., and Malpica, J.M. 2003. Variation and evolution of plant virus populations. *Int. Microbiol*. 6: 225–232.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31:3784-3788.

Geering, A.D.W and Thomas, J.E. 1997. Search for alternative hosts of banana bunchy top virus in Australia. *Australian Plant Pathol*. 26: 250–254.

Geering, A.D.W. and Thomas, J.E. 1996. A comparison of four serological tests for the detection of banana bunchy top virus in banana. *Australian J. Agric. Res*. 47: 403–412.

Geourjon, C. and Deleage, G. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* 11 (6): 681–684.

Geourjon, C. and Deléage, G. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci*. 11(6):681-4.

Gibbs, A. and Ohshima, K. 2010. Potyviruses and the digital revolution. *Annu. Rev. Phytopathol.* 48: 205–223.

Gibbs, M,J. and Weiller, G.F. 1999. Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. *Proc. Natl. Acad. Sci. USA.* 96(14):8022-8027.

Goodman, R.M., McDonald, J.G., Horne, R.W., and Bancroft, J.B. 1976. Assembly of flexuous plant viruses and their proteins. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. 276: 173–179.

Gough, K.H. and Shukla, D.D. 1981. Coat protein of potyviruses 1. Comparison of the four Australian strains of sugarcane mosaic virus. *Virology,* 111: 455-462.

Gouy, M. and Gautier, C. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res*. 10(22): 7055-7074.

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R. 1981. *Nucleic Acids Res*. 9: r43-r79.

Green, M. R. and Sambrook, J. 2012. *Molecular cloning A laboratory manual. Vol 1*. (4th Ed.). Cold Spring Harbour Laboratory Press, New York. Pp. 11-19.

Grigoras, I., Timchenko, T., Grande-Pérez, A., Katul, L., Vetten, H.J., and Gronenborn, B. 2010. High Variability and Rapid Evolution of a Nanovirus. *J. Virol*. 84: 9105–9117.

Gulati, A., Alapati, K., Murthy, A., Savithri, H.S., Murthy, M.R.N., 2016a. Structural studies on tobacco streak virus coat protein: Insights into the pleomorphic nature of ilarviruses. *J. Struct. Biol*. 193: 95–105.

Gulati, A., Murthy, A., Abraham, A., Mohan, K., Natraj, U., Savithri, H.S., Murthy, M.R.N., 2016b. Structural studies on chimeric Sesbania mosaic virus coat protein: Revisiting SeMV assembly. *Virology*, 489: 34-43.

Gundry, R.L., White, M.Y., Murray, C.I., Kane, L.A., Fu, Q., Stanley, B.A., and Van-Eyk, J.E. 2010. Preparation of Proteins and Peptides for Mass Spectrometry Analysis. *Curr. Prot. Mol. Biol.* 90(1): 342–355.

Ha, C., Coombs, S., Revill, P.A., Harding, R.M., Vu, M., and Dale, J.L. 2008. Design and application of two novel degenerate primer pairs for the detection and complete genomic characterization of potyviruses. *Arch. Virol*. 153: 25–36.

Hafner, G. J., Harding, R. M., and Dale, J. L. 1997a. A DNA primer associated with banana bunchy top virus. *J. Gen. Virol*.78: 1795–1799.

Hafner, G.J., Stafford, M.R., Wolter, L.C., Harding, R.M., and Dale, J.L. 1997b. Nicking and joining activity of banana bunchy top virus replication protein in vitro. *J. Gen. Virol*. 78: 1795–1799.

Hafner, J., Harding, R.M., Dale, J.L. 1995. Movement and transmission of banana bunchy top virus DNA component one in bananas. *J. Gen. Virol*. 76: 2279–2285.

Hamdayanty, H., Hidayat, S.H. and Damayanti, T.A. 2016. Expression of Recombinant Sugarcane Streak Mosaic Virus Coat Protein Gene in Escherichia coli. *HAYATI J. Biosci.* 23: 111–116.

Harding, R.M., Burns, T.M., and Dale, J.L. 1991. Virus-like particles associated with banana bunchy top disease contain small single-stranded DNA. *J. Gen. Virol*. 72: 225–230.

Harper, G., Dahal, G., Thottappilly, G., and Hull, R. 1999. Detection of episomal banana streak badnavirus by IC-PCR. *J. Virol. Methods* 79: 1–8.

Harrison, B.D., Robinson, D.J. 1988. Molecular variation in vector-borne plant viruses: Epidemiological [and Discussion] significance. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. 321: 447–462.

He, Z.F., Li, H.P., Xiao, H.G. and Fan, H.Z. 2000. Cloning and sequencing of DNA component 1 of two BBTV strains. *Acta Phytopathol. Sin*. 30:364–369.

Hema, M., Kirthi, N., Sreenivasulu, P., Savithri, H.S. 2003. Development of recombinant coat protein antibody based IC-RT-PCR for detection and discrimination of sugarcane streak mosaic virus isolates from Southern India. *Arch. Virol.* 148: 1185–1193.

Hershberg, R. and Petrov, D.A. 2008. Selection on codon bias. *Annu. Rev. Genet.* 2008, 42: 287–299.

Hiebert, E. and McDonald, J. G. 1973. Characterization of some proteins associated with viruses in the potato Y group. *Virol.* 56: 349-361.

Hiebert, E., Tremaine, J.H., and Ronald, W.P. 1984. The effect of limited proteolysis on the amino acid composition of five potyviruses and on the serological reaction and peptide maps of the tobacco etch virus capsid protein. *Phytopathol*. 74: 411-416.

Hollings, M. and Brunt, A.A. 1981. Potyvirus. In: Kurstak E. (ed.), *Handbook of Plant Virus Infections: Comparative Diagnosis*. Elsevier/North Holland, Amsterdam, pp. 147–170.

Holmes, E.C. 2009. In: Harvey, P.M. and May, R.M. (eds.), *The Evolution and Emergence of RNA Viruses*. Oxford University Press, Oxford, 249p.

Holmes, E.C., Worobey, M., and Rambaut, A. 1999. Phylogenetic Evidence for Recombination in Dengue Virus. *Mol. Biol. Evol*. 16: 405–409.

Homer, R.B. and Goodman, R.M. 1975. Circular dichroism and fluorescence studies on *Potato virus X* and its structural components. *Biochimica et Biophysica Acta*. 378: 296-304.

Hooks, C.R.R., Manandhar, R., Perez, E.P., Wang, K.H., and Almeida, R.P.P. 2009. Comparative susceptibility of two banana cultivars to banana bunchy top virus under laboratory and field environments. *J. Econ. Entomol*. 102: 897–904.

Hooks, C.R.R., Wright, M.G., Kabasawa, D.S., Manandhar, R., and Almeida, R.P.P. 2008. Effect of banana bunchy top virus infection on morphology and growth characteristics of banana. *Ann. Appl. Biol*. 153: 1–9.

Horser, C.L., Harding, R.M. and Dale, J.L. 2001a. Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. *J. Gen. Virol*. 82: 459-464.

Horser, C.L., Karan, M., Harding, R.M., and Dale, J.L. 2001b. Additional Rep-encoding DNAs associated with banana bunchy top virus. *Arch. Virol*. 146: 71–86.

Hu, J. S., Wang, M., Sether, D., Xie, W., and Leonhardt, K. W. 1996. Use of polymerase chain reaction (PCR) to study transmission of banana bunchy top virus by the banana aphid (Pentalonia nigronervosa). *Ann. Appl. Biol*.128: 55–64.

Hu, JM., Fu, H.C., Lin, C.H., Su, H.J., and Yeh, H.H. 2007. Re-assortment and concerted evolution in Banana bunchy top virus genomes. *J. Virol*. 81:1746–1761.

Hudson, R.R. and Kaplan, N.L. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, 111(1):147-64.

Hull, R. 2009. *Comparative Plant Virology* (2nd Ed.). Academic press. China. 255p.

Huttinga, H., Mosch, W.H.M. 1974. Properties of viruses of the potyvirus group. 2. Buoyant density, S-value, particle morphology and molecular weight of the coat protein subunit of bean yellow mosaic virus, pea mosaic virus, lettuce mosaic virus and potato virus Y^N. *Neth. J. Plant Pathol*. 80: 19-27.

Hyder, M.Z., Shah, S.H., Hameed, S., and Naqvi, S.M.S. 2011. Evidence of recombination in the Banana bunchy top virus genome. Infect. *Genet. Evol*. 11: 1293–1300.

Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol*. 2(1): 13-34.

Iskra-Caruana, M.L., Galzi, S., and Laboureau, N. 2008. A reliable IC One-step RT-PCR method for the detection of BBrMV to ensure safe exchange of Musa germplasm. *J. Virol. Methods*. 153: 223–231.

Iskra-Caruana, M.L., Garnier, M., and Bove, J. M. 1989. Purification of banana bunchy top virus. *Fruits*. 44: 63–66.

Jacob, T. and Usha, R. 2002. Expression of Cardamom mosaic virus coat protein in Escherichia coli and its assembly into filamentous aggregates. *Virus Res.* 86: 133–141.

Jagadish, M.N., Huang, D., and Ward, C.W. 1993. Site-directed mutagenesis of a potyvirus coat protein and its assembly in Escherichia coli. *J. Gen. Virol*. 74: 893–896.

Jagadish, M.N., Ward, C.W., Gough, K.H., Tulloch, P.A., Whittaker, L.A., and Shukla, D.D. 1991. Expression of potyvirus coat protein in Escherichia coli and yeast and its assembly into virus-like particles. *J. Gen. Virol*. 72: 1543–1550.

Jawdah, A., Sobh, H., Cordahi, N., Kawtharani, H., Nemer, G., Maxwell, D. P., and Nakhla, M. K. 2004. Immunodiagnosis of Prune dwarf virus using antiserum produced to its recombinant coat protein. *J. Virol. Method.* 121(1): 31- 38.

Jenkins, G.M., Rambaut, A., Pybus, O.G., and Holmes, E.C. 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J. Mol. Evol*. 54**:**156-165.

Jenny, R.J., Mann, K.G., and Lundblad, R.L. 2003. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr. Purif*. 31: 1–11.

Jespersen, M.C., Peters, B., Nielsen, M., and Marcatili, P. 2017. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res*. 45: W24–W29.

Jones, D.T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol*. 292: 195–202.

Jose, P.C. 1981. Reaction of different varieties of banana against bunchy top disease. *Agric. Res. J. Kerala.* 9: 96–97.

Joseph, J. and Savithri, H.S. 1999. Determination of 3'-terminal nucleotide sequence of pepper vein banding virus RNA and expression of its coat protein in Escherichia coli. *Arch. Virol.* 144: 1679–1687.

Kagy, V., Thomas, J.E., Sharman, M., and Mademba-Sy, F. 2001. First record of banana bunchy top disease in New Caledonia. *Aust. Plant Pathol*. 30:71.

Kan, J.H., Van Resandt, R.W.W., and Dekkersf, H.P.J.M. 1986. Fluorescence studies on the coat protein of alfalfa mosaic virus. *J. Biomol. Struct. Dyn*. 3: 827–842.

Kapust, R. B. and Waugh, D. S. 2000. Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr Purif.* 19: 312-318.

Karan, M., Harding, R.M., Dale, J.L., 1994. Evidence for two groups of banana bunchy top virus isolates. *J. Gen. Virol.* 75, 3541–3546.

Kauffman, C. and Karypis, G. 2008. An analysis of information content present in protein–DNA interactions. *Pac Symp Biocomput*. 13: 477–488.

Kawano, S., H. J. Su. 1993. Occurrence of banana bunchy top virus in Okinawa. *Ann. Phytopathol. Soc. Japan*, 59:53.

Kehoe, M.A., Coutts, B.A., Buirchell, B.J. and Jones, R.A. 2014. Split personality of a Potyvirus: to specialize or not to specialize? *PLoS ONE*. 9(8): e105770.

Keightley, P. D. and Otto, S. P. 2006 Interference among deleterious mutations favours sex and recombination in finite populations. *Nature*, 443, 89–92.

Kendall, A., McDonald, M., Bian, W., Bowles, T., Baumgarten, S.C., Shi, J., Stewart, P.L., Bullitt, E., Gore, D., Irving, T.C., Havens, W.M., Ghabrial, S.A., Wall, J.S., Stubbs, G., 2008. Structure of Flexible Filamentous Plant Viruses. *J. Virol*. 82, 9546–9554.

Kenyon, L., Brown, M. and Khonje, P. 1997 First report of Banana bunchy top virus in Malawi. *Plant Dis*. 81:1096

Kežar, A., Kavčič, L., Polák, M., Nováček, J., Gutiérrez-Aguirre, I., Žnidarič, M.T., Coll, A., Stare, K., Gruden, K., Ravnikar, M., Pahovnik, D., Žagar, E., Merzel, F., Anderluh, G., Podobnik, M., 2019. Structural basis for the multitasking nature of the potato virus Y coat protein. *Sci. Adv.* 5: 1–14.

Khalid, S., Soomro, M. H. and Stover, R. H. 1993. First report of banana bunchy top virus in Pakistan. Plant Dis. 77(1):101.

Khan, S., Tasleem Jan, A., Aquil, B., Mohd Rizwanul Haq, Q., 2011. Coat Protein Gene Based Characterization of Cucumber Mosaic Virus Isolates Infecting Banana in India. *J. Phytopathol.* 32, 94–101.

Kim, H. S.; Kim, D. W.; Jung, Y. T. Development of serological procedures for sensitive, rapid detection of Cucumber mosaic virus in Lilium. *Hortic. Environ. Biotechnol.* 2016, 57(6), 633-639.

Kimura, M., 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111-120.

Kiranmai, G., Kumar, P. L., Hema, M., Venkatramana, M., Kirshna-Prasadji, Mandheva-Rao, J., et al. (2005). Partial characterization of a potyvirus causing bract mosaic of banana in Andhra Pradesh. *Indian J. Virol.* 16: 7–11.

Kneller EL, Rakotondrafara AM, and Miller, W.A. 2006. Cap-independent translation of plant viral RNAs. *Virus Res.* 119: 63–75.

Knight, R.D., Freeland, S.J. and Landweber, L.F. 2001. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol.* 2(4): RESEARCH0010.

Koolivand, D., Sokhandan Bashir, N., and Rostami, A. 2017. Preparation of polyclonal antibody against recombinant coat protein of Cucumber mosaic virus isolate B13. *J. Crop Protect*. 6(1): 25-34.

Kosack, C.S., Page, A.L., and Klatser, P.R. 2017. A guide to aid the selection of diagnostic tests. WHO, 95(9): 636-645.

Kringelum, J.V., Lundegaard, C., Lund, O., and Nielsen, M. 2012. Reliable B Cell Epitope Predictions : Impacts of Method Development and Improved Benchmarking. *PLoS Comput. Biol.* 8(12): 1-10. https://doi.org/10.1371/journal.pcbi.1002829

Krishnamoorthy, V. 2002. Breeding for resistance to Sigatoka leaf spot and nematodes in banana (Musa spp.). Ph.D. (Hort.) thesis. Tamil Nadu Agricultural University, Coimbatore.

Ksenofontov, A.L., Dobrov, E.N., Fedorova, N. V., Serebryakova, M. V., Prusov, A.N., Baratova, L.A., Paalme, V., Järvekülg, L., and Shtykova, E.V. 2018. Isolated Potato Virus A coat protein possesses unusual properties and forms different short virus-like particles. *J. Biomol. Struct. Dyn.* 36: 1728–1738.

Ksenofontov, A.L., Paalme, V., Arutyunyan, A.M., Semenyuk, P.I., Fedorova, N. V., Rumvolt, R., Baratova, L.A., Järvekülg, L., and Dobrov, E.N. 2013. Partially Disordered Structure in Intravirus Coat Protein of Potyvirus Potato Virus A. *PLoS ONE*, 8(7): 1-7. https://doi.org/10.1371/journal.pone.0067830

Kumar, N., Bera, B.C., Greenbaum, B.D., Bhatia, S., Sood, R., Selvaraj, P., Anand, T., Tripathi, B.N., and Virmani, N. 2016. Revelation of influencing factors in overall codon usage bias of Equine influenza viruses. *PLoS ONE*, 11(4): e0154376.

Kumar, P. L., Ayodele, M. Oben, T.T., Mahungu, N.M., Beed, F., Coyne, D., Londa, L., Mutunda, P., M., Kiala, D. and Maruthi, M.N. 2009. First report of Banana bunchy top virus in banana and plantain (Musa spp.) in Angola. *Plant Pathol*. 58(2): 402.

Kumar, P.L., Selvarajan, R., Iskra-Caruana, M.L., Chabannes, M., and Hanna, R. 2015a. Biology, etiology, and control of virus diseases of banana and plantain In: *Advanced Virus Res.* (1st Ed.), Elsevier Inc.

Kumar, P.V., Sharma, S.K., Rishi, N., and Baranwal, V.K. 2018. Efficient immunodiagnosis of Citrus yellow mosaic virus using polyclonal antibodies with an expressed recombinant virion-associated protein. 3 *Biotech.* 8: 1–7.

Kumar, S., Karmakar, R., Garg, D.K., Gupta, I., and Patel, A.K. 2019. Elucidating the functional aspects of different domains of bean common mosaic virus coat protein. *Virus Res*. 273: 197755.

Kumar, S., Rai, R., and Baranwal, V.K. 2015. Development of an immunocapture–reverse transcription–polymerase chain reaction (IC-RT-PCR) using modified viral RNA release protocol for the detection of Grapevine leafroll-associated virus 3 (GLRaV-3). *Phytoparasitica*. 43: 311–316.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol*. 35:1547-1549.

Kurland, C. and Gallant, J. 1996. Errors of heterologous protein expression. *Curr. Opin. Biotechnol.* 7: 489–493.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nat.* 227: 680–685.

LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., and McCoy. J.M. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Biotechnol*.; 11: 187–193.

Lefeuvre, P., Lett, J. M., Reynaud, B., Martin, D.P. 2007. Avoidance of protein fold disruption in natural virus recombinants. *PLoS Path*. 3: e181.

Lefeuvre, P., Lett, J.M., Varsani, A., and Martin, D.P. 2009. Widely conserved recombination patterns amongst single stranded DNA viruses. *J. Virol.* 83: 2697–2707.

Lesley, S.A. and Wilson, I.A. 2005. Protein production and crystallization at the joint center for structural genomics. *J. Struct. Funct. Genomics*. 6(2-3):71-9.

Li, X., Song, B., Hu, D., Wang, Z., Zeng, M., Yu, D., Chen, Z., Jin, L., and Yang, S. 2012. The development and application of new crystallization method for tobacco mosaic virus coat protein. *Virol. J.* 9: 1–12.

Li, Z., Jaroszewski, L., Iyer, M., Sedova, M., and Godzik, A. 2020. FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Res*. 48: W60–W64.

Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J., Tan, S. 2005. Comparison of affinity tags for protein production. Protein Expression Purification. 41: 98–105.

Ling, K.S., Zhu, H.Y., Jiang, Z.Y., and Gonsalves, D. 2000. Effective application of DAS-ELISA for detection of grapevine leafroll associated closterovirus-3 using a polyclonal antiserum developed from recombinant coat protein. *Eur. J. Plant Pathol*. 106(4): 301-309.

Liu, F., Feng, L., Chen, X., Han, Y., Li, W., Xu, W. 2012. Simultaneous detection of four banana viruses by multiplex PCR. *J. Phytopathol*. 160: 622–627.

Lowe, S., Browne, M., Boudjelas, S., De, P.M. 2000. 100 of the world's worst invasive alien species a selection from the global invasive species database. Published by the Invasive Species Specialist Group (ISSG) a specialist group of the Species Survival Commission (SSC) of World Conservation Union (IUCN)

Magee C.J.P. 1948. Transmission of banana bunchy top to banana varieties. J. Aust. Agric. Sci. 14: 18–24.

Magee, C.J.P. 1927. Investigation on the bunchy top disease of the banana. Council for Scientific and Industrial Research Bulletin.

Magee, C.J.P. 1967. Report to the government of Western Samoa: the control of banana bunchy top. South Pacific Commission, Technical Paper No. 150.

Magnaye, L.V. and Espino, R.R.C. 1990. Note: Banana bract mosaic, a new disease of banana I. Symptomatology. *Philipp. Agric*. 73: 55–59.

Magnaye, L.V. and Valmayor, R.V. 1995. BBTV, CMV and other viruses affecting banana in Asia and the Pacific. *Food Fer. Technol. Cent.* 6: 1-6.

Mahadev, S.R., Thamilarasan, S.K., and Kathithachalam, A. 2013. PCR Detection of Banana Bunchy Top Virus (BBTV) at Tissue Culture Level for the Production of Virus-free Planting Materials. *Int. Res. J. Biol. Sci. Int. Res. J. Biol. Sci*. 2: 2278–3202.

Maity, R., Pauty, J., Krietsch, J., Buisson, R., Genois, M.M., and Masson, J.Y. 2013. GST-his purification: A two-step affinity purification protocol Yielding full-length purified proteins. *J. Vis. Exp*. (80): e50320. doi: 10.3791/50320.

Mandal, B. 2010. Advances in small isometric multicomponent ssDNA viruses infecting plants. *Indian J. Virol.* 21: 18–30.

Mandal, B., Shilpi, S., Barman, A.R., Mandal, S., and Varma, A. 2013. Nine novel DNA components associated with the foorkey disease of large cardamom: Evidence of a distinct babuvirus species in Nanoviridae. *Virus Res.* 178: 297–305.

Mansoor, S., Qazi, J., Amin, I., Khatri, A., Khan, I.A., Raza, S., Zafar, Y. 2005. Briddon, R.W. A PCR-based method, with internal control, for the detection of Banana bunchy top virus in banana. *Mol Biotechnol.* 30(2):167-170.

Martin, D. and Rybicki, E. 2000. RDP: Detection of recombination amongst aligned sequences. *Bioinformatics*, 16: 562–563.

Martin, D.P., Posada, D., Crandall, K.A. and Williamson, C. 2005. A Modified Bootscan Algorithm for Automated Identification of Recombinant Sequences and Recombination Breakpoints. *AIDS Res. Human Retroviruses*, 21 (1): 98–102.

Maynard-Smith, J. 1992. Analyzing the Mosaic Structure of Genes. *J. Mol. Evol.* 34: 126–129.

McDonald, J.G., Bancroft, J.B., 1977. Assembly studies on potato virus Y and its coat protein. *J. Gen. Virol.* 35, 251–263.

McGuffin, L.J. and Jones, D.T. 2003. Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics*, 19(7):874-81.

McKern, N.M., Edskes, H.K., Ward, C.W., Strike, P.M., Barnett, O.W., and Shukla, D.D. 1991. Coat protein of potyviruses 7. Amino acid sequence of peanut stripe virus. *Arch. Virol.* 119: 25–35.

McLachlan, A.D., Staden, R. and Boswell, D.R. 1984. A method for measuring the non-random bias of a codon usage table. *Nucleic Acids Res.* 12: 9567-9575.

Mezulis, S., Yates, C.M., Wass, M.N., E Sternberg, M.J., and Kelley, L.A. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc*. 10(6): 845-858.

Moghal, S.M. and Francki, R.I.B. 1976. Towards a system for the identification and classification of potyviruses. 1. Serology and amino acid composition of six distinct viruses. *Virol.* 73: 350-362.

Muhire, B.M., Varsani, A., and Martin, D.P. 2014. SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One* 9 (9): e108277 https://doi.org/10.1371/journal.pone.0108277

Mumford, R.A. and Seal, S.E. 1997. Rapid single tube immunocapture RT-PCR for the detection of two yam potyviruses. *J. Virol. Methods.* 69: 73–79.

Munez, A. R. 1992. Symptomatology, transmission and purification of banana bract mosaic virus (BBMV) in 'giant cavendish' banana. Faculty of graduate school, University of Philippines, pp 1–57.

Nallamsetty, S., Kapust, R. B., Tözsér, J., Cherry, S., Tropea, J. E., Copeland, T.D., Waugh, D.S. 2004. Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. *Protein Expr Purif.* 38: 108-115.

Nee, S., 1987. The evolution of multicompartmental genomes in viruses. *J. Mol. Evol*. 25: 277–281.

Neet, K.E. and Lee, J.C. 2002. Biophysical characterization of proteins in the post-genomic era of proteomics. *Mol. Cell. Proteomics*. 1: 415–420.

Nei, M. and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol*. 3: 418-426.

Nelson, M.I., Viboud, C., Simonsen, L., Bennett, R. T., Griesemer, S. B., George, K., Taylor, J., Spiro, D. J., Sengamalay, N. A., Ghedin, E., Taubenberger, J. K., and Holmes, E. C. 2008. Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918. *PLoS Pathol.*. 4 (2): e1000012.

Nelson, S. C. 2004. Banana bunchy top: Detailed signs and symptom. In: *Cooperative extension service college of tropical agriculture and human resources,* Ma¯noa: University of Hawai‘i. 22p.

Nguyen, H.D., Tran, H.T., Ohshima, K. 2013. Genetic variation of the Turnip mosaic virus population of Vietnam: a case study of founder, regional and local influences. *Virus Res.* 171: 138–149.

NHB [National Horticultre Board]. NHM home page [on line]. http://nhb.gov.in/report_files/ banana/BANANA.htm [Accessed on 25.11.2020].

Nigam, D., LaTourrette, K., Souza, P.F.N., and Garcia-Ruiz, H. 2019. Genome-Wide Variation in Potyviruses. *Front. Plant Sci*. 10: 1–28.

Nikolaeva, O.V., Karasev, A.V., Gumpf, D.J., Lee, R.F., and Garnsey, S.M. 1995. Production of polyclonal antisera to the coat protein of citrus tristeza virus expressed in E. coli: application for immunodiagnosis. *Phytopathol*. 85(6): 691-694.

Niyongere, C., Ateka, E. A., Losenge, T., Lepoint, P., & Blomme, G. (2011). Screening Musa genotypes for banana bunchy top disease resistance in Burundi. *Acta Horticulturae,* 897, 439–447.

Nolasco, G., De Blas, C., Torres, V., and Ponz, F. 1993. A method combining immunocapture and PCR amplification in a microtiter plate for the detection of plant viruses and subviral pathogens. *J. Virol. Methods*. 45: 201–218.

Nugent, T. and Jones, D.T. 2010. Predicting transmembrane helix packing arrangements using residue contacts and a force-directed algorithm. *PLoS Comput. Biol*. 6: e1000714.

Oben, T.T., Hanna, R., Ngeve, J., Alabi, O.J., Naidu, R.A., and Lava Kumar, P. 2009. Occurrence of banana bunchy top disease caused by the banana bunchy top virus on banana and plantain (Musa sp.) in Cameroon. *Plant Dis*. 93(10): 1076.

Ojosnegros, S., Garcı´a-Arriaza, J., Escarmı´s, C., Manrubia, S.C., Perales, C., Armando, A., Mateu, M.G., and Domingo, E. 2011. Viral genome segmentation can result from a trade-off between genetic content and particle stability. *PLoS Genet.* 7 (3): e1001344.

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

Pandey, M. 2015. Immunodiagnosis of Cucumber mosaic virus (CMV), with polyclonal antibodies raised against coat protein (CP) gene of the virus. M.Sc. (Ag.) thesis, University of Agricultural Sciences GKVK, Bengaluru, 88 p.

Panhwar, M.H. 1991. The banana epidemic in Sindh: Imported disease or deliberate sabotage?. *Sindh Q.* 19(4): 12-23.

Peng, J., Zhang, J., Xia, Z., Li, Y., Huang, J., and Fan, Z. 2012. Rapid and sensitive detection of banana bunchy top virus by loop-mediated isothermal amplification. J. Virol. Methods. 185: 254–258.

Perrier, X., De, Langhe, E., Donohue, M., Lentfer, C., Vrydaghs, L., Bakry, F., Carreel, F., Hippolyte, I., Horry, J.P., Jenny, C., Lebot, V., Risterucci, A.M., Tomekpe, K., Doutrelepon, H., Ball, T., Manwaring, J., De-Maret, P., Penham, T. 2011. Multidisciplinary perspectives on banana (Musa spp.) domestication. *Proc Natl Acad Sci USA*. 108: 11311–11318.

Perrier, X., Flori, A., and Bonnot, F. 2003. Data analysis methods. In: Hamon, P., Seguin, M., Perrier, X., Glaszmann, J. C. (eds), Genetic Diversity of Cultivated Tropical Plants. Enfield, Science Publishers, Montpellier, pp. 43- 76.

Perrier, X., Jacquemoud-Collet, J.P., 2006. DARwin software http://darwin.cirad.fr/

Posada, D. and Crandall, K.A. 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* 98: 13757–13762.

Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I., and Sussman, J.L. 2005. FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinforma.* 21: 3435–3438.

Provine, W.B. 2004. "Ernst Mayr: Genetics and speciation". *Genet.* 167 (3): 1041–1046.

Puigbo, P., Bravo, I.G., and Garcia-Vallve, S. 2008. E-CAI: A novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinforma.* 9: 1-7.

Qazi, J. 2016. Banana bunchy top virus and the bunchy top disease. *J. Gen. Plant Pathol*. 82: 2–11.

Quito-Avila, D.F., Ibarra, M.A., Alvarez, R.A., Ratti, M.F., Espinoza, L., Cevallos-Cevallos, J.M., and Peralta, E.L. 2013. First report of banana bract mosaic virus in 'cavendish' banana in Ecuador. *Plant Dis*. 97: 1003.

Ram, R.D. and Summanwar, A.S. 1984. Colocasia esculenta (L) Schott: a reservoir of bunchy top disease of banana. *Curr. Sci*. 53: 145–146.

Rao, A. A., Qureshi, M. A., Khanzada, A. L., Khanzada, M. A., Rajput, M. A. and Rajput, K. I. 2002. Occurrence and incidence of banana bunchy top disease in Southern part of Sindh. *Pakistan J. Plant Pathol.* 1:74-75.

Revers, F. and Garcia, J. A. 2015. Molecular biology of potyviruses. *Adv. Virus Res*. 92: 101–199.

Riechmann, J.L., Lain, S., and Garcia, J.A. 1992. Highlights and prospects of potyvirus molecular biology. *J. Gen. Virol*. 73: 1–16.

Rodoni, B.C., Ahlawat, Y.S., Varma, A., Dale, J.L., and Harding, R.M. 1997. Identification and characterization of banana bract mosaic virus in India. *Plant Dis.* 81: 669–672.

Rodoni, B.C., Dale, J.L., and Harding, R.M. 1999. Characterization and expression of the coat protein-coding region of banana bract mosaic potyvirus, development of diagnostic assays and detection of the virus in banana plants from five countries in southeast Asia. *Arch. Virol*. 144: 1725–1737.

Roldão, A., Mellado, M.C.M., Castilho, L.R., Carrondo, M.J.T., and Alves, P.M. 2010. Virus-like particles in vaccine development. *Expert. Rev. Vaccines*. 9(10):1149-1176.

Roossinck, M.J. 2003. Plant RNA virus evolution. *Curr. Opin. Microbiol*. 6: 406–409.

Rosano, G.L. and Ceccarelli, E.A. 2014. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* 5: 1–17.

Rostami, A., Bashir, N.S., Pirniakan, P., Masoudi, N. 2014. Expression of Cucumber mosaic virus coat protein and its assembly into virus like particles. *Biotechnol. Health Sci*. 1(3): 180.

Sabharwal, P. 2017. Molecular insight in to the structure and function of Pepper vein banding virus encoded proteins and endocytic uptake pathway of virus like particles in to mammalian cells. Ph.D. thesis, Indian Institute of Science, Bangalore, 115p.

Sabharwal, P., Amritha, C.K., Sushmitha, C., Natraj, U., and Savithri, H.S. 2019. Intracellular trafficking and endocytic uptake pathway of Pepper vein banding virus-like particles in epithelial cells. *Nanomedicine*. 14: 1247–1265.

Sabharwal, P., Sushmitha, C., Amritha, C.K., Natraj, U., Murthy, M.R.N., and Savithri, H.S. 2020. Development of pepper vein banding virus chimeric virus-like particles for potential diagnostic and therapeutic applications. *Arch. Virol.* 165: 1163–1176.

Sairam, S., Selvarajan, R., Handanahalli, S., and Venkataraman, S. 2020. Towards understanding the structure of the capsid of Banana Bunchy Top Virus. *BioRXIV*. [accessed on 2.2.2021]

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

Salem, R., Arif, I.A., Salama, M., and Osman, G.E.H. 2018. Polyclonal antibodies against the recombinantly expressed coat protein of the Citrus psorosis virus. *Saudi J. Biol. Sci*. 25: 733–738.

Sankaralingam, A., Baranwal, V., Ahlawat, Y., Devi, R., and Ramiah, M. 2006. RT-PCR detection and molecular characterization of Banana bract mosaic virus from the pseudostem and bract of banana. *Arch. Phytopathol. Plant Prot.* 39: 273–281.

Savithri, H.S. and Murthy, M.R.N. 2010. Structure and assembly of Sesbania mosaic virus. *Curr. Sci*. 98: 346–351.

Schägger, H. 2006. Tricine-SDS-PAGE. *Nat. Protoc.* 1: 16–22.

Sebasigari, K. and Stover, R. H. 1988. Banana diseases and pests in East Africa. Report of a Survey.in November 1987. INIBAP, Montpellier, France.

Seetharam, R., Heeren, R.A.,Wong, E.Y., Braford, S.R., Klein, B.K., Aykent, S.,Kotts, C.E., Mathis, K.J., Bishop, B.F., and Jennings, M.J. 1988. Mistranslation in IGF-1 during over-expression of the protein in Escherichia coli using a synthetic gene containing low frequency codons. *Biochem. Biophys. Res. Commun*. 155: 518–523.

Sekiya, M.E., Lawrence, S.D., McCaffery, M., and Cline, K. 1991. Molecular cloning and nucleotide sequencing of the coat protein gene of citrus tristeza virus. *J. Gen. Virol*. 72: 1013–1020.

Selvarajan, R. and Balasubramanian, V. 2008. Banana viruses. In: Rao, G.P., Myrta, A., and Ling, K. (eds), Characterization, Diagnosis and Management of Plant Viruses (2nd Volume). Studium press LLC, Texas, USA, pp. 109–124.

Selvarajan, R., Balasubramanian, V., and Sathiamoorthy, S. 2006. Vector transmission of banana bract mosaic and banana streak viruses in India [abstract]. In: *Abstracts of XVI annual convention and international symposium on management of vector-borne viruses*; 7–10, February, 2006, Hyderbad. International Crops Research Institute for Semi-Arid Tropics, Hyderbad, 110p.

Selvarajan, R., Balasubramanian, V., Dayakar, S., Sathaimoorthy, S., and Ahlawat, Y.S. 2010a. Evaluation of immunological and molecular techniques for the detection of different isolates of Banana bunchy top virus in India. *Indian Phytopathol*. 63: 333–336.

Selvarajan, R., Balasubramanian, V., Dayakar, S., Sathaimoorthy, S., and Ahlawat, Y. S. 2004. Simultaneous detection of Indian isolates of banana bunchy top virus, banana streak virus and banana bract mosaic virus by multiplex-PCR technique. In: Bandyopadhyay, B., Sarkar, S.K., Misra, D.K. and Bauri, F.K. (eds), *Banana Research in India.* Proceedings of national seminar on banana industry- present scenario and future strategies, FTC Kalyani, West Bengal. Bidhan Chandra Krishi Viswavidyala, Mohanpur, pp. 237–240.

Selvarajan, R., Balasubramanian, V., Priyanka, P., Jebakumar, R.M., Selvam, K.P., and Uma, S. 2020a. Evidence of seed transmission of Banana bract mosaic virus in Musa synthetic diploid H-201, a possible threat to banana breeding. *Eur. J. Plant Pathol.* 156: 1181–1187.

Selvarajan, R., Balasubramanian,V. and Gayathrie, T. 2016. Highly efficient immunodiagnosis of episomal banana streak MY virus using polyclonal antibodies raised against recombinant Viral-Associated Protein. *J. Phytopathol.* 164: 497–508.

Selvarajan, R., Jeyabaskaran, K.J. 2006. Effect of Banana bract mosaic virus (BBrMV) on growth and yield of cultivar Nendran (Plantain, AAB). *Indian Phytopathol.* 59: 496–500.

Selvarajan, R., Kanichelvam, P.S., Balasubramanian, V., and Sethurama Subramanian, S., 2020b. A rapid and sensitive lateral flow immunoassay (LFIA) test for the on-site detection of banana bract mosaic virus in banana plants. *J. Virol. Methods*. 284: 113929.

Selvarajan, R., Sheeba, M., Balasubramanian, V., Rajmohan, R., Dhevi, N.L., and Sasireka, T. 2010b. Molecular characterization of geographically different banana bunchy top virus isolates in India. *Indian J. Virol*. 21: 110–116.

Sharma, P., Sahu, A.K., Verma, R.K., Mishra, R., Choudhary, D.K., and Gaur, R.K. 2014. Current status of Potyvirus in India. *Arch. Phytopathol. Plant Prot*. 47: 906–918.

Sharma, S.K., Kumar, P.V., and Baranwal, V.K. 2014. Immunodiagnosis of episomal Banana streak MY virus using polyclonal antibodies to an expressed putative coat protein. *J. Virol. Methods*. 207: 86–94.

Sharman, M., Thomas, J.E., and Dietzgen, R.G. 2000. Development of a multiplex immunocapture PCR with colourimetric detection for viruses of banana. *J. Virol. Methods*. 89: 75–88.

Sharman, M., Thomas, J.E., Skabo, S., and Holton, T.A. 2008. Abacá bunchy top virus, a new member of the genus Babuvirus (family Nanoviridae). *Arch. Virol.* 153: 135–147.

Sharp, P.M. and Li, W.H. 1986. Codon usage in regulatory genes in Escherichia coli does not reflect selection for “rare” codons. *Nucleic Acids Res.* 14: 7737–7749.

Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H. and Wright. F. 1988. Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces

cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity. *Nucleic Acids Res*., 16:8207-8211.

Sharp, P.M., Tuohy, T.M.F. and Mosurski, K.R. 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res*. 14 (13): 5125-5143.

Shibaei, N., Majidi, J., Sokhandan Bashir, N., Karkhaneh, A. A., and Razavi, K. 2018. Production and partial purification of the Grapevine fanleaf virus coat protein 42 polyclonal antibody against inclusion body expressed in *Escherichia coli*. *Iranian J. Biotechnol.* 16(4): 303-310.

Shilpa, B.S., Moger, N., Vani, A., Rabinal, C., Hedge, R.V., and Byadagi, A. S. 2016. Cloning and Expression of Banana Bunchy Top Virus (BBTV) Coat Protein Gene in *E. coli*. *Elixir Agric.* 94: 40000-40004.

Shirbaghaee, Z. and Bolhassani, A. 2016. Different applications of virus-like particles in biology and medicine: Vaccination and delivery systems. *Biopolymers*, 105(3):113-32.

Shukla, D.D. and Ward, C.W. 1989. Identification and classification of potyviruses on the basis of coat protein sequence data and serology. Arch. Virol. 106: 171-200.

Shukla, D.D., Gough, K.H., and Ward C.W. 1987. Coat protein of potyviruses. 3. Comparison of amino acid sequences of the coat proteins of four Australian strains of sugarcane mosaic virus. *Arch. Virol.* 96: 59-74.

Shukla, D.D., Strike, P.M., Tracy, S.L., Gough, K.H., and Ward, C.W. 1988. The N and C Termini of the Coat Proteins of Potyviruses Are Surface-located and the N Terminus Contains the Major Virus-specific Epitopes. *J. Gen. Virol.* 69: 1497–1508.

Shukla, D.D., Tribbick, G., Mason, T.J., Hewish, D.R., Geysen, H.M. and Ward, C.W. 1989. Localization of virus-specific and group-specific epitopes of plant potyviruses by systematic immunochemical analysis of overlapping peptide fragments. *Proc. Natl. Acad. Sci. USA.* 86(21):8192-6.

Siljo, A., Bhat, A.I., and Biju, C.N. 2014. Detection of Cardamom mosaic virus and Banana bract mosaic virus in cardamom using SYBR Green based reverse transcription-quantitative PCR. *Virus Dis.* 25: 137–141.

Siljo, A., Bhat, A.I., Biju, C.N., and Venugopal, M.N. 2012. Occurrence of Banana bract mosaic virus on cardamom. *Phytoparasitica.* 40: 77–85.

Simmonds, J.H. 1935. Bunchy top of banana and its control. *Qd Agric. J.* 41(3): 241-244.

Simon-Loriere, E. and Holmes, E.C. 2011. Populations of RNA viruses habitually harbour abundant. *Nat. Publ. Gr*. https://doi.org/10.1038/nrmicro2614

Singh, S.J. 2003. *Viral Diseases of Banana*. Kalyani Publisher, New Delhi, pp 62-79.

Smith, G.R., Ford, R., Bryant, J.D., Gambley, R.L., McGhie, T.K., Harding, R.M., and Dale, J.L. 1995. Expression, purification, and use as an antigen of recombinant sugarcane mosaic virus coat protein. *Arch. Virol*. 140: 1817–1831.

Sochor, J., Babula, P., Adam, V., Krska, B., and Kizek, R. 2012. Sharka: the past, the present and the future. *Viruses.* 4: 2853–2901.

Sorensen, H.P. and Mortensen, K.K. 2005a. Advanced genetic strategies for recombinant protein expression in Escherichia coli. *J. Biotechnol*. 115: 113–128.

Sorensen, H.P. and Mortensen, K.K., 2005b. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. Microb. *Cell Fact.* 4: 1–8.

Sorensen, H.P., Sperling-Petersen, H.U., and Mortensen KK 2003. Pro- duction of recombinant thermostable proteins expressed in Escherichia coli: completion of protein synthesis is the the bottle- neck. *J. Chromatogr. B.* 786: 207–214.

Stainton, D., Kraberger, S., Walters, M., Wiltshire, E. J., Rosario, K., and Halafihi, M., 2012. Evidence of inter-component recombination, intra-component recombination and reassortment in banana bunchy top virus. *J. Gen.Virol*. 93: 1103–1119.

Stanley, W. 1935. Isolation of a crystalline protein possessing the properties of tobacco-mosaic virus. *Sci*. 81: 644-645.

Steinhauer, D. A., Domingo, E., and Holland, J. J. 1992. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene*. 122 (2): 281–288.

Studier, F.W. and Moffatt, B.A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189: 113–130.

Szathmary, E. 1992. Viral sex, levels of selection, and the origin of life. *J. Theor. Biol.* 159: 99–109.

Tajima, F. 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genet.* 123 (3): 585-595.

Tamura K., Nei M., and Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Nat. Acad. Sci. USA.* 101: 11030-11035.

Taylor, A.L. 1969. The Fiji Banana. *Proc. Helminthol. Soc. Wash*. 36(2): 157-163.

Templeton, A.R. 1980. The theory of speciation via the founder principle. Genet. 94 (4): 1011–1038.

Thangavelu, R., Selvarajan, R., and Singh, H. P. 2000. Status of banana streak virus and banana bract mosaic virus diseases in India. In: Singh, H.P. and Chadha, K.L. (eds), *Banana: Improvement, Production and Utilization*. Proceedings of the conference on challenges for banana production and utilization in 21st century, Trichy, India. AIPUB, NRCB, pp. 364–376.

Thiribhuvanamala, G., Doraiswamy, S., and Ganapathy, T. 2005. Detection of banana bunchy top virus in the aphid vector using double antibody sandwich (DAS) ELISA. *Indian J. Virol.* 16: 12–14.

Thomas, J.E. and Dietzgen, R.G. 1991. Purification, characterization and serological detection of virus-like particles associated with banana bunchy top disease in Australia. *J. Gen. Virol.* 72: 217–224.

Thomas, J.E. and Iskra-Caruana, M.L. 2000. Bunchy top. In: Jones, R.D. (ed.), Diseases of banana, abaca and ensete. CAB International, Wallingford, pp 241–253.

Thomas, J.E., Geering, A.D.W., Dahal, G., Lockhart, B.E.L., and Thottappilly, G. 2003. Banana and plaintain. In: Loebenstein, G. and Thottappilly, G. (eds), Virus and

virus-like diseases of major crops in developing countries. Kluwer, Dordrecht, pp 477–496.

Thomas, J.E., Geering, A.D.W., Gambley, C.F., Kessling, A.F., and White, M. 1997. Purification, properties, and diagnosis of banana bract mosaic potyvirus and its distinction from abaca mosaic potyvirus. *Phytopathol.* 87: 698–705.

Thomas, J.E., Iskra-Caruana, M.L., and Jones, D.R. 1994. Banana bunchy top disease. Musa Disease Fact Sheet No. 4. International Network for the Improvement of banana and plantain (INIBAP). Montpellier,pp 2

Thomas, J.E., Magnaye, L.V. 1996. Banana bract mosaic disease. Musa Disease Fact Sheet - 1. INIBAP, Montpellier, France.

Tian, Y.P., Liu, J.L., Zhang, C.L., Liu, Y.Y., Wang, B., Li, X.D., Guo, Z.K., and Valkonen, J.P.T. 2011. Genetic diversity of Potato virus Y infecting tobacco crops in China. *Phytopathol*. 101: 377–387.

Tillett, D. and Neilan, B. A. 1999. Enzyme-free cloning: a rapid method to clone PCR products independent of vector restriction enzyme sites. *Nucleic Acids Res*. 27: e26.

Tomimura, K., Spak, J., Katis, N., Jenner, C.E., Walsh, J.A., Gibbs, A.J., and Ohshima, K. 2004. Comparisons of the genetic structure of populations of Turnip mosaic virus in West and East Eurasia. *Virol.* 330: 408–423.

Towbin, H., Staehelin, T. and Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA*, 76(9):4350-4354.

Tripathi, S., Patil, B. L., and Verma, R. 2016. Viral Diseases of Banana and Their Management. In: Gaur, R. et al. (eds), *Plant Viruses: Evolution and Management*. Springer, Singapore, pp. 289-308.

Tsao, T. 2008. Towards the development of transgenic banana bunchy top virus (BBTV)-resistant banana plants: interference with replication Ph.D thesis, Centre for Tropical Crops and Biocommodities, Queensland, 198p.

Unger, T., Jacobovitch, Y., Dantes, A., Bernheim, R., and Peleg, Y. 2010. Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *J. Struct. Biol*. 172: 34–44.

Urcuqui-Inchima, S., Haenni, A.L., and Bernardi, F. 2001. Potyvirus proteins: A wealth of functions. *Virus Res*. 74: 157–175.

Vaira A.M., Vecchiati M., Masenga V., and Accotto G.P. 1996. A polyclonal antiserum against a recombinant viral protein combines specificity with versatility. *J. Virol. Meth.* 56: 209–219.

Vakili, N.G. 1969. Bunchy top disease of bananas in the Central Highlands of South Vietnam. *Plant Dis. Rep.* 53: 634-638.

Valekunja, R.B., Kamakoti, V., Peter, A., Phadnis, S., Prasad, S., and Nagaraj, V.J. 2016. The detection of papaya ringspot virus coat protein using an electrochemical immunosensor. *Anal. Methods.* 8*(*48): 8534-8541.

van der Walt, E., Rybicki, E.P., Varsani, A., Polston, J.E., Billharz, R., Donaldson, L., Monjane, A.L. and Martin, D.P. 2009 Rapid host adaptation by extensive recombination. *J. Gen. Virol.* 90: 734–46.

Vedadi, M., Arrowsmith, C.H., Allali-Hassani, A., Senisterra, G., and Wasney, G.A. 2010. Biophysical characterization of recombinant proteins: A key to higher structural genomics success. *J. Struct. Biol*. 172: 107–119.

Vetten, H.J., Chu, P.W.G., Dale, J.L Harding, R., Hu, J., Katul, L., Kojima, M., Randles, J.W., Sano, Y. and Thomas, J. E. 2005. Nanoviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (eds), *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. Virus taxonomy. Nineth report of the International Com- mittee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, 343–352pp.

Vishnoi, R., Raj, S.K., and Prasad, V. 2009. Molecular characterization of an Indian isolate of Banana bunchy top virus based on six genomic DNA components. *Virus Genes.* 38: 334–344.

Walker, J.M. 2009. *The Protein Protocols Handbook* (3rd Ed.). Springer-Verlag, New York, 1974p.

Walls, D. and Loughran, S.T. 2011. Tagging recombinant proteins to enhance solubility and aid purification. *Methods Mol. Biol.* 681:151-75.

Wang, I.C., Sether, D.M., Melzer, M.J., Borth, W.B., and Hu, J.S. 2010. First report of banana bract mosaic virus in flowering Ginger in Hawaii. *Plant Dis.* 94(7): 921.

Wang, Y., Gong, G.H., Wei, C.X., Liang, L., and Zhang, B. 2014. Recombinant expressed vector pET32a (+) S constructed by ligation independent cloning. *Molecules*. 19: 16179–16189.

Wanitchakorn, R., Hafner, G.J., Harding, R.M., and Dale, J.L. 2000a. Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *J. Gen. Virol.* 81: 299–306.

Wanitchakorn, R., Harding, R.M., and Dale, J.L. 2000b. Sequence variability in the coat protein gene of two groups of banana bunchy top isolates. *Arch. Virol.* 145: 593–602.

Wanitchakorn, R., Harding, R.M., Dale, J.L. 1997. Banana bunchy top virus DNA-3 encodes the viral coat protein. *Arch. Virol.* 142: 1673–1680.

Wardlaw, C.W. 1961. Banana diseases including plantains and abaca. Longman, Green, and Co. Ltd., London, UK, 648 pp.

Wardlaw, C.W. 1972. Banana Diseases, Including Plantains and Abaca, William Clowes and Sons Ltd. London, Beccles and Colchester, pp. 68–11.

Waugh DS. 2005. Making the most of affinity tags. *Trends Biotechnol*. 23: 316–320.

Wetzel, T., Candresse, T., Macquaire, G., Ravelonandro, M., Dunez, J. 1992. A highly sensitive immunocapture polymerase chain reaction method for plum pox potyvirus detection. *J Virol Methods*. 39: 27–37.

Wickramaarachchi, W.A.R.T., Shankarappa, K.S., Rangaswamy, K.T., Maruthi, M.N., Rajapakse, R.G.A.S., and Ghosh, S. 2016. Molecular characterization of banana bunchy top virus isolate from Sri Lanka and its genetic relationship with other isolates. *Virus Dis*. 27: 154–160.

Wiederstein, M. and Sippl, M.J. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. (Web Server issue):W407-W410.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall, W.B., Snoeyink, J., Adams, P.D., Lovell, S.C., Richardson, J.S. and Richardson, D.C. 2018. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* 27(1):293-315.

Wingfield, P.T. 2016. Overview of purification of recombinant proteins. *Curr. Protoc. Protein Sci.* 176: 100–106.

Wright, F. 1990. The effective number of codons used in a gene. *Gene*. 87: 23–29.

Wu, R.Y. and Su, H. J. 1990a. Purification and characterization of banana bunchy top virus. *J. Phytopathol*.128: 153–160.

Wu, R.Y. and Su, H.J. 1990b. Production of monoclonal antibodies against banana bunchy top virus and their use in enzyme-linked immunosorbent assay. *J. Phytopathol.*128: 203–208.

Xie, W.S. and Hu, J.S. 1995. Molecular cloning, sequence analysis and detection of banana bunchy top virus in Hawaii. *Phytopathol*. 85: 339–347.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. 2014. The I-TASSER suite: Protein structure and function prediction. *Nat. Methods*. 12: 7–8.

Yang, Z. and Nielsen, R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol*. 19: 908–917.

Yasin, H., Javed, M.A., Nazir, S., Bashir, R., and Khan, M.A. 2015. In-Vitro Propagation and Elisa Based Screening for the Development of Banana Bunchy Top Virus Free Plants. *ARPN J. Agric. Biol. Sci.* 10 (7): 274–279.

Yasmin, T., Khalid, S., Soomro, M. H., Malik, S. A., Shah, H., and Ahmad, I. 2001. Specificity of host-pathogen interaction of banana bunchy top disease. *J. Biol. Sci*. 1: 212-213.

Yates, C.M., Filippis, I., Kelley, L.A., and Sternberg, M.J.E. 2014. SuSPect: Enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J. Mol. Biol.* 426(14): 2692–2701.

Yu, N.T., Zhang, Y.L., Feng, T.C., Wang, J.H., Kulye, M., Yang, W.J., Lin, Z.S., Xiong, Z., and Liu, Z.X., 2012. Cloning and sequence analysis of two banana bunchy top virus genomes in Hainan. *Virus Genes.* 44: 488–494.

Yusibov, V., Kumar, A., North, A., Johnson, J.E., and Loesch-Fries, L.S. 1996. Purification, characterization, assembly and crystallization of assembled alfalfa mosaic virus coat protein expressed in Escherichia coli. *J. Gen. Virol.* 77: 567–573.

Zamora, M., Mendez-Lopez, E., Agirrezabala, X., Cuesta, R., Lavín, J.L., Amelia Sanchez-Pina, M., Aranda, M.A., and Valle, M. 2017. Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses. *Sci. Adv.* 3: 2–8.

Zhang, J., Borth, W.B., Lin, B., Dey, K.K., Melzer, M.J., Shen, H., Pu, X., Sun, D., and Hu, J.S. 2016. Deep sequencing of banana bract mosaic virus from flowering ginger (Alpinia purpurata) and development of an immunocapture RT-LAMP detection assay. *Arch. Virol.* 161: 1783–1795.

Zhang, Y., Olsen, D.R., Nguyen, K.B., Olson, P.S., Rhodes, E.T., and Mascarenhas, D. 1998. Expressoin of eukaryotic proteins in soluble form in Escherichia coli. *Protein Expr. Purif.* 12: 159–165.

Appendix I

Model of questionnaire for the study "Development of recombinant coat protein for immunodiagnosis of banana bunchy top and bract mosaic diseases"

Name & Address :

Location details :
Weather parameters :
Stage of the crop :
Propagation :

Varieties :

Disease & symptoms

Leaf lamina & petiole :

Midrib :

Pseudostem :

Bract :

Bunches :

Variability if any :

Vector : ☐ Present ☐ Absent Stage: Nymph ☐ Adul ☐

Observations

Similar symptoms observed before in the field ☐ Y ☐ N

Month/ Season:

Yield reduction: ☐ Y ☐ N

Management strategies employed

Preventive measures taken, if any

Signature

Appendix II

1. Preparation of Yeast extract Tryptone broth

Yeast extract - 12 g

Tryptone - 6 g

Dissolved in 500 ml water

2. Terrific Broth

TB solution A was prepared by adding 24 g yeast extract, 12 g tryptone and 4 g glycerol to 900 ml water. It was autoclaved at 121 ºC for 15 min.

TB solution B was prepared by adding 2.3 g potassium dihydrogen phosphate (KH_2PO_4) and 16.4 g dipotassium hydrogen phosphate (K_2HPO_4) in 100 ml water. It was sterilised by 0.22 µm syringe filter and added to TB-A under aseptic condition while inoculating culture.

3. SDS PAGE

Acrylamide solution (30 %)

Acrylamide - 29 g

Bis acrylamide - 1 g

Dissolved in 100 ml water

Sealing gel was prepared by mixing 6 ml water and 6 ml acrylamide solution and 200 µl APS and 15 µl TEMED

Composition of stacking and resolving gel

Components	**Big gel**	**Small gel**
Stacking gel		
Acrylamide solution	1.17 ml	67 µl
Sterile distilled water	4.7 ml	2.7 ml
Tris (pH 6.8)	875.0 µl	0.5 ml
10% SDS	70 µl	40 µl

10% APS	70 µl	40 µl
TEMED	5 µl	10 µl
Resolving gel		
Acrylamide solution	10.0 ml	4.25 ml
Sterile distilled water	8.2 ml	3.5 ml
Tris (pH 8.8)	6.3 ml	2.5 ml
10% SDS	250 µl	150 µl
10% APS	250 µl	150 µl
TEMED	20 µl	30 µl

5X running buffer was prepared by adding 15.1 g Tris 94 g glycine and 1 % SDS in 1 l water. pH was adjusted to 8.3.

4. Tricine SDS PAGE

Acrylamide solution

Acrylamide - 48 g

Bis acrylamide - 1.5 g

Dissolved in 100 ml water

10X Anode buffer, pH 8.9

1 M Tris

Dissolved in 1 l water

10X Cathode buffer, pH 8.25

1 M Tris

1 M Tricine

1 % SDS

Dissolved in 1 l water

3X Gel buffer, pH 8.45

3 M Tris

0.3 % SDS

Dissolved in 100 ml water

pH of the buffers was adjusted by concentrated HCl

Composition of stacking and resolving gel

Components	**Stacking gel (4 %)**	**Resolving gel (10 %)**
Acrylamide solution	1 ml	6 ml
3X Gel buffer	3 ml	10 ml
Glycerol (80 %)	-	0.6 ml
Water	7.4 ml	13.4 ml
APS	90	100
TEMED	9	10
Total volume	~12 ml	~30 ml

5. **Transfer buffer**

   5X Tris glycine buffer - 20 ml

   20 % methanol - 20 ml

   Volume was made upto 100 ml using water

6. **Staining solution**

   Methanol - 40 ml

   Acetic acid - 10 ml

   CBB R- 250 - 0.1 g

   Total volume was made up to 100 ml using distilled water

7. **Destaining solution**

   Methanol - 40 ml

   Acetic acid - 10 ml

   Total volume was made up to 100 ml using distilled water

8. **Reagents and solutions for sample preparation for mass spectroscopy**

   Destaining solution

   50 % v/v acetonitrile (ACN) - 2.5 ml

   50 % 25 mM ammonium bicarbonate (NH_4HCO_3; ABC) - 2.5 ml

Prepare 50mM ABC and dilute it with Acetonitrile in 1:1 ratio to obtain 25 mM ABC

Stock solution of 1 M Dithiothreitol (DTT) and Iodoacetamide was prepared in 50 mM ABC. Required concentration of the reagents were prepared in 50 mM ABC.

Gel extraction solution (10 ml)

50 % v/v ACN	- 5 ml
1 % v/v trifluoroacetic acid	- 0.1 ml
49 % water	- 4.9 ml

Appendix III

Sequence details of BBTV CP gene of Kerala isolates generated in the present study

>MT174314_KAN1_Nendran
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAAGCCAAATAACATAC
AACACGCTATGAAATACAAGACGCTATGACAAAAGTACGGGTATCTGATTAG
GTATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAA
TGTTTAATAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGA
GGTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGA
GTATATTTGTGTCGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGAT
GTCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGC
TCGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCG
TTTAGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGG
GAACCGTCCCAAATTCAAACTTCGATGGCTTGCCCTGCAAGCCATCTAGAAGT
CTATAAATACCAGTGTCTAGATAGTGATTCAGACAACAAATGGCTAGGTATC
CGAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCA
AGGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGA
AAACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGA
TATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAG
AATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACA
ACCTGTCTGGAACCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGT
TAAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCA
GATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAG
AGGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAAC
TACCAGAACAGAATTACATATCATGTTTGA

> MT174315_KKD1_Mysorepoovan
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAAGTACGGGTATCTGATTAGG
TATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGGGAGAG
TATATTTGTGTCGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTAGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TTTAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCCG
AAGAAATCCATCAAGAAGAGGCGGGTTGGGCGTCGGAAGTATGGCAGCAAG
GCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAAA
ACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGATAT
TTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGAAT
ACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAACCT
GTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTTAAA
CTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAGATG
TTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGAGGT
GGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACTACC
AGAACAGAATTACATATCATGTTTGA

>MT174316_WYD1_Nendran
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAATGTACGGGTATCTGATTAGG
TATCGTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACTAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGATACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATAACGGGCCCAAAGGGCCCGT
CTAGTAGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAATTTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAACCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

> MT174317_WYD2_Nendran
TATGTTTATGTAAACATAAACTTCTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAAGTACGGGTATCTGATTAGG
TATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTAGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAAAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGTAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTTGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

>MT174318_WYD3_Njalipoovan
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAAGTACGGGTATCTGATTAGG
TATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACTAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT

TTAGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACTGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTTTCTAGATAGTGATTCAGACAACAAATGGCTAGGTATCCG
AAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAAG
GCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAAA
ACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGATAT
TTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGAAT
ACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAACCT
GTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTTAAA
CTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAGATG
TTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGAGGT
GGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACTACC
AGAACAGAATTACATATCATGTTTGA

>MT174319_WYD4_Nendran
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAAGTACGGGTATCTGATTAGG
TATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTAGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCAACAA
CCTGTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

> MT174320_WYD5_Nendran
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAATGTACGGGTATCTGATTAGG
TATCGTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACTAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGATACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATAACGGGCCCAAAGGGCCCGT
CTAGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
ACCCGTCCCAATTTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAACCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT

AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

> MT174321_TCR1_Kunnan
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAAGTACGGGTATATGATTAGG
TATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTTGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTAGATGGGTTTTGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

> MT174322_TCR2_Palayangodan
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAAGTACGGGTATTTGATTAGG
TATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAAAAAACAAATATACATGATACGGATAGTTGAATACATAAACAACAAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGAAACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTAGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAAAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

> MT174323_TCR3_Nendran
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAATAACGGGTATCTGATTAGG
TATCGTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT

GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTTGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCCGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAACCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

> MT174324_TCR4_Chengalikodan
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACAAAC
AACACGCTATGAAATACAAGACGCTATGACAAAAGAACGGGTATCTGATTAG
GTATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAA
TGTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGA
GGTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGA
GTATATTTGTGTCGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGAT
GTCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGC
TCGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCG
TTTTGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGG
GAACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGT
CTATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATC
CGAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCA
AGGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGA
AAACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGA
TATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAG
AATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACA
ACCTGTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGT
TAAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCA
GATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAG
AGGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAAC
TACCAGAACAGAATTACATATCATGTTTGA

> MT174325_TCR5_Kasaka
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAAGAACGGGTATTTGATTAGG
TATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGAAACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTTGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAATTTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA

GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTCGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

> MT174326_PKD1_Nendran
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAAGAACGGGTATCTGATTAGG
TATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTTGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

>MT174327_PKD2_Nendran
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAAGAACGGGTATCTGATTAGG
TATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTTGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

>MT174328_PKD3_Nendran
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAAAGTACGGGTATCTGATTAGG
TATCCTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACGAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGACACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTAGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTAGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

>MT174329_KOT1_Nendran
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAATGTACGGGTATCTGATTAGG
TATCTTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATAAATAAACAACGAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATAACACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATATAGGGCCCAAAGGGCCCGT
TTAGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAAATTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAAGCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

>MT174329_KOL1_Nendran
TATGTTTATGTAAACATAAACTATTGTATGGAATGAAATCCAAATAACATACA
ACACGCTATGAAATACAAGACGCTATGACAAATGTACGGGTATCTGATTAGG
TATCGTAACGATCTAGGGCCGAAGGCCCGTGAGCAATATGCGTCGAAATAAT
GTTTAACAAACAAATATACATGATACGGATAGTTGAATACATAAACAACTAG
GTATACAATACAACAAACTGTTGTAAAGAAATAAAAAATAAGAAGAGAGAG
TATATTTGTGTCGGATAAGCATGATACCCACCACTTTAGTGGTGGGTCAGATG
TCCCGAGTTAGTGCGCCACGTAAGCGCTGGGGCTTATTATTACCCCCAGCGCT
CGGGACGGGACATGGGCTAATGGATTGTGGATAACGGGCCCAAAGGGCCCGT

CTAGATGGGTTTGGGCTTATGGGCTTTATCCAGAAGACCAAAAACAGGCGGG
AACCGTCCCAATTTCAAACTTCGATTGCTTGCCCTGCAAGCCATCTAGAAGTC
TATAAATACCAGTGTCTAGATAGTAGTTCAGACAACAAATGGCTAGGTATCC
GAAGAAATCCATCAAGAAGAGGCGGGTTGGGCGCCGGAAGTATGGCAGCAA
GGCGGCAACGAGCCACGACTACTCGTCGTTAGGGTCAATATTGGTTCCTGAA
AACACCGTCAAGGTATTTCGGATTGAGCCTACTGATAAAACATTACCCAGAT
ATTTTATCTGGAAAATGTTTATGCTTCTTGTGTGCAAGGTGAAGCCCGGAAGA
ATACTTCATTGGGCTATGATCAAGAGTTCTTGGGAAATCAACCAGCCGACAA
CCTGTCTGGAACCCCCAGGTTTATTTATTAAACCTGAACATAGCCATCTGGTT
AAACTGGTATGTAGTGGGGAACTTGAAGCAGGAGTCGCAACAGGGACATCAG
ATGTTGAATGTCTTTTGAGGAAGACAACCGTGTTGAGGAAGAATGTAACAGA
GGTGGATTATTTATATTTGGCATTCTATTGTAGTTCTGGAGTAAGTATAAACT
ACCAGAACAGAATTACATATCATGTTTGA

Sequence details of BBrMV CP gene of Kerala isolates generated in the present study

>MT818176_TVMBr5_Rasakadali
TCTGGAACCGGGAGTCAACCAACGATGATGACCAAGCGCACTATTGATGCAG
GTGGCAGTGTAAGAGGAACTCAACCATCAACAACCACTACAACAGCTCAAAG
CACATTTGGACAACCCACGACTATGTCAGCTCCATCTTCATCAAGCACGCCAC
CCAGAGCTTCTACGCAAATTGCCCCAGTTCGGGATCGTGACGTTGATGCTGGT
AGTACAAACTTCATCATCCCTCGAATTAAACCGATGACTGGGAAAATGCGTCT
TCCAAAGTATCGAGGAAAAACTGCAATTAACGTCGAGTTCTTGCTCCAATATA
AGCCTGATCAGTTTGATTTGTCGAATGCTATCGCAACTAGGGAGCAATATGAC
GCATGGTGTGATGCTGTAAAACGTGAGTATGCCATAGAGGACGAAGAACAGT
TCACAACCTTGTTAGGTGGTTTAATGGTATGGTGCATAGAGAATGGGACATCA
CCAAATTTGAATGGTACATGGAGTATGATGGATAAGGGTGAGCAATTAATTT
ACCAGTTAAAGCCTATTATTGAGAACGCTCAGCCTACTTTTCGGCAAATTATG
GCACATTTTTCTGATGCCGCTGAGGCATATATAACAATGCGCAATGTCACGGA
GAGATATATGCCCAGGTGGGGAGCACTTAGAGGATTGAATGACATAAGCTTA
GCCCGATATGCATTTGATTTTTACGTAGTCACATCAAAAACTACAAACAGGGC
TAGAGAAGCACACACGCAGATGAAAGCTGCAGCTATTCGTGGGTCAAACACT
CGGCTGTTCGGTTTGGATGGAAATCTTGGACCTGGCGAAGAGAATACAGAGA
GGCATACTGTTGAAGATGTGAAGCGTGATATGCACTCTCTGCTTGGGATGAA
ACATGAATAAATAAATAGTTATCTGGAGCTTGCTCCCTATAATTATGTGTGCT
TTATGATATTGTGATAATCGTAGTGTGAGCTTCTCACCTAAGTACCTACATGT
ATTGTGTGTGGTATTTGTATATCGCCTCTCGCC

>MT818177_EKMBr1_Nendran
TCTGGAACTGGAGTCATTCATGATGATGAATCCATCTATATTGATGCATGTGC
ATGTGCTAGGACTCAACTATCATTAATCATACATCAGTCTCATGCACATTTGG
ACACTCATCACTACGTCAGCTGCATCTCTCAGCACGTCAACCAGAGCTTCGAC
GCAAATTGCACAGTTCGGGATCGTGACGTTGACGCTGGTAGTCAAACTTTATT
CATCCCTCGAACTAAACCGATGACTTGGGATAAATGCGTCTCCAATGGTATCG
AGGGAAAACTGCAATTAATGTCGAGTTCTTGCTTCAATATAAGCCTGATCAGT
TTGATTTGTCAAGTGCTATCGCAATTTAGGGAGCAATATGATGCATGGTGTGA
TGCTGTAAAACGTGAGTGTGCCATAGAGGACGAAGAACAGTTCTCATCCTTG
TTAGGTGGTTTGATGGTATGGTGCATAGAGAATGGGTCATCTCCAAATTTAAA
TGGCACATGGAGTATGATGGACAAGGGTGAGCAGTTAGTTTACCAGTTAAAG
CCTATTATTGAGAACGCTCAGCCTACTTTTCGGCAAATTATGGCACATTTTTCT
GATGCTGCTGAGGCAAAAATAACAATGCGCAATGTCACGAAAAGATATATGC

CCAGGTGGGGAGCACTTAGAGGATTGAATGACATAAGCTTAGCCCGATATGC
ATTCGATTTTTACGTAGTCACATCAAAAACTACAAACAGGGCTAGAGAAGCA
CACACGCAGATGAAAGCTGCAGCTATTCGTGGGTCAACACTCGGTTATTCGGT
TTGGACGGAAATCTTGGACCTGATGAAGAGATCACAGAGAGGCATACTGAAT
GAAGATGTGAAGCGTGATATGCACTCTCTGCTGGATGAACATGATAATAATA
GTATCCGGAGCTGCTCCTATATAGTGGGCTTAAGAATGGATATTGAGTGAGCT
CTCACTAGTACTACTGAGTGAGTGCATTAATATCGCATATGCAAGGACCGCCT
GGGAAGGTATATGATCAA

>MT818178_TCRBr1_Kunnan
TCTGGAACGGAGTCAACCAACGATGATGACCCAAGCCGCACTATTGATGCAG
GTGGTAGTGCAAGAGGAACTCAACCATCAATAACCACCACAACAGCTCCGAG
CACATTTGAACAACCCACGACTATGTCAGCTCCATCTTCATCAAGCACGCCAC
CCAGAGCTTCCACGCAAATTGCACCAGTTCGGGATCGTGACGTTGATGCTGGT
AGTACAAACTTCATTATCCCTCGAATTAAACCGATGACTGGGAAAATGCGTCT
CCCAAGGTATCATGGAAAAACTGCAATTAATGTCGAGTTCTTGCTTCAATATA
AGCCCGATCAGTTTGATTTGTCAAATGCTATCGCAACTAGGGAGCAATATGAT
GCATGGTGTGATGCTGTCAAACGTGAGTATGCCATAGAGGACGAGGAACAGT
TCACAACTTTGTTAGGTGGTTTGATGGTATGGTGCATAGAGAATGGGACATCA
CCAAATTTGAATGGTACATGGAGTATGATGGATAAGGGTGAGCAATTAGTTT
ACCAGTTAAAGCCTATTATTGAGAACGCTCAGCCTACTTTTCGGCAAATTATG
GCACATTTTTCTGATGCTGCTGAGGCATACATAACAATGCGCAACGTCACGGA
AAGATATATGCCTAGGTGGGGAGCACTTAGAGGATTGAATGACATAAGCTTA
GCCCGATATGCATTTGATTTTTATGTAGTCACATCAAAAACTACAAACAGGGC
TAGAGAAGCACACACGCAGATGAAAGCTGCAGCTATTCGTGGGTCAAACACT
CGGTTATTCGGTTTGGATGGAAATCTTGGACCTGGTGAAGAGAACACAGAGA
GGCATACTGTTGAAGATGTGAAGCGTGATATGCACTCTCTGCTTGGGATGAA
ACATGAATAAATAAATAGTTATCTGGAGCTTGCTCCCTATAATTATGTGTGCT
TTATGATATTGTGATAATTGTAGTGTGAGCTTCTCACCTAAGTACCTACATGT
ATTGTGTGTGGTATTTATATTTTCACAATAAACAAAGGGACCGCCTGTGAGGT
ATTGATCGAG

>MT818179_TCRBr2_Rasakadali
TCTGGAACGGAGTCAACCAATGATGATGACCCAAGTCGCACTATTGATGCAG
GTCGCAGTGCTAGACGAACTCAACCATCAATAACCACTACAACAGCTCCAAG
CACATTTGGACAACCCACGACCACGTTAGCTCCATCTTCATCAAGTACGCCAC
CCAGAGCTTCCACGCAAATTGCACCAGTTCGGGATCGTGACGTTGACGCTGGT
AGTACAAACTTTATCATCCCTCGGATTAAACCGATGACTGGGAAAATGCGTCT
CCCAAGGTATCGAGGAAAAACTGCAATTAATGTCGAGTTCTTGCTTCAATATA
AGCCTGATCAGTTTGATTTGTCAAATGCTATCGCAACTAGGGAGCAATATGAT
GCATGGTGTGATGCTGTAAAACGCGAGTATGCCATGGAGGACGAAGAACAGT
TCACAACCTTGTTAGGTGGTTTGATGGTATGGTGTATAGAAAATGGAACATCA
CCAAACTTGAATGGTACATGGAGTATGATGGATAAGGGTGAGCAATTAGTTT
ACCAGTTAAAGCCTATTATTGAGAACGCTCAGCCTACTTTTCGGCAAATTATG
GCACATTTTTTTGATGCTGTTGAGGCATACATAACAATGCTCAATGTCACGGA
GAGGTATATTCCCAGGTGGGGAGCACTTAGAGGATTGAATGACATAAGCTTA
TCCCGATATGCATTTGATTTTTACGTAGTCACATCAAAAATTACAAACAGGGC
TAGAGAAGCACACACGCAGATGAAATCTGCAGCTATTCGTGGGTCAAACACT
CGGTTATTCGGTTTGGATGGAAATCTTGGACCTGGTGAAGAGAATACAGAGA
GGCATACTGTTGAAGATGTGAGGCGTGATATGCACTCTCTGCTTGGTATGAAA
CATGAATAAATAAATAGTTATCTGGAGCTTGCTCCCTATAATTATGTGTGCGT
TATGATATTGTGATAATTGTAGTGTGAGCTTCTCACCTAAGTACCTACTATGT
ATTGTGTGTGGTATTTTATATATTACCTGTCA

>MT818180_TCRBr3_Nallabontha
TCTGGAACGGAGTCAACCAATGATGATGACCCAATCTGCCCTATTGATGCAG
GTGGCAGTGCAAGAGGAACTCAACCCTCAATAACCACTACAACAGCTCCAAG
TACATTTGGACAACCCACAAATATGTCAGCTCCATCATCATCAAGCACTCCAC
CCAGAGGTTCCACGCAAATTGCACCAGTTCGGGATCGTGACGTTGATGTTGGT
AGTACAAACTTCATCATCCCTCGAATTAAACCGATGACTGGGAAAATGCGTCT
CCCCAAAGGTATCGAGGAAAAACTGCAAATTAATGTCGAGTTCTTGCTTCAAT
ATAAGCCTGATCAGTTCGATTTGTCAAATGCTATTGCAACTAGGGAGCAATAT
GATGCATGGTGTGATGCTGTAAAACGTGAGTATGCCATAGAGGACGAAGAAC
AGTTCACAACCTTGTTAGGTGGTTTGATGGTATGGTGCATAGAGAATGGGAC
ATCACCAAATTTGAATGGTACATGGAGTATGATGGATAAGGGTGAGCAATTA
GTTTACCAGTTAAAGCCTATTATTGAAAACGCTCAACCTACTTTTCGGCAAAT
TATGGCACATTTTTCTGATGCTGCTGAGGCATACATAACAATGCGCAATGTCA
CGGAGAGATATATGCCCAGGTGGGGAGCACTTAGAGGATTGAATGACATAAG
CTTAGCCCGATATGCATTTGATTTTTACGTAGTCACATCAAAAACTACAAACA
GGGCTAGAGAAGCACACACGCAGATGAAAGCTGCAGCTATTCGTGGGTCAAA
CACTCGGTTATTCGGTTTGGATGGAAATCTTGGACCTGGTGAAGAGAACACA
GAGAGGCATACTGTTGAAGATGTGAAGCGTGATATGCACTCTCTGCTTGGGA
TGAAACATGAATAAATAAATAGTTATGTGGAGCTTGCTCCCTATAAGATGTGT
GCTTTATGATATGGTATAATGTAGTGTGAGCTTCTCACCTATGTAC

>MT818181_TCRBr5_Njanlipoovan
TCTGGAACGGGGAACACCAATGATGATGACCCAAGCCGCACTATTGATGCAG
CTGGCAGTGCAAGAGGAACTCAACCATCAATAACCACTACAACAGCTCCAAG
CACATTTGGACAACCAACGAATACGTCAGCTCCATCTTCATCAAGCACGCCAC
CCAGAGCTTCCACACAAATTGCACAAATTCGGGATCGTGACGTTGATGCTGGT
AGTACAAACTTCATTATCCCTCGAATTAAACCGATGACTGGGAAAATGCGTCT
CCCAAGGTATCATGGAAAAACTGCAATTAATGTCGAGTTCTTGCTTCAATATA
AGCCCGATCAGTTTGATTTGTCAAATGCTATCGCAACTAGGGAGCAATATGAT
GCATGGTGTGATGCTGTCAAACGTGAGTATGCCATAGAGGACGAGGAACAGT
TCACAACTTTGTTAGGTGGTTTGATGGTATGGTGCATAGAGAATGGGACATCA
CCAAATTTGAATGGTACATGGAGTATGATGGATAAGGGTGAGCAATTAGTTT
ACCAGTTAAAGCCTATTATTGAGAACGCTCAGCCTACTTTTCGGCAAATTATG
GCACATTTTTCTGATGCTGCTGAGGCATACATAACAATGCGCAACGTCACGGA
AAGATATATGCCTAGGTGGGGAGCACTTAGAGGATTGAATGACATAAGCTTA
GCCCGATATGCATTTGATTTTTATGTAGTCACATCAAAAACTACAAACAGGGC
TAGAGAAGCACACACGCAGATGAAAGCTGCAGCTATTCGTGGGTCAAACACT
CGGTTATTCGGTTTGGATGGAAATCTTGGACCTGGTGAAGAGAACACAGAGA
GGCATACTGTTGAAGATGTGAAGCGTGATATGCACTCTCTGCTTGGGATGAA
ACATGAATAAATAAATAGTTATCTGGAGCTTGCTCCCTATAATTATGTGTGCT
TTATGATATTGTGATAATTGTAGTGTGAGCTTCTCACCTAAGTACCTACATGT
ATTGTGTGTGGTAGTTGATATTTTCACAATAAACAAAGGGACGCAGTAGATTT
TGATCACTA

>MT818182_PKDBr1_Nendran
TCGGGAACGGAGATCACCATGATTGATGACCCAAGCCGAACTATTGATGCAG
GTGGCAGTGCAAGAGGAACTCAACCATCAATAACCACTACAACAGCTCCAAG
CACTTTTGGACAACCCACGACTACGTCAGCTCCATCTTCATCAAGCACGCCAC
CCAGAGCTTCCACGCAAATTACACCAGTCCGGGATCGTGACGTTGATGCTGGT
AGTACAAACTTTATCATCCCCCGAATTAAACCGATGACTGGGAAAATGCGTCT
CCCAAGGTATCGTGGAAAAACTGCAATTAACGTCGAGTTCTTGCTCCAATATA
AGCCTGATCAGTTTGACCTGTCAAACGCTATCGCAACTAGGGAGCAATATGA
TGCATGGTGTGATGCTGTAAAACGTGAGTATGCCATAGAGGACGAAGAACAG
TTCACAACCTTGTTAGGTGGTTTGATGGTATGGTGCATAGAGAATGGGACATC
ACCAAATTTGAATGGTACATGGAGTATGATGGATAAGGGTGAGCAATTAATT

TACCAGTTAAAGCCTATTATTGAGAACGCTCAGCCTACTTTTCGGCAAATCAT
GGCACATTTTTCTGATGCTGCTGAGGCATACATAACAATGCGCAATGTCACGG
AGAGATATATGCCCAGGTGGGGAGCACTTAGAGGATTGAATGACATAAGCTT
AGCCCGATATGCATTTGATTTTTACGTAGTCACATCAAAAACTACAAACAGGG
CTAGAGAAGCACACACGCAGATGAAAGCTGCAGCTATTCGTGGGTCAAACAC
TCGGTTATTCGGTTTGGATGGAAATCTTGGACCTGGTGAAGAGAATACAGAG
AGGCATACTGTTGATGATGTGAAGCGTGATATGCACTCTCTGCTTGGGATGAA
ACATGAATAGATAAATAGTTATCTGGGGCTTGCTCCTTATAATTATGTGTGCT
TTATGATATTGTGATTATTGTAGTGTGAGCTTCTCACCTAAGTACCTACATGTA
TTGTGTGTGGTATGTATATATTCGCATATATAAGGGACCGCCTGTGAGAGTGG
ATCGAC

>MT818183_KKDBr12_Mysorepoovan
TCTGGAACCGAGAGTCAATCAACGATGATGACCCAAGCTGCACTAATGATGC
AGGTGGTAGTGCAAGAAGAACTCAACCATCAACAAACCACTACAACAGCTCC
AAGCACATTTGGACAACCCACAACTACGTCAGCTCCATCTTCATCAAGCACGC
CACCCAGAGCTTCCACGCAAATTGCACCAGTTCGGGATCGTGACGTTGATGCT
GGTAGTACAAACTTTACCATCCCTCGAATTAAACCGATGACTGGGAAAATGC
GTCTCCCAAGGTATCGTGGAAAAACTGCAATTAACGTCGAGTTCTTGCTTCAA
TATAAGCCTGATCAGTTTGATTTGTCAAATGCTATCGCAACTAGGGAGCAATA
TGATGCATGGTGTGATGCTGTAAAACGTGAGTATGCCATAGAAGACGAAGAA
CAGTTCACAACCTTGTTAGGTGGTTTGATGGTATGGTGCATAGAGAATGGGAC
ATCACCAAATTTGAATGGTACATGGAGTATGATGGATAAGGGTGAGCAATTA
GTTTACCAGTTAAAACCTATTATTGAGAACGCTCAGCCTACTTTTCGGCAAAT
TATGGCACATTTTTCTGATGCCGCTGAGGCATACATAACAATGCGCAATGTCA
CGGAGAGATATATGCCCAGGTGGGGAGCACTTAGAGGATTGAATGACATAAG
CTTAGCCCGATATGCATTTGATTTTTACGTAGTCACATCAAAAACTACAAACA
GGGCTAGAGAAGCACACACGCAGATGAAAGCTGCAGCTATTCGTGGGTCAAA
CACTCGGTTATTCGGTTTGGATGGAAATCTTGGACCTGGTGAAGAGAATACA
GAGAGGCATACTGTTGAAGATGTGAAGCGTGATATGCACTCTCTGCTTGGGA
TGAAACATGAATAAATAAATAGTTATCTGGAGCTTGCTCCCTATAATTATGTG
TGCTTTATAATATTGCGGTAATTGTAGTGTGAGCTTCTCACCTAAGTACCTAC
ATGTATCGTGTGTGGTATTTATATATTCGCAATATGCAAAGGGACCGCCTGTG
AGTATTG

> MT818184_KKDBr15_Nendran
TCTGGAACCGAGAGTCAATCAACGATGATGACCCAAGCTGCACTAATGATGC
AGGTGGTAGTGCAAGAAGAACTCAACCATCAACAAACCACTACAACAGCTCC
AAGCACATTTGGACAACCCACAACTACGTCAGCTCCATCTTCATCAAGCACGC
CACCCAGAGCTTCCACGCAAATTGCACCAGTTCGGGATCGTGACGTTGATGCT
GGTAGTACAAACTTTACCATCCCTCGAATTAAACCGATGACTGGGAAAATGC
GTCTCCCAAGGTATCGTGGAAAAACTGCAATTAACGTCGAGTTCTTGCTTCAA
TATAAGCCTGATCAGTTTGATTTGTCAAATGCTATCGCAACTAGGGAGCAATA
TGATGCATGGTGTGATGCTGTAAAACGTGAGTATGCCATAGAAGACGAAGAA
CAGTTCACAACCTTGTTAGGTGGTTTGATGGTATGGTGCATAGAGAATGGGAC
ATCACCAAATTTGAATGGTACATGGAGTATGATGGATAAGGGTGAGCAATTA
GTTTACCAGTTAAAACCTATTATTGAGAACGCTCAGCCTACTTTTCGGCAAAT
TATGGCACATTTTTCTGATGCCGCTGAGGCATACATAACAATGCGCAATGTCA
CGGAGAGATATATGCCCAGGTGGGGAGCACTTAGAGGATTGAATGACATAAG
CTTAGCCCGATATGCATTTGATTTTTACGTAGTCACATCAAAAACTACAAACA
GGGCTAGAGAAGCACACACGCAGATGAAAGCTGCAGCTATTCGTGGGTCAAA
CACTCGGTTATTCGGTTTGGATGGAAATCTTGGACCTGGTGAAGAGAATACA
GAGAGGCATACTGTTGAAGATGTGAAGCGTGATATGCACTCTCTGCTTGGGA
TGAAACATGAATAAATAAATAGTTATCTGGAGCTTGCTCCCTATAATTATGTG
TGCTTTATAATATTGCGGTAATTGTAGTGTGAGCTTCTCACCTAAGTACCTAC

ATGTATCGTGTGTGGTATTTATATATTCGCAATATGCAAAGGGACCGCCTGTG
AGTATTG

> MT818185_KKDBr19_Nendran
TCTGGAACGGAGTCAACCAATGATGATGACCCAAGCCGCACTATTGATGCAG
GTGGCAGTGCAAGAGGAACTCAACCATCAATAACCACTACAACAGCTCCAAG
CACATTTGGACAACCCACGACTACGTCAGCTCCATCTTCATCAAGCACGCCAC
CCAGAGCTTCCACACAAATTGCACCAATTCGGGATCGTGACGTTGATGCTGGT
AGTACAAACTTCATTATCCCTCGAATTAAACCGATGACTGGGAAAATGCGTCT
CCCAAGGTATCATGGAAAAACTGCAATTAATGTCGAGTTCTTGCTTCAATATA
AGCCCGATCAGTTTGATTTGTCAAATGCTATCGCAACTAGGGAGCAATATGAT
GCATGGTGTGATGCTGTCAAACGTGAGTATGCCATAGAGGACGAGGAACAGT
TCACAACTTTGTTAGGTGGTTTGATGGTATGGTGCATAGAGAATGGGACATCA
CCAAATTTGAATGGTACATGGAGTATGATGGATAAGGGTGAGCAATTAGTTT
ACCAGTTAAAGCCTATTATTGAGAACGCTCAGCCTACTTTTCGGCAAATTATG
GCACATTTTTTCTGATGCTGCTGAGGCATACATAACAATGCGCAACGTCACGGA
AAGATATATGCCTAGGTGGGGAGCACTTAGAGGATTGAATGACATAAGCTTA
GCCCGATATGCATTTGATTTTTATGTAGTCACATCAAAAACTACAAACAGGGC
TAGAGAAGCACACACGCAGATGAAAGCTGCAGCTATTCGTGGGTCAAACACT
CGGTTATTCGGTTTGGATGGAAATCTTGGACCTGGTGAAGAGAACACAGAGA
GGCATACTGTTGAAGATGTGAAGCGTGATATGCACTCTCTGCTTGGGATGAA
ACATGAATAAATAAATAGTTATCTGGAGCTTGCTCCCTATAATTATGTGTGCT
TTATGATATTGTGATAATTGTAGTGTGAGCTTCTCACCTAAGTACCTACATGT
ATTGTGTGTGGTATTTATATTTTCACAATAAACAAAGGGACCGCCTGTGAGGT
ATTGATCGAG

> MT818186_KANBr2_Nendran
TCTGGAACGGAGTCAACCAATGATGATGACCCAAGGCCGCACATTGATGCAG
GTGGCAGCGCAAGAGGAACTCAACCATCAATAACCACTACAACAGCCCCAAG
CACATTTGGACAACCCACGACCACGTCAGCTCCATCTTCATCAAGCACGCTAC
CCAGAGCCTCCACGCAAATTGCGCCAGTTCGGGATCGTGATGTTGATGCTGGT
AGTACAAATTTTATTATCCCTCGAATCAAATCAATGACAGGGAAAATGCGTCT
CCCAAGGTATCGGGGAAAAACTGCAATTAATGTTGAGTTCTTGCTTCAATACA
AGCCTGACCAGTTTGATTTGTCAAATGCTATCGCAACTAGGGAGCAATATGAT
GCATGGTGTGATGTCGTAAAACGTGAGTATGCCATAGAGGACGAAGAACAGT
TCACAACTTTGTTAGGTGGTCTGATGGTATGGTGCATAGAGAATGGGACATCA
CCAAATTTGAATGGTACATGGAGTATGATGGATAAGGGTGAGCAAATAGTTT
ACCAGCTAAAGCCTATTATTGAGAACGCTCAACCTACTTTTCGGCAGATTATG
GCACATTTTTTCTGATGCCGCTGAGGCATACATAACGATGCGCAATGCCACGG
AGAGATATATGACCACGTGGGGAGCACTCAAAGAATTGAATGACATAAGCTT
AGCCCGATATGCATTCGATTTTTATGTAGTCACATCAAAAACTACAAACAGGG
CTAGAGAAGCACACACGCAGATGAAAGCTGCAGCTATTCGTGGATCAAACAC
TCGTTTATTTGGTTTGGATGGAAATCTTGGACCTGGTGAAGAGAATACAGAGA
GGCATACTGTTGAAGATGTGAAGCGTGATATGCATTCTCTGCTTGGGATGAAA
CATGAATAAATAAATAGTTATCTGGAGCTTGCTCCCTATAATTATGTGTGCTT
TATGATATTGTGATAATTATAGTGTGAGCTTCTCACCTAAGTACCTACATGTA
TTGTGTGTGGTATTTATA

> MT818187_KASBr1_Nendran
TCGGGAACGGATTCATCCAATGATGATGATCTAAGCCGTGGCATTGATGCAG
GTGGCAGTGTTAGCGGAACTCAACCAGGAATCACCACTAATCACTACTACAG
CTCCAAGTACATTGGATCAATCACATCTACGTCAGCTGCATCTTCAATCAAGC
ACGGCATCCATGAGCTTTCATGCAAATTGCACCAGGTCGGGATTCGTGACGTT
GATGTTGTTAGTACAAACTTCATCATCCCTCGAATTATATCGATGACTGGGAA

AATGCGTCTCTCAAGGTATCGAGGAAAATCTGCAATTAATGTCGAGTTCTTGC
TTAATTATAAGCCTGATCAGTTCGATTTGTCAAATGCTATTGCAACTAGGGAG
CAATATGATGCATGGTGTGATGCTGTAAAACGTGAGTATGCCATAGAGGACG
AAGAACAGTTCACAACCTTGTTAGGTGGTTTGATGGTATGGTGCATAGAGAA
TGGGACATCACCAAATTTGAATGGTACATGGAGTATGATGGATAAGGGTGAG
CAATTAGTTTTCCAGTTAAAGCCTATTATTGAAAACGCTCAACCTTCTTTTCGG
CAAATTATGGCACATTTTTCTGATGCTGCTGAGGCATACATAACAATGCGCAA
TGTCACGGAGAGATATATGCCCAGGTGGGGAGCTCTTAGAGGTTTGAATGAT
ATAAGCTTAGCCCGATATGCATTTGATTTTTTTGTAGTCACATCAAAAACTAC
AAACAGGGCTAGAGAAGCACTCACGCAGATGAAAGCTGCAGCTATTCGTGGG
TCAAACACTCGGTTATTCGGTTTGGATGGAAATCATGGACCTGGTGAAGAGA
ACACAGAGAGGCACTCTGTCGAAGACAGAAGCGTGATAACCACTCTTATGTT
TCTGATGAAACATGAATAAATAAACAGTTATCTGGCGATCATACCTATATTTA
CGTGTGCTTTATGATATTGGGATAAATGGAGCAGACCTTCAAGCCTAAGTCCA
CCACAGGTATGATGTCGGTAC

Appendix 1V

Details of BBTV CP gene sequences retrieved from GenBank

Sl. No	Accession number	Location	Country
1	GU125413.1	Kerala	India
2	GU085261.1	Tamil Nadu	India
3	GU085260.1	Tamil Nadu	India
4	FJ664271.1	Tamil Nadu	India
5	JX171699.1	Coimbatore	India
6	GU125411.1	Karnataka	India
7	JN250599.1	Karnataka	India
8	GU125412.1	Andhra Pradesh	India
9	GU125409.1	Maharashtra	India
10	GU085262.1	Maharashtra	India
11	GU125408.1	Gujrat	India
12	DQ515970.2	Uttar Pradesh	India
13	GU125414.1	Bihar	India
14	FJ168538.1	Bihar	India
15	GU125407.1	Delhi	India
16	GU125406.1	West Bengal	India
17	EU190968.1	Arunachal Pradesh	India
18	EU190967.1	Meghalaya	India
19	EU190965.1	Assam	India
20	MN380801.1	Tripura	India
21	KT180290.1	Tripura	India
22	KC466374.1	Umiam	India
23	GU125410.1	Andaman	India
24	AF246122.1		China
25	AF238877.1		China
26	MF688998.1		China

27	MG545612.1		China
28	AF148944.1		Fiji
29	AB252644.1		Myanmar
30	JN250595.1		Sri Lanka
31	KT220180.1		Egypt
32	MF039868.1		Thailand
33	AB848112.1		Indonesia
34	AB848111.1		Indonesia
35	EF593169.1		Pakistan
36	AM418566.1		Pakistan
37	AB250956.1		Philippines
38	AB189068.1		Philippines
39	EU366171.1		Taiwan
40	DQ826393.1		Taiwan
41	AB078023.1		Japan
42	AF148945.1		Vietnam
43	KT923137.1		Brazil
44	JF755985.1		Nigeria
45	JF755983.1		Angola
46	JF755980.1		Malawi
47	AF148943.1	Burundi	Africa
48	JF755979.1		Cameroon

Details of BBrMV CP gene sequences retrieved from GenBank

Sl. No	Accession number	Location	Country	Host
1	AF071582.1	Kerala	India	*Musa sp.*
2	AF071583.1	Tamil Nadu	India	*Musa sp.*
3	AF071584.1	Tamil Nadu	India	*Musa sp.*
4	AF071585.1	Tamil Nadu	India	*Musa sp.*
5	AF071586.1	Coimbatore	India	*Musa sp.*
6	AF071587.1	Karnataka	India	*Musa sp.*
7	AF071588.1	Karnataka	India	*Musa sp.*
8	AF071589.1	Andhra Pradesh	India	*Musa sp.*
9	AF071590.1	Maharashtra	India	*Musa sp.*
10	KY369923.1	Maharashtra	India	*Musa sp.*
11	KF385470.1	Gujrat	India	*Musa sp.*
12	KF385473.1	Uttar Pradesh	India	*Musa sp.*
13	KF385474.1	Bihar	India	*Musa sp.*
14	KF385475.1	Bihar	India	*Musa sp.*
15	KF385476.1	Delhi	India	*Musa sp.*
16	KF385477.1	West Bengal	India	*Musa sp.*
17	KF385478.1	Arunachal Pradesh	India	*Musa sp.*
18	KF385479.1	Meghalaya	India	*Musa sp.*
19	KF385480.1	Assam	India	*Musa sp.*
20	KF385481.1	Tripura	India	*Musa sp.*
21	KF385482.1	Tripura	India	*Musa sp.*
22	KF385483.1	Umiam	India	*Musa sp.*
23	KF385484.1	Andaman	India	*Musa sp.*
24	KF385485.1		China	*Musa sp.*
25	KF385486.1		China	*Musa sp.*
26	KF385487.1		China	*Musa sp.*
27	KF385488.1		China	*Musa sp.*
28	KF385489.1		Fiji	*Musa sp.*

29	KF385490.1		Myanmar	*Musa sp.*
30	KF385491.1		Sri Lanka	*Musa sp.*
31	HM348780.1		Egypt	*Musa sp.*
32	HM348779.1		Thailand	*Musa sp.*
33	HM348782.1		Indonesia	*Musa sp.*
34	HM348781.1		Indonesia	*Musa sp.*
35	AY953427.1		Pakistan	*Musa sp.*
36	EF654655.1		Pakistan	*Musa sp.*
37	AY494979.1		Philippines	*Musa sp.*
38	EU414267.1		Philippines	*Musa sp.*
39	HQ709162.1	Kerala (Wayanad)	India	Cardamom
40	HQ709163.1	Kerala (Idukki)	India	Cardamom
41	HQ709164.1	Karnataka	India	Cardamom
42	HQ709165.1	Karnataka	India	Cardamom
43	HQ709166.1	Karnataka	India	Cardamom
44	KT456531.1		Hawaii	Flowering ginger

Appendix V

Vector maps

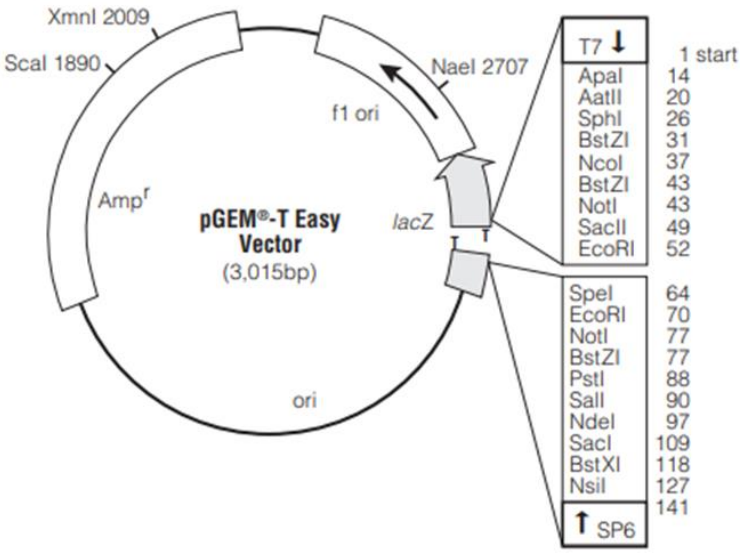


Fig I. Vector map of pGEM-T easy vector

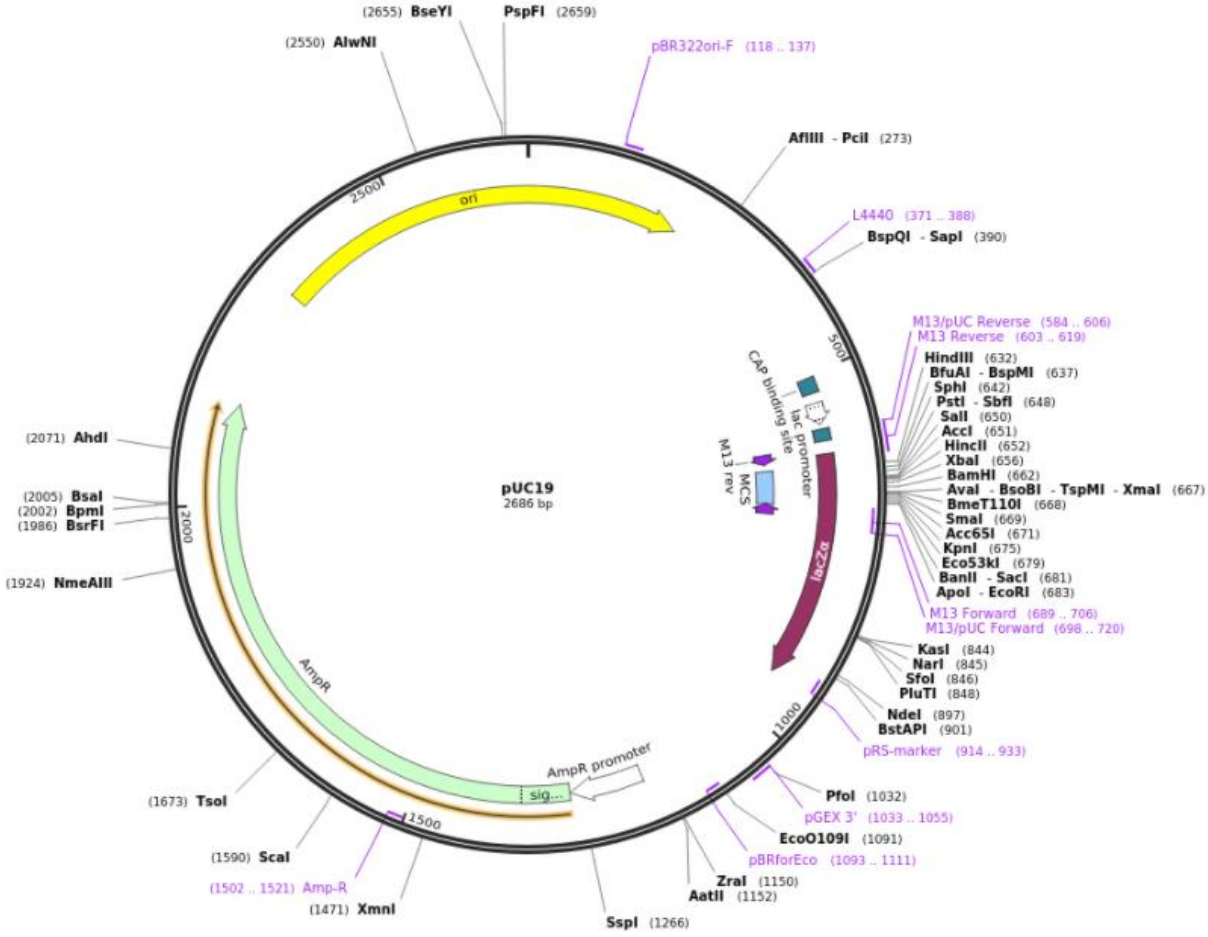


Fig II. Vector map of pUC19

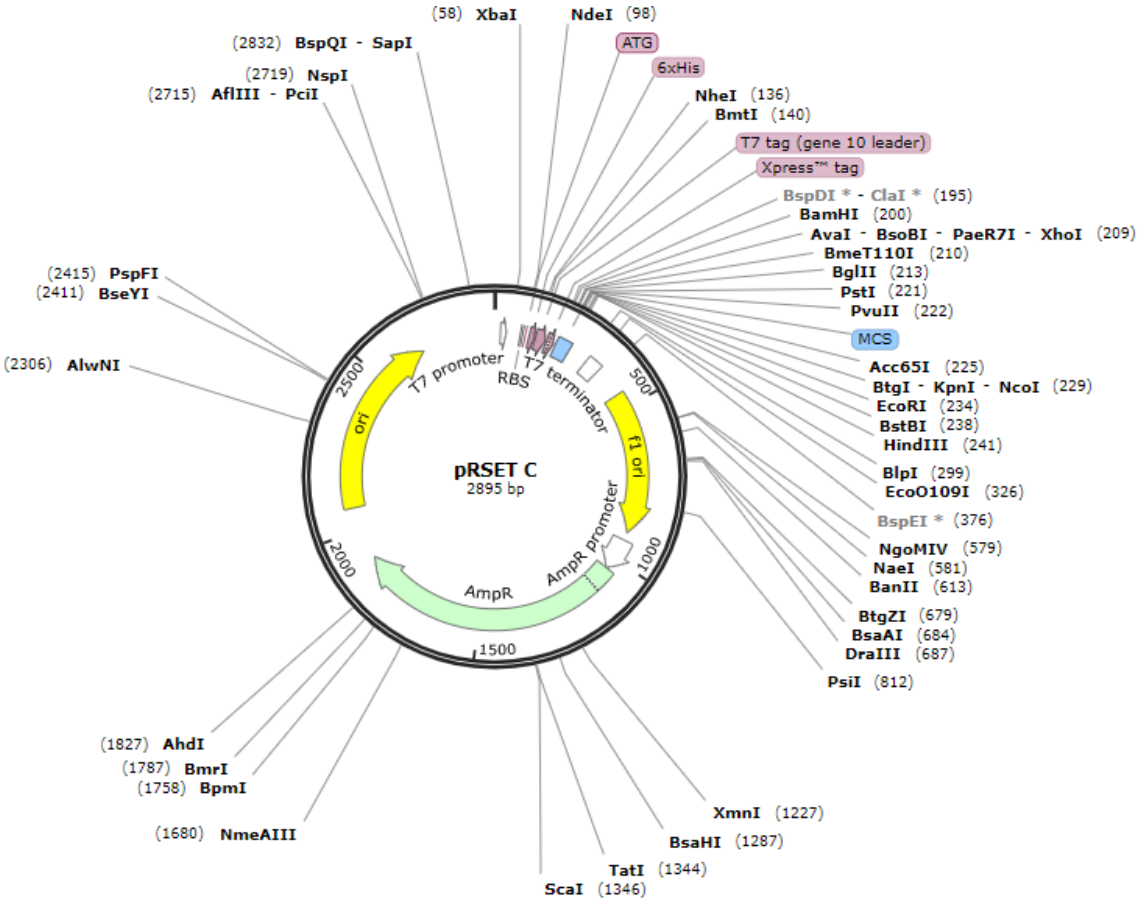


Fig III. Vector map of pRSET-C

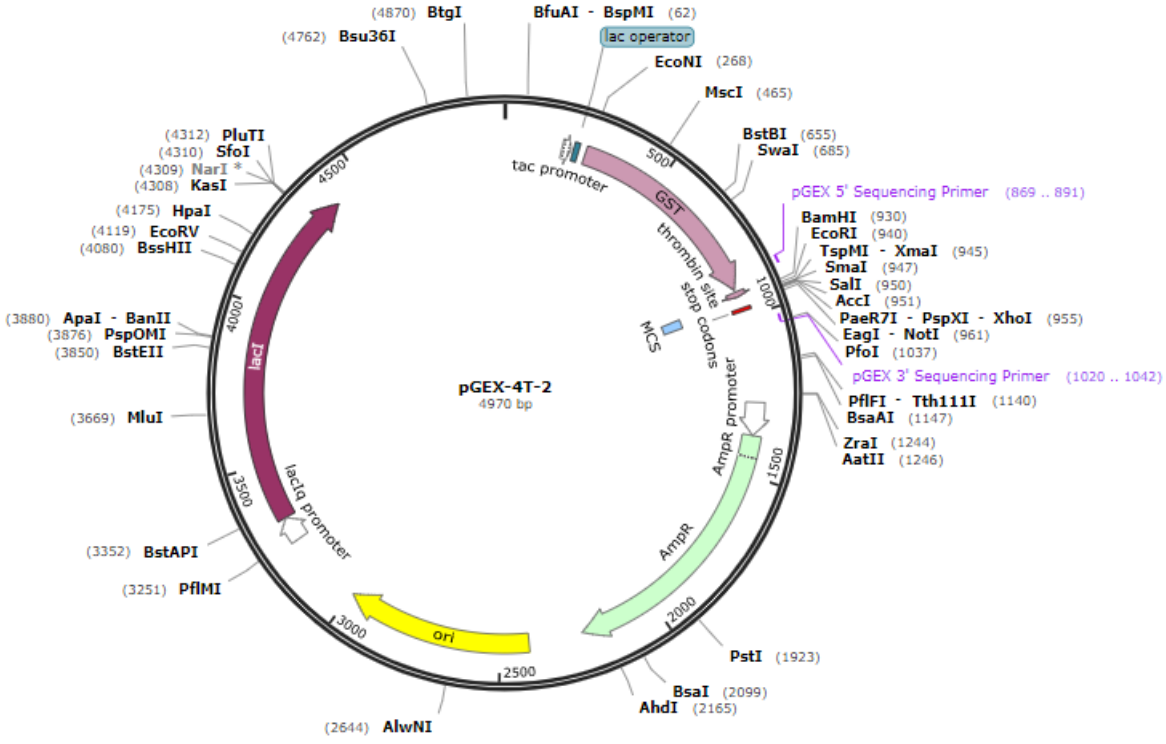


Fig IV. Vector map of pGEX-4T-2

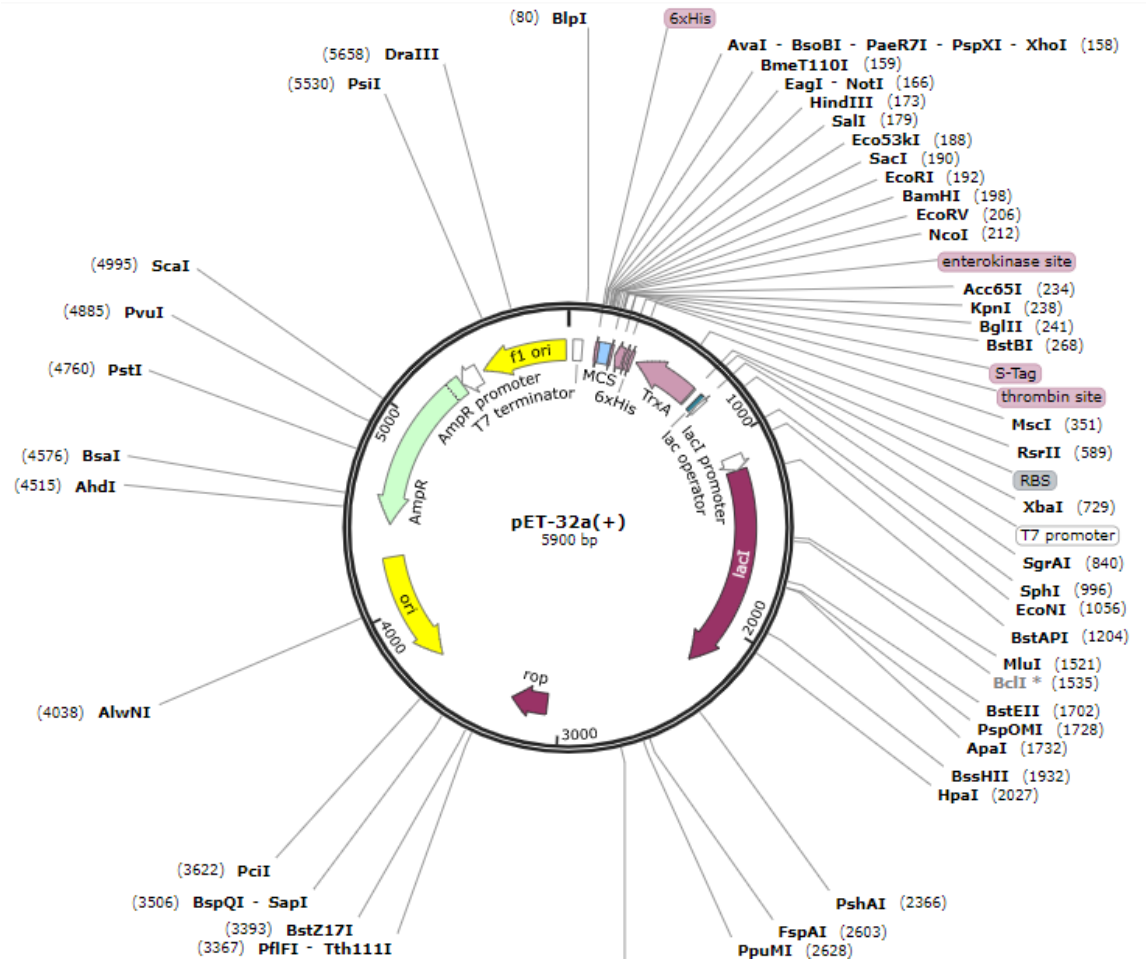


Fig V. Vector map of pET-32a(+)

DEVELOPMENT OF RECOMBINANT COAT PROTEIN FOR IMMUNODIAGNOSIS OF BANANA BUNCHY TOP AND BRACT MOSAIC DISEASES

by

DARSANA DILIP K.C.
2016-21-024

ABSTRACT OF THE THESIS

Submitted in partial fulfillment of the

requirement for the degree of

DOCTOR OF PHILOSOPHY IN AGRICULTURE

Faculty of Agriculture

Kerala Agricultural University

DEPARTMENT OF PLANT PATHOLOGY

COLLEGE OF AGRICULTURE

VELLANIKKARA, THRISSUR - 680 656

KERALA, INDIA

2021

ABSTRACT

The present investigation was undertaken to develop recombinant coat protein (rCP) of *Banana bunchy top virus* (BBTV) and *Banana bract mosaic virus* (BBrMV) for immunodetection of the viruses. The experiments were conducted at the Virology Lab, Banana Research Station, Kannara; Department of Plant Pathology, College of Agriculture, Vellanikkara, Kerala Agricultural University and Indian Institute of Science, Bengaluru during the period of 2016-2020.

A roving survey in 10 districts of Kerala, divided into population subsets *viz.*, North, Central and Southern zones were conducted for sample collection. After a preliminary DAC-ELISA, 17 and 12 representative samples respectively were selected and carried forward for further evaluations. The CP gene of BBTV was amplified from the total DNA isolated using reported primers by Polymerase Chain Reaction (PCR) and that of BBrMV by Reverse Transcriptase-PCR (RT-PCR). The CP gene sequences of these isolates were determined and submitted in the NCBI-GenBank Database. The 17 BBTV isolates were designated as MT174314-MT174330 and the 12 BBrMV isolates as MT818176-MT818187.

It was inevitable to evaluate the molecular diversity of the viruses prior to devising nucleic-acid based and serological detection methods. The phylogeographic analysis depicted a clear demarcation of BBTV Kerala isolates based on geography whereas no such clustering was observed in the case of BBrMV isolates. Being an RNA virus, the molecular diversity of BBrMV (ranging between 1-12 %) was higher than BBTV. However, the 5' and 3' terminal of BBrMV CP gene was hypervariable and found unsuitable to be targeted for nucleic-acid based detection. Hence, forward primer was designed from the NIb region of ssRNA genome of BBrMV and reverse primer from 3' UTR region upstream and downstream to the CP gene respectively. For nucleic-acid based detection of BBTV, highly conserved non-coding region of DNA-S upstream and downstream to the CP ORF was targeted. The primers were validated by detecting virus from the field samples collected from various parts of the state.

The rCPs were chosen as a potential antigen for raising antibodies in order to develop serodiagnostic assays for the early detection of the viruses. The BBTV CP gene was cloned

in to three expression vectors *viz.*, pRSET-C, pGEX-4T-2 and pET32a(+) and transformed to expression hosts like BL21 (DE3) pLysS, Rosetta (DE3) pLysS and C41 strains of *E. coli* after amplification in DH5α. The 20 kDa recombinant BBTV CP (rBBTV CP) cloned in to pRSET-C, and overexpressed in various *E. coli* hosts had a hexa histidine (6X His) tag at the N terminal. Similarly, a 37 kDa fusion protein (pET/rBBTV CP) was overexpressed from pET/BBTVCP clone had a thioredoxin (Trx) tag (17 kDa) along with the 6X His tag. Whereas, a 45 kDa fusion protein (pGEX/rBBTV CP) with GST tag was overexpressed from pGEX/BBTVCP clone. These affinity tags in the fusion rCP enabled purification from other *E. coli* proteins. Although pRSET/rBBTV CP was soluble, the 20 kDa protein was highly unstable and partially degraded during purification at 4 °C. Curiously, pGEX/rBBTV CP dissociated from its GST affinity tag and the rCP without the tag degraded. On evaluating the protease cleavage sites in the fusion protein, trypsin cleavage sites were present between the C terminal of GST and N terminal of BBTV CP which might be the reason for cleavage of the ~20 kDa protein from its affinity tag. Thus, it was impossible to purify the protein from the pool of *E. coli* proteins.

Restriction free (RF) cloning of BBTV CP to pGEX-4T-2 was attempted not only to replace these trypsin cleavage sites but also the thrombin cleavage site present in the vector with Tobacco etch virus (TEV) NIa protease site. Thrombin is a specific enzyme used to cleave off the tag from the fusion protein after purification. However, its specificity is not universal. Furthermore, the commercially available enzyme is costly. TEV protease on other hand was produced in the laboratory and was highly specific. However, the cleavage using TEV protease was unsuccessful apparently because of a steric hindrance contributed by the two extremely ordered regions flanking the TEV cleavage site present in the disordered region of the fusion protein. pET/rBBTV CP was highly soluble like ΔpGEX/rBBTVCP.

Likewise, BBrMV CP gene was cloned into pRSET-C and pGEX-4T-2 to obtain pRSET/rBBrMV CP and pGEX/rBBrMV CP of size 34 kDa and 60 kDa respectively. The 34 kDa pRSET/rBBrMV CP was insoluble. Overexpression and purification of the protein was standardized in various conditions to increase solubility. On the contrary, pGEX/rBBrMV CP was highly soluble and was purified by GSH Sepharose affinity column chromatography. 360 µg/ml of untagged protein was obtained from 1 l culture. However, like any other potyviral CP, the exposed N and C terminal of BBrMV CP was also prone to proteolytic cleavage. It partially degraded when incubated with thrombin at

room temperature for GST tag cleavage. All these bands were detected by potyviral CP specific antibody in Western blot. Further on storage complete degradation of the protein was observed. Further standardisation of the protocol is necessary to either stabilise monomeric CP or develop BBrMV VLPs *in vitro* for immunising animal in order to raise the antiserum.

The immunogenicity of the antigens (rBBTV CP and rBBrMV CP) was confirmed by Western blot using BBTV CP specific and potyvirus CP specific antibody procured from NRC, Banana and IISc, Bangalore respectively. The rCPs were also characterized by fluorescence spectroscopy, sucrose gradient ultra centrifugation and electron microscopy. The fluorescent spectra of tagged and tag less rBBrMV CP deviated from 330 nm which is typical for a partially disordered protein. However, the spectra of pET/rBBTV CP and ΔpGEX/rBBTV CP were different. The former depicted the spectra of a mostly globular protein. There were two λmax for the fluorescence spectra of ΔpGEX/rBBTV CP.

The epitope prediction of BBTV CP with Trx tag gave interesting insights. A single linear epitope of 80 residues were detected in pET/rBBTV CP comprising of C terminal of the affinity tag and the N terminal of BBTV CP. This was expected to increase the immunogenicity of the antigen and administered for production of antiserum.

The titre value of polyclonal antiserum produced against the 37 kDa pET/rBBTV CP was evaluated by DAC-ELISA and was found to be 1:128000. Titre value for serological assays of field samples was standardized as 1:10000 to be more inclusive for detecting virus even at early stages of infection. A total of 247 tissue culture samples and 10 field samples were screened for the presence of the virus using the antiserum and was compared with the procured antiserum. Seemingly, the latter non-specifically reacted with plant proteins which gave a higher absorbance value in negative control and correspondingly high absorbance in the infected samples. The polyclonal antiserum raised against rBBTV CP was used to standardize serological detection assays like IC-PCR, DIBA and TAS-ELISA apart from DAC-ELISA. DIBA and TAS-ELISA were the most sensitive assays which could detect up to 1:80 dilution of the antigen.

In conclusion, due to the higher nucleotide variability of the CP gene, serological detection is preferred over nucleic acid based assays. However, the quality of antigen used for raising the antibody plays a major role in serodiagnostics. Hence, high quality rCPs of both BBTV and BBrMV were developed in the laboratory in various vector/host systems. The

pET/rBBTV CP overexpressed in C41 strain of *E.coli* (1.1 mg/ ml obtained from 1 L culture) was used for immunisation of the animal. A highly sensitive antiserum specific to BBTV with a titre ten fold higher than that of the commercially available antiserum was obtained. Using this antiserum raised against rBBTV CP, various serodiagnostic assays were standardised in the laboratory. Among these, TAS-ELISA was the most sensitive, detecting antigen even at higher dilution.