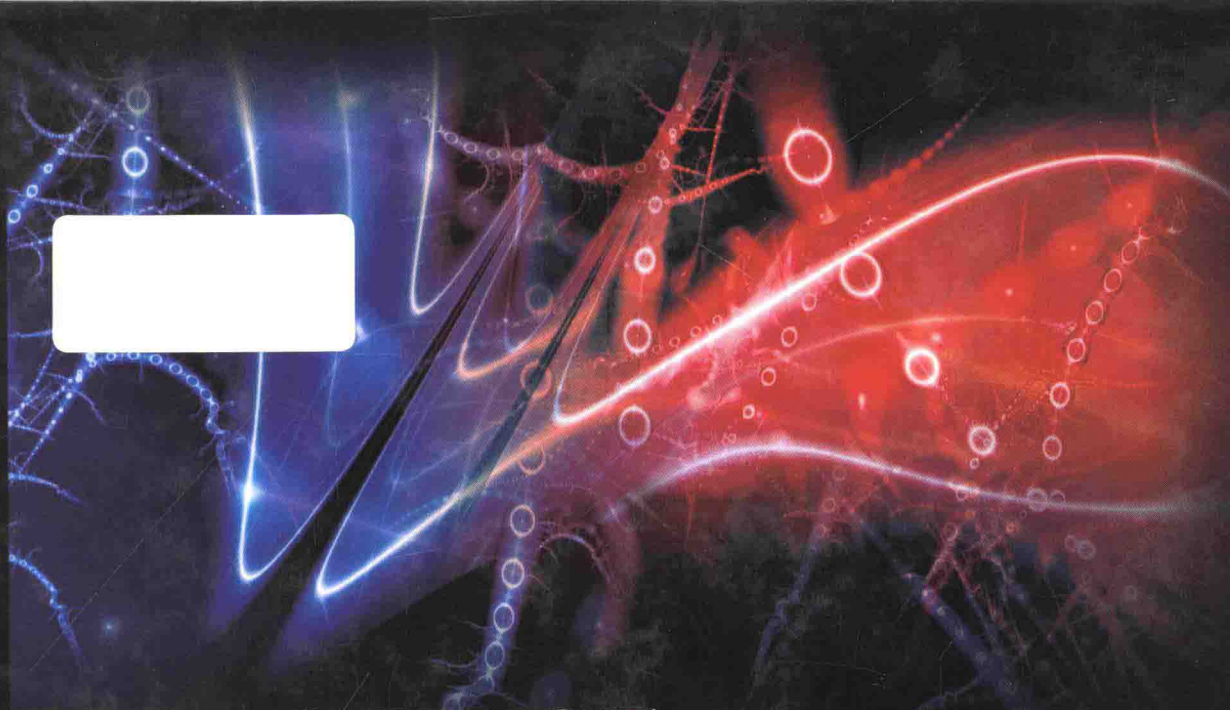

BIOLOGY OF USEFUL PLANTS AND MICROBES

Editor
Arnab Sen

Narosa

BIOLOGY OF USEFUL PLANTS AND MICROBES

Editor
Arnab Sen

Narosa Publishing House
New Delhi Chennai Mumbai Kolkata

Biology of Useful Plants and Microbes
412 pgs. | 30 figs. | 40 tbls.

Editor
Arnab Sen
Department of Botany
University of North Bengal
Raja Rammohunpur
Siliguri, West Bengal

NAROSA PUBLISHING HOUSE PVT. LTD.

22 Delhi Medical Association Road, Daryaganj, New Delhi 110 002
35-36 Greams Road, Thousand Lights, Chennai 600 006
306 Shiv Centre, Sector 17, Vashi, Navi Mumbai 400 703
2F-2G Shivam Chambers, 53 Syed Amir Ali Avenue, Kolkata 700 019

www.narosa.com

Printed from the camera-ready copy provided by the Editor.

ISBN 978-81-8487-264-4

Published by N.K. Mehra for Narosa Publishing House Pvt. Ltd.,
22 Delhi Medical Association Road, Daryaganj, New Delhi 110 002

Printed in India

BIOLOGY OF USEFUL PLANTS AND MICROBES

Dedicated to the memory of
my father
Sri Paresh Chandra Sen
(1940-1994)

Foreword

The importance of research in the field of economically important plants especially medicinal plants and plants for herbal products and cosmetics and whole lot of bacterial community used for biotechnological purposes for the production of medicines and vaccines etc. are much more relevant today than the last century. Bioinformatics, on the other hand is fast emerging as an important discipline for academic research and industrial applications. Research and development in bioinformatics and computational biology requires the co-operation of specialized personnel from the field of general biology, Computer science, Statistics etc. The overall goal of the subject is to elucidate cell function and physiology from a comprehensive set of measurements. Various computational methods are being widely used to address a broad spectrum of problems including the research development and management of economically important plants and microbes

Keeping this in view, in the last winter, Dr. Arnab Sen organized a conference on "Biology and Bioinformatics of Economically Important Plants and Microbes" at University of North Bengal, Siliguri. In this conference, hundreds of eminent scientists gathered and expressed their views on useful plants and microbes of the planet. The idea of the present volume was put forward in a brainstorming season in that conference.

The edited volume "Biology of useful plants and microbes" (edited by Dr. Arnab Sen) consists of 28 articles and virtually has two parts. In one part, modern biological research and development of plants and microbes were discussed and in the other part various techniques of computational biology and bioinformatics were covered. The chapters were written by renowned scientists and science writers in the field of biology and bioinformatics.

However, the best part of the book is perhaps its editing. I must congratulate this young editor for this compilation and for bringing out this book within short time. Dr. Arnab Sen with all his editorial skill and experience including the experience of editing Indian Journal of Microbiology (a Springer publication) has brilliantly abridged the chapters in such a way that the present volume has become a treasure in the field of economic biology.

Dr. T. Madhan Mohan
Advisor
Department of Biotechnology
Ministry of Science & Technology
Govt. of India

Preface

There is perhaps an everlasting quest in biology and that is to learn about ourselves, human being in particular and the living community as a whole. However, through this learning procedure man has also learnt to exploit the living world for its well being. Since the dawn of the civilization, he started to realize about the importance of plants, animals and microbes.

This has become more apparent as we entered in to the 21st Century. More than ever before, human communities are now dependent on plants and microbes. Therefore, research in the field of medicinal, cosmetics and other products from native plants and microbes has once again taken the centre stage of biological research.

Another important aspect is the coordinated interaction between microbes and plants. This interaction is of paramount importance for their healthy association in almost all agricultural settings.

Realizing the importance of plants, microbes and their everlasting association, modern techniques are widely used these days which has led to the development of new disciplines in biology. One such discipline is Bioinformatics.

Bioinformatics is the field of science in which biology, computer science and information-technology merge to form a single discipline. Its ultimate goal is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned. These days, techniques and principles of bioinformatics have been widely used in analyzing the genomes of various beneficial and harmful microbes as well as economically important plants. This way bioinformatics has come up as a major tool for studying the modern day biology. It is in this context Department of Botany, University of North Bengal in association with NBU Bioinformatics Facility organized a three day National Conference on "Biology and Bioinformatics of Economically Important Plants and Microbes". In that conference it was decided that a proceeding volume will be published on the papers presented. The present volume is a culmination of that effort. There are all together 28 articles in this book which may be broadly divided into two parts. One part is biology of economically important plants and microbes and the other part is bioinformatics study of various plants and microbes along with description of various bioinformatics tools.

This book couldn't have been materialized without the active support of University of North Bengal, financial support from Department of Biotechnology, Government of India and Narosa Publication House, New Delhi for agreeing to publish this volume. I am also thankful to Dr. T. Madhanmohan, Adviser, DBT, Government of India for providing help as and when required and also for writing a foreword for this book. I am also thankful to the contributors for conferring

considerable confidence on me and contributing in this volume. I would like to thank my colleagues in the Department of Botany, NBU as well as in the laboratory, especially Mr. Arvind K. Goyal for giving me great editorial support. Thanks are also due to my mother, wife and children for allowing me to spend most of the time working on this book which I should have spent with them. Above all I would like to thank the Almighty God, without whose blessings no work can be done, I believe.

Arnab Sen

Contents

Biology of useful plants and microbes
Editor: Arnab Sen

Chapter 1

Molecular characterization of the *Citrus tristeza* virus (CTV) genome

Avijit Tarafdar[1], KK Biswas[2] and PD Ghosh[1*]
[1]Cytogenetics and Plant Breeding Section, Plant Biotechnology research Unit, Department of Botany, University of Kalyani; [2]Division of Plant Pathology, IARI, New Delhi

Abstract

Citrus tristeza virus (CTV) is a viral species of the *Closterovirus* genus that causes the most economically damaging disease to *Citrus* plants. CTV is with a long flexuous virion of 2000 X 11 nm size containing positive sense, mono-partite, single stranded RNA about 19.2-19.3kb long enclosed by two types of capsid proteins. The size of its genome makes CTV one of the largest RNA viruses known. The CTV genome contains 12 open reading frames, which could encode at least 17 proteins. CTV exhibits a wide range of diversity in the natural environment. Analyses of complete genomes sequence of many CTV isolates available in GenBank database have shown that the 3' terminal half of the genome is well conserved, whereas the 5' terminal half exhibits high sequence diversity. Genetic diversity studies involving Indian isolates have confirmed that Indian isolates form distinguishable phylogenetic groups, and majoritiy of them are in phylogenetic association with VT genotype.

Keywords

Citrus tristeza virus (CTV), virion, genome, genetic diversity

* *Corresponding author:*
Email: pdggene@rediffmail.com

Introduction

Citrus tristeza virus (CTV) is the longest known, phloem-limited, plant virus under genus *Closterovirus* in the family Closteroviridae. CTV is with a long flexuous virion of 2000 X 11 nm size containing positive sense, mono-partite, single stranded RNA of 19.2-19.3 kb with 12 open reading frames (ORFs)[1]. Virions contains about 6% RNA content that are helically constructed with a basic pitch of about 3.7 nm[2], about 8-9 capsids per helix turn and a central hole of 3-4 nm. CTV was first *Closterovirus* to be biophysically characterized. Unlike the virions of other elongated plant viruses they posses cylindrical nucleocapsids and like the other members of *Closteroirida*[3], CTV virions are bipolar and contain two coat proteins of 25 kDa (major CP) and 27kDa (minor CP, CPm) that encapsulate about 97 and 3% of virion length respectively[4]. Virions consist of long helical body and short tail. The tail corresponds to 5' end region of the viral genome and the particle tail of other Closteroviruses has been associated with small amount of p61 and p65.

Molecular characterization of the CTV genome

The 19.3 kb single stranded and positive sense gnomic RNA (gRNA) of CTV is divided almost equally into two parts, the 5' part consisting of ORF lab and lb harboring the viral replication machinery and the 3' half harboring ten ORFs encoding a range of structural proteins and other gene products involved in virion assembly and host and vector interactions[1]. Complete nucleotide sequences of atleast 18 distinct CTV isolates have been determined[5]. Interestingly while the sequences of the 3' half of all the sequenced CTV isolates shared 97% and 89% identity when comparing the A' non translated regions (NTRs) and rest of the 3' half respectively, the 5' half sequences often differ considerably with 60-70% identities in 5' NTR and coding region[1,5].

A remarkable feature of CTV isolates are the close identities (97%) of their 3'NTR and considerable divergence of their 5' NTR sequences. Sequences of 5' NTR are of three types, but surprisingly even dissimilar sequences of the 5' NTRs from different strains are predicted to contain two stem loops (SLR and SL2) separated by a short spacer region[6]. Mutations disrupting secondary structure abolished replication whereas compensatory mutations allowed replication to resume, suggesting the essentiality of the predicted secondary structure for replication than primary structure. However, some sequences of the 5'-UTR were necessary for both replication and virion assembly[4]. The 3' UTR, a highly conserved region[6] is critical for recognition by the replicase complex. The secondary structure of this region comprises a series of ten stem-loops with variable ability to support mutations affecting the primary or the secondary structure. Core of the 3' replication signal might be located in three of the central stem-loops (SL) mapped to be 230 nt within 3'

NTR, mutations introduced in this region completely prevented replication suggesting critical recognition signal for replicase complex in this particular region[7]. A terminal 3' triplet CCA seems essential for efficient replication[7].

Genome organization and Function

The single stranded positive sense genomic RNA of CTV is organized into twelve open reading frames (ORF's), potentially encoding at least 17 protein products, and two untranslated regions (UTR's) of about 107 and 273 nt at the 5' and 3' termini, respectively[8]. These 12 ORFs of CTV are expressed through a variety of mechanisms including proteolytic processing of the poly protein, translational frameshifting, and production of ten 3'- terminal genomic RNAs[9]. The first two mechanisms are used to express proteins encoded by the 5' half the genome while the third mechanism is use to express ORF's 2-11[10]. ORFs 1a and 1b encode proteins of replicase complex, whereas ORF 2-11, spaning the 3' terminal half of JRNA, encode proteins p33, p6, p65, p61, p27, p25, p18, p13, p20 and p23. The coding region of the gRNA is divided into, replication module (ORFs la and lb), a five-gene module (p6, p65, p61, p27 and p25) encoding proteins involved in virion assembly and transport, which is unique signature block, conserved in all members of the family *Closteroviridae,* the p20 gene, a homologue of the p21 gene of *Beet yellow virus* (BY), and four genes encoding proteins with no homologue in other *Closteroviruses* (p33, p18, p13 and p23)[11]. The ORF la encodes a 349 kDa poly protein containing two papain-like protease and methyl transferase like and helicase like domains. Translation of the poly protein could also continue through the 54 kDa, RNA dependent RNA polymerase (RdRp) like domain (ORF lb) by +1 frameshift[12]. The function of p33 (ORF2) is unknown, but is required for infection of a subset of the viral host range. The next protein products encoded by the 3' five -gene module include a transmembrane protein (p6), a homologue of the HSP70 plant heat-shock proteins (p65), two diverged copies of the capsid protein (p25 and p27)[13], and an additional protein (p61) also regarded as a diverged CP copy[11]. The p6 homologue in BYV has been shown to be a movement protein and required for systemic invasion of host plants[14]. The coordinate action of p65 (ORF4, cellular heat –shock protein homologue, HSP70) and p61 (ORF5), in addition to the CP and CPm coat proteins, is required for proper virion assembly[15].

These two proteins, p65 and p61 probably bind to the transition zone between CP and CPm and restrict CPS to the virion tail[4]. The p20 protein is a major component of CTV-induced amorphous inclusion bodies[16], and p23, an RNA binding protein with a Zn finger domain, regulates asymmetrical accumulation of the positive and negative strands during RNA replication[7]. Both p20 and p23, in addition to the CP, act as RNA silencing suppressors in *N.*

benthamiana and *N. tabacum* plants, with p23 inhibiting intercellular silencing, CP intracellular silencing, and p20 both inter- and intracellular silencing[17]. Deletion mutants lacking genes p33, p18 and p13 were capable of replication and assembly[15,18], indicating that they are not required for these functions thus their role in CTV biology remains unknown. Protein p23 ORF 11 is unique in CTV as no homologue found in other *Closteroviruses*[19], it is a multifunctional protein which binds cooperatively both ss RNA and ds RNA, contains a zinc-finger domain that regulate the synthesis of minus strand molecules, controls the level of genomic and subgenomic negative stranded RNAs and is an inducer of CTV like symptoms in transgenic *Citrus aurantifolia* plants[1].

Replication of the CTV gRNA involves synthesis of negative strands that serve as template for the generation of new positive strands, although the latter accumulate 10-20 times more than the negative strands[7]. During replication large number of less than full length RNAs are produced. These include ten 3' co-terminal sg RNAs[20], and ten negative stranded sg RNAs corresponding to the ten 3' sg mRNAs, plus ten 5'-coterminal sgRNAs that apparently are produced by termination just 5' of each of the ten ORFs[21,22].

Infected cells also contain huge amount of two other small 5-coterminal positive stranded sg RNAs of about 600 and 800 nt designated as low molecular weight tristeza (LMT1 and LMT2) produced by premature termination of the gRNA synthesis, thus totaling more than thirty sg RNAs in infected cells [9,22,23].

Frequently, CTV-infected tissues also accumulate large amounts of positive- and negative-stranded defective RNAs (D-RNAs) that contain the 3' and 5' termini of the gRNA but lack variable portions of the central region with extensive internal deletions. These viral RNAs, which are easily observed by electrophoretic analysis of plant extracts enriched in double-stranded RNA (dsRNA), are very common in field isolates[23-26]. CTV D-RNAs accumulate abundantly even when their genome contains less than 10% of the viral genome. Most D-RNAs are 2.0-5.0 kb in size, but large D-RNAs comprising ORFs la and lb in their 5' proximal moiety, or ORFs 2-11 in their 3' terminal moiety, have been reported[27]. Sequence analysis of the junction site and flanking regions suggest that most D-RNAs must be generated by a template- switching mechanism induced by different factors[28]. The minimal replication signals required for D-RNA replication in trans are located in the 5' proximal 1 kb and at the 3'-UTR of the D-RNA sequence[29,30].

The biological role of CTV D-RNAs is presently unknown, as no specific associations have been established for any of the D-RNAs. Their wide occurrence in CTV isolates suggests that they may provide some advantage. At least in one case, the presence of a D-RNA was reported to modulate CTV symptom expression[31]. The interaction of CTV with host factors

is largely unknown, but examination of D-RNA from Alemow plants infected with seedling yellow (SY) and non-SY inducing isolates revealed mostly a major single D-RNA of 4.5 or 5.1 kb in non-SY plants and two different D-RNAs of 2.4 or 2.7 kb in SY plants. These results suggested the possibility that D-RNAs might play a role in suppression of SY symptoms[31]. Most CTV isolates occur in mixed population of different variants[32] and defective RNAs. However, some CTV isolate consists principally of one genotype with minor concentrations of other variants.

Manjunath *et al.*[33] cloned and sequenced coat protein (CP) gene of four selectively Indian CTV isolates. Eight ORFs of the total 12 ORFs in CTV genome of Florida severe isolate T36 have been cloned and sequenced for the first time (Figure 1)[34]. After comparison in the annotated amino acid sequences of CTV ORFs with other annotated amino acid of viral ORFs available in database it was reported that genome of CTV have been arisen as a result of multiple recombination events[34]. Complete genome of 18 CTV isolates from different geographical areas inducing different phenotypes has been sequenced and deposited in GenBank. They are T36 (U16304), and T30 (AF260651) from Florida, VT (AF001623) from Israel, T385 (Y18420 and T318A (DQ151548)) from Spain, SY568 (AF001623) from California, NuagA (AB046398) from Japan, Qaha (AY340974) from Egypt, B165 (EU076703) from India, one isolate Mexico (DQ272579) from Mexico, five isolates, NZ-B18 (FJ525436), NZRB-G90 (FJ525432), NZRB-M12 (FJ525431), NZRB-M17 (FJ525435), NZRB-TH28 (FJ525433) and NZRB-TH30 (FJ525434) from New Zealand,

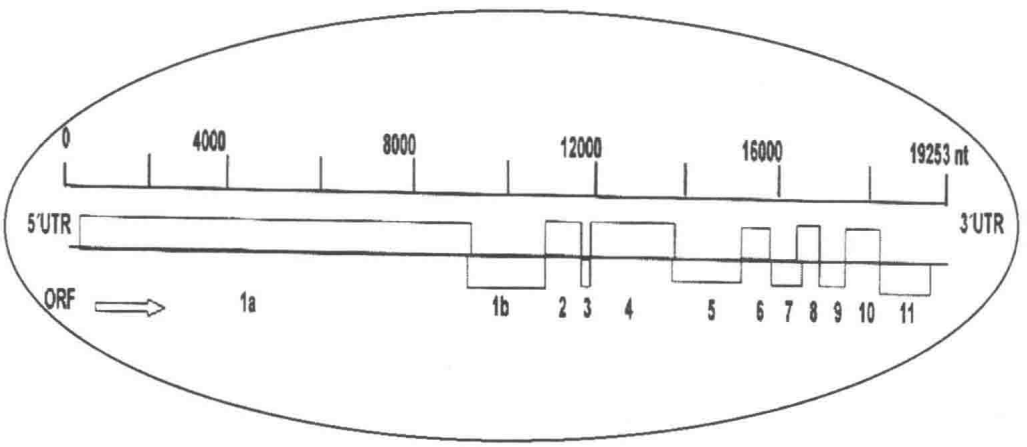

Figure1: Diagram of the organization of CTV complete genome (20- kb) with 12 open reading frames

and two isolates, HA16-5 (GQ454870) and HA18-9 (GQ454869) from Hawaii.

Genetic diversity

CTV has been shown to exhibits a wide range of diversity in the natural environment[35,36] and existence of several divergent CTV isolates have been reported from many citrus growing countries[37]. Nucleotide sequence analysis has been proved to be an accurate procedure for estimation of genetic diversity of CTV[38] and using this method several divergent isolates have been reported from citrus growing countries[39,40]. Various procedures has detected variations in viral RNA, including: (1) differences in dsRNA pattern between isolates[25,41], later shown to be due to the presence of D-RNAs in CTV isolates; (2) hybridization pattern with cDNA or cRNA probes of several gRNA regions[42], (3) restriction fragment length polymorphism (RFLP) analysis of the CP gene[43], (4) RT-PCR amplification patterns with primer sets specific for several CTV genotypes[37]), for 5-UTR sequence types I. II and III[44,45] or for three. groups of isolates differing by their p23 sequence[46], and (5) single-strand conformation polymorphism (SSCP) analysis of different gRNA regions[47,48]. This SSCP technique has been used to characterize the population structure of CTV isolates and select specific variants for sequencing, thus allowing estimates of the genetic diversity within and between isolates[49]. The values obtained were in the range of those calculated for other plant viruses[50], but higher than those estimated for other members of the family Closteroviridae[38].

Genetic variation is unevenly distributed along the CTV ERNA, the most conserved region being the 5'UTR, with over 95% identity between various isolates, and most variable being the 5UTR with identity values as low as 44-55 % between sequence types of different groups[6]. Analysis of genomes of two CTV isolates VT and T36, more similarity (89 %) at 3' half than that at the 5' half of the CTV genomes were noticed which suggests that diversity of CTV isolates was not resulted from a recent recombination event[51]. Later on studies on the genetic diversities of CTV isolates began in citrus growing countries of the world. Sequence diversity in some CTV isolates is shown to be uniform throughout the whole genomes, while in other isolates it is asymmetrical[18]. By genome comparison of various CTV isolates it was found that CTV exhibits about 30% sequence diversity in the 5' half, while 10% in 3' half of the genomes[52]. Analyses of complete genomes sequence of many CTV isolates available in GenBank database have shown the pattern of sequence variation among the isolates with > 90% in 3' proximal 8 kb, and much greater divergence in the 5' proximal 11 kb of the genome[37], concluding the 3' terminal half of the genome is well conserved, whereas the 5' terminal half exhibits high sequence diversity[53].

Overall, sequence comparisons between CTV isolates of different geographical origin and

pathogenicity characteristics showed a high degree of conservation between CTV genomes separated in time and space, with a limited repertoire of genotype[45], and a population structure variable between isolates, with some consisting of a predominant sequence and some closely related variants, and others having a complex structure with highly divergent sequence variants[49]. CTV genome contains three RNA silencing suppressors viz., P20, P23, and CP[54].

CTV isolates based on sequence diversity in various gnomic regions were divided in to different groups viz., three groups[6]. Sequence variation is comparatively uniform in genome of CTV-VT, whereas it is very divergent with other CTV isolates like T36 and primers based on diverged sequences detected only a few CTV isolates[55]. Complete nucleotide sequences of Florida mild isolate (T30) which are identical to a Spain mild isolate (T385), suggesting parents of these isolates might have common origin probably in Asia, followed by dispersion throughout the world by movement of citrus[42].

Classification scheme based on sequence analysis of 5' and 3'end of CTV genome has been proposed, and two major groups VT and T36 have been identified by Hilf *et al.*[37]. The VT group members had sequence divergences which were distributed in constant proportion throughout the genome, while the T36 group members had highly divergent and conserved sequences at 5' and 3' ends respectively.

Rubio *et al.*[38] analyzed sequences of four ORFs of 30 CTV isolates from Spain and California and concluded that sequence variants originated due to effect of recombination between diverged sequences variants. Roy *et al.*[43] in USA sequenced CP genes four Indian CTV isolates and compared the nucleotide, and amino acid sequence, with other international isolates and considerable genetic variability among them was not found. Roy and Brlansky[56] based on the ORFla, overlapping region of RdRp and P33 sequence analysis of 21 Indian CTV isolates, divided them in to five distinct groups. Most of the analyzed isolates were found of VT genotype and first time reported T36, T30 (occurring in mixed infections) like isolates under Indian conditions. Population structure of CTV isolate SY568 revealed that it composed of sequence types occurring with varying nucleotide frequency[57].

Conclusion

Genetic diversity of Indian CTV isolates has been studied earlier using limited number of isolates[40]. Analyzing 15 Indian isolates with specific primers, it is proposed Indian isolates are of VT genotype[37]. Using 21 Indian CTV isolates based on sequences from 5'ORFla (position 697-1484 nt) it is reported that Indian isolates form distinguishable phylogenetic groups, and majorities of them are in phylogenetic association with VT genotype[58].