Plant virus genome organization

PN Sharma
Department of Plant Pathology
CSK HPKV, Palampur-176062
No of sequence data on plant virus genome generated over the last one decade

1st viral genome sequenced was DNA of CaMV (Franck et al., 1980) and then RNA of TMV (Goelet et al., 1982)

By 1990, genome sequences of ~40 spp. From 20 groups have been determined.

Genome of > 250 spp. have been fully sequenced.
Viral Genomes

- Most plant viruses are RNA although some 80+ viruses are DNA viruses.
  - 540 SS RNA
  - 40 DS RNA
  - 30 DS DNA
  - 50 SS DNA
- Small genome ~ 4-20 kb
- Simple - normally just a few genes
- Compact - use “compressing” strategies
- Use cellular translation machinery
- Encode their own RdRp
Flow of genetic information

ssDNA ↔ dsDNA ↔ ss(+)RNA ↔ dsRNA

ss(-)RNA

Protein(s)
Genome Properties

- Nature of the genome
  - Circular or linear

- Number of genome components
  - DNA or RNA
  - Single or multiple (up to 11) ranging in size from 1 kb in satellite virus STNV (Nanovirus) to 20 kb (Closterovirus)
  - Positive sense, negative sense, or ambisense
Genome Properties

- Infection versus transfection
- Number of genes
  - Most plan viruses have at least 3 genes
    - 1 or more for the capsid protein – genome protection
    - 1 or more – movement, cell to cell and long-distance
    - 1 or more - nucleic acid replication
Genome Properties

- Biological properties – may not correlate with the genome properties
- Genome relatedness
  - Has replaced serological properties as a tool
- Translation strategy
  - Directly or from intermediates, sub-genomic RNAs
Positive Sense RNA Viruses

- Purified (+) sense RNAs are directly infectious
- Genome size may be limited due to: fragility of the RNA molecule or packing constraints
  - Generally have a 5’ untranslated region (UTR) and a 3’UTR
  - 5’ nucleotide may be covalently attached to a protein (VPg) or to a methylated cap; 3’ end – may be poladenylated (Poly A’tail)
Negative Sense RNA Viruses

- More diverse than positive sense RNA viruses
- Genome may be segmented
- Carries a virus-specific polymerase in the virion
- Makes (+) sense mRNA or double stranded RNA intermediates called either replicative intermediate (RI) or replicative form (RF)
DNA Virus Genomes

- Bacteriophages
- Some are circular
- Multipartite
  - Found only in plants
  - Different genome segments are separated into different particles
  - Requires large input of inoculum
General properties of plant viral genome

- Basically plant virus genome comprises of;
  - Coding regions that
    - expresses the proteins required for viral infection cycle
    - Movement through the plant
    - Interaction with host
    - Movement between hosts
  - Non-coding regions that control the expression and replication of the genome
  - Control sequences that can also be found in coding regions
Information contents

- Same nucleotide sequence in a viral genome could code for up to 12 or more polypeptides.
- There could be an ORF in each of the three reading frames of both +ve & -ve sense strands, that give rise to six polypeptides
  - ORF is a sequence that commences with AUG initiation codon and is capable of expressing a protein of 10KDa or more.
- Read through and frame shift are quite common
• No. of genes found in viruses ranges from 1 (in STNV) to 12 (in Clostero & Reoviruses).

• Most of the ss +ve sense RNA genome code for ~4-7 proteins.

• In addition to coding regions for proteins, genomic n/a contains nucleotide sequences with recognition and control function that are important for virus replication. These control and recognition functions are found mainly in the 5’ & 3’ non-coding sequences of the ssRNA viruses, however, they may also occur internally even in coding sequences.

• Viruses make very efficient use of the limited amount of genomic n/a they possess.
Functions of viral gene products

The known function of viral gene products may be classified as:

- **Structural proteins**: are the CP of small viruses, nucleoproteins of reovirus and tenuiviruses.
- **Enzymes**
  - **Proteases**: are coded for by those virus groups in which whole genome or segment of the genome is first transcribed into a single polyproteins.
Polyproteins: a single ORF

VPg = Virus Protein, genome-linked Cap-independent translation

TEV (Potyviridae)

VPg

P1  HC-Pro  P3  CI (HEL)  NIa  NIb (RdRp)  CP

Poly(A)
Enzymes involved in n/a synthesis: it is now well accepted that all plant viruses code for one or more proteins that have an enzymatic function in n/a synthesis either genomic n/a or mRNA or both. These sequences are polymerases:

- **RdRp**: polymerases that catalyze transcription of RNA from an RNA template is RdRp (RNA dependent RNA polymerase)
- **Replicase**: the enzyme complex that makes copies of an entire RNA genome & the subgenomic mRNA.
- **Transcriptase**: in an RdRp is found as a functional part of the virus particle it is called transcriptase. E.g. in Rhabdoviridae & Reoviridae.
- **RdDp**: the enzyme which copies a full length viral RNA into genomic DNA is called RNA dependent DNA polymerase (RdDp) or reverse transcriptase (RT). E.g. in Caulimoviridae
- **DdDp**: DNA dependent DNA polymerase is of host origin which is utilized by the DNA viruses for their replication e.g. geminiviruses
Sub-genomic RNAs

*Tombusviridae: TBSV*

4.2-kb genome

2.2-kb sgRNA

0.9-kb sgRNA
Virus movement and Transmission

- A specific virus coded protein has been identified as a essential requirement for cell to cell movement and with in the plant
- For virus vector specificity

Non-enzymatic role in RNA synthesis: the 5'VPg protein found in some virus genera is believed to act as a primer in RNA synthesis

CP of AMV

- Cp of AMV and the corresponding protein in Ilarviruses has an essential role in the initiation of infection by the viral RNA, possibly by priming (-) strand synthesis.

Host cell recognizing proteins
Non-coding regions:

a. **End structures:** many plant viral ssRNA genome contains specialized structures at their 5’ and 3’ termini

- **The 5’ cap:**
  - many mammalian cellular and animal virus mRNA have a methylated blocked 5’ terminal group of the form:
    - $m^7G^5pp^5X^{(m)}pY^{(m)}p$, where $X^{(m)}$ & $Y^{(m)}$ are the two methylated bases. Some plant viral RNAs have this type of 5’ end k.a. ‘CAP’ but the bases $X$ and $Y$ are not methylated in such known plant viral RNAs.

- The capping activity is virus coded & differ from the host capping activity.

- Viral & cellular enzymes are involved in the synthesis of cap structures e.g. in TMV 126 kDA TMV protein and in 1a encoded by RNA1 of BMV.
**b. 5’ linked protein (VPgs)**

- Plant viruses have a small protein (~3.5-24 kDa) covalently linked to the 5’ end of the genome k.a. VPgs.
- All the VPGs are coded for by the virus concerned & has been identified in many viruses.
- In multipartite RNA genome all the RNAs genome have the same proteins.
- The VPGs is attached to the genomic RNA by a phosphodiester bond between the b-OH group of serine or tyrosine residue of the genomic RNAs.
- VPGs are involved in virus replication.
C. 3’ poly (A) Tracts

- Polyadenylated sequences present in mRNA of eukaryotes have also been found in the 3’terminus of viral RNAs that can act as messengers.
- The length of poly A tail varies from virus to virus e.g. 25-170 for B RNA of CPMV.
- Internal poly A tracts are found in bormoviruses and hordeiviruses.
D. 3’ t-RNA like structures

- Pink et al., 1970 reported that TYMV RNA when incubated with various $^{14}$C-labeled a/a in the presence of appropriate cell free extracts from yeast or *E. coli*, bound valine which become attached to the 3’ terminal adenosine by an ester linkage.
- Later same acceptance of valine was observed in other Tymoviruses.
- Also present in ds replicative form of RNA and in this state RNA was resistant to RNase attack.
- These experiments demonstrated that the a/a accepting activity was integral part of the viral +ve sense ssRNA
- Later 3’ terminus of TMV RNA accept histidine
- MBV RNAs and BSMV RNAs accept tyrosine
- The tertiary structures of these RNAs-like 3’ termini involve Pseudoknots
E. Complementary 5’ & 3’ sequences

- The RNA genome segments of the TOSPO an tenuiviruses have complementary sequences at 5’ & 3’ ends, that enable the termini of the RNAs to anneal to form pan-handle structures.

- These sequences are conserved across genome segments. E.g. the three tospoviruses genome segment have a 5’ consensus sequence of AGAGCAAU--- & a 3’ sequence of UCUCGUUA--.
Intergenic regions

- Intergenic sequences also involved in both RNA synthesis and translation of downstream ORFs.

- The initiation of synthesis of sub genomic RNAs is in these regions and these RNAs are the messengers for translation of non -5’ ORFs in many viruses.

- Many such interactions have been recognized between terminal and internal sequence regions in the control of expression of the genomic information from +ve strans RNAs (may also be in –ve strands and dsRNA genomes and the genomes of ss and ds DNA viruses.)
1. DNA and RNA are polymers composed of monomers called nucleotides.

2. Each nucleotide has three parts:
   a. A pentose (5-carbon) sugar.
   b. A nitrogenous base.
   c. A phosphate group.

3. The pentose sugar in RNA is ribose, and in DNA it’s deoxyribose. The only difference is at the 2’ position, where RNA has a hydroxyl (OH) group, while DNA has only a hydrogen.
Case study
Cauliflower mosaic virus - A DNA virus (*Caulimoviridae*)

- Most have narrow host ranges
- Most are relatively unimportant as pathogens; exception is *Rice tungro bacilliform virus*, part of the most important rice virus complex
- Most are transmitted by invertebrate vectors
- Viruses do not replicate in vector; use virus-coded helper protein to aid transmission
- Promoter elements commonly used in genetic engineering of plants
- Caulimoviruses are not very versatile as plant gene expression vectors because of packaging constraints/instability
Cauliflower mosaic virus - A DNA virus (*Caulimoviridae*)

- Isometric 50 nm T=7 particles or bacilliform particles; no envelope
- Nicked dsDNA genomes ~ 8 kb
- First plant virus shown to have DNA genome
- Replication is by reverse transcription
- Transcription is in nucleus; DNA replication in cytoplasm
- Most do not integrate into host genome
Cauliflower mosaic virus genome structure

- Seven ORFs on CaMV genome
- Translation of seven proteins from two transcripts
- ORF 2 is the only dispensable ORF
- ORFs 6 and 7 are involved in translation regulation
- Packaged genomic DNA has discontinuities on both strands
- Replication is from tRNA$^\text{met}$ primer
Genome Organization of *Caulimoviridae*

**Petuvirus**
- PVCV

**Caulimoviruses**
- CaMV
- FMV
- CERV

**Soymoviruses**
- SbCMV
- BRRV

**Cavemoviruses**
- CsVMV

**Tungroviruses**
- RTBV

**Badnaviruses**
- ComYMV
**GENOME ORGANIZATION OF BEGOMOVIRUSES**

**AV 1**: Coat Protein (CP)

**AC 1**: Replication Initiation Protein (Rep)

**AC 2**: Transcription Activator Protein (TrAP)

**AC 3**: Replication Enhancer Protein (Ren)

- Share > 95% homology with all Begomoviruses - probe for group detection

**IR**: Common region between A and B, specific for a virus

**IMYMV-Bg DNA A**

(2745 bp)

**IMYMV-Bg DNA B**

(2616 bp)

**BC 1**

**BV 1**

Movement Protein

- Share only 40-45% homology - probe for virus specific detection
A consensus potyviral genome map (Xu and Hampton, 1996)
The genome of TMV is

- 6395 nucleotides long and
- contains four large open reading frames.
Acknowledgements

- I gratefully acknowledge the use of text book “Matthews Plant Virology” by Roger Hull.

- I acknowledge the scientists who spent valuable time in generating information on various aspects of plant Virology and displayed the same on internet for use by students, teachers and researchers.

PN Sharma