RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF AGRICULTURAL SCIENCES & ALLIED INDUSTRIES

Lecture 6

Genome Organization

The first viral genome to be sequenced was the DNA of CaMV followed by the **RNA of TMV**. By 1990, the genome sequences from about 40 species from about 20 groups (some of the groups have been reclassified since) had been determined. In the year 2000, the genomes of about 250 species had been fully sequenced including representatives of most plant virus genera. There are also numerous partial sequences mainly of viral coat protein genes.

GENERAL PROPERTIES OF PLANT VIRAL GENOMES

A. Information content

In theory, the same nucleotide sequence in a viral genome could code for up to 12 or more polypeptides. There could be an open reading frame (ORF) in each of the three reading frames of both the positive (+)- and negative (-)-sense strands, giving six polypeptides. Usually an ORF is defined as a sequence commencing with an AUG initiation codon and capable of express- ing a protein of 10 kDa or more.

The number of genes found in plant viruses ranges from I for the satellite virus STNV, to 12 for some closteroviruses and reoviruses. Most of the ss (+)-sense RNA genomes code for about four to seven proteins. In addition to coding regions for proteins, genomic nucleic acids contain nucleotide sequences with recog- nition and control functions that are important for virus replication. These control and recogni- tion functions are mainly found in the 5' and 3' non-coding sequences of the ssRNA viruses, but they may also occur internally, even in cod- ing sequences.

B. Economy in the use of genomic nucleic acids

Viruses make very efficient use of the limited amount of genomic nucleic acids they possess. Eukaryote genomes may have a content of introns that is 10-30 times larger than that of the coding sequences.

1. Coding sequences are usually very closely packed, with a rather small number of non- coding nucleotides between genes.

2. Coding regions for two different genes may overlap in different reading frames (e.g. in BNYVV; or one gene may be contained entirely within another in a different reading frame.

3. Read-through of a 'leaky' termination codon may give rise to a second, longer read- through polypeptide that is co-terminal at the amino end with the shorter protein. This is quite common among the virus groups with ss (+)-sense RNA genomes. Frameshift proteins in which the ribosome avoids a stop codon by switching to another reading frame have a result that is similar to a 'leaky' termination signal.

4. A functional viral enzyme may use a host- coded protein in combination with a virus- coded polypeptide (e.g. the replicase of TMV).

5. Regulatory functions in the nucleotide sequence may overlap with coding sequences (e.g. the signals for subgenomic RNA synthesis in TMV).

C. The functions of viral gene products

The known functions of plant viral gene products may be classified as follows:

1. Structural proteins

These are coat proteins of the small viruses, the matrix, core or nucleoproteins proteins of the reoviruses, tenuiviruses and those viruses with a lipoprotein membrane, and proteins found within such membranes.

2. Enzymes

a. Proteases

These are proteases coded for by those virus groups in which the whole genome or a segment of the genome is first transcribed into a single polyprotein.

b. Enzymes involved in nucleic acid synthesis

It is now generally accepted that all plant viruses, except some satellite sequences, code for one or more proteins that have an enzymatic function in nucleic acid synthesis, either genomic nucleic acid or mRNAs or both. The general term for these enzymes is polymerase. Polymerases that catalyze transcription of RNA from an RNA template have the general name RNA-dependent RNA polymerase (RdRp). The enzyme complex that makes copies of an entire RNA genome and the subg nomic mRNAs is called a replicase. If an RdRp is found as a functional part of the virus particle as in the Rhabdoviridae and Reoviridae it is often called a transcriptase. The enzyme coded by members of the Caulimoviridae, which copies a full-length viral RNA into genomic DNA, is called an RNA-dependent DNA polymerase or reverse transcriptase. In the Geminiviridae the viral gene product(s)associate(s) with the host DNA-dependent DNA polymerase.

3. Virus movement and transmission

For many plant viruses, a specific virus-coded protein has been identified as an essential requirement for cell-to-cell movement and for systemic movement within the host plant. Other gene products have been identified as essential for successful transmission by invertebrate vectors, and viral gene products may also be involved in trans- mission by fungi.

4. Non-enzymatic role in RNA synthesis

The 5' VPg protein found in some virus genera is thought to act as a primer in RNA synthesis.

5. Coat protein of AMV

The AMV coat protein, and the corresponding protein in ilarviruses, has an essential role in the initiation of infection by the viral RNA, possibly by priming (-)-strand synthesis. This protein is discussed in more detail in Chapter 8 (Section IV.G)

D. Non-coding regions

1. End structures

The structures at the 5' and 3' ends of viral nucleic acids

2. 5' and 3' non-coding regions

The 5' and 3' non-coding regions control both translation and replication. These two regions interact in the initiation of translation of, at least, the 5' open reading frames (ORFs). The 3' non-coding region is the site of initiation of (-)-strand RNA synthesis and the 5' non-coding region (the 3' end of (-)-strand RNA) is the site of initiation of (+)-strand synthesis.

3. Intergenic regions

Sequences in intergenic regions are also involved in both RNA synthesis and the trans- lation of downstream ORFs. The initiation of synthesis of subgenomic RNAs is in these regions and these RNAs are the messengers for translation of non-5' ORFs in many viruses.

PLANT VIRAL GENOME ORGANIZATION

DOUBLE.STRANDED DNA VIRUSES

A. Family Caulimoviridae

The basic features of the genomes of members of this family is that they are circular dsDNA molecules with one discontinuity in one strand and one or more discontinuity in the other; the discontinuities are associated with the replication of the viruses. Transcription is asymmetric with all the coding information on one strand.

1. Genus *Caulimovirus*

a. Genome structure

The DNA nucleotide sequences of the type member of the genus Caulimovirus, CaMV, and consists of a circular dsDNA molecule of about 8 kb. The circular ds DNA of CaMV has a single gap in one strand and two in the complementary strand. All caulimoviruses have a single gap in the- (or minus) strand. Viruses in the group other than CaMV may have one, two or three discontinuities in the (+) strand. The DNA encodes six and possibly eight genes. These are closely spaced but with very little overlap except for the possible gene VIII.

SINGLE.STRANDED DNA VIRUSES

A. Family Geminiviridae

1. Genome structure,

Four genera have been described within the family Geminiviridae - the mastreviruses, the curtoviruses, the begomoviruses and the topocuviruses. The genomes of all four genera comprise circular ssDNA and are either as one species or divided between two species. The genera are distinguished on their genome organizations. However, all geminiviruses have a conserved genome sequence in common. There is a large (-200 base) non-coding inter- genic region, termed the common region, in the two-component geminiviruses, features of which are found in the similar genomic posi- tion in the single-component geminiviruses. This region has sequences capable of forming a hairpin loop. Within this loop is a conserved sequence, TAATATTAC found in all geminiviruses.

DOUBLE.STRANDED RNA VIRUSES

Double-stranded RNA viruses have genomes that consist of multiple linear dsRNA molecules. In most cases the dsRNA segments are monocistronic. There are some segments that apparently have two ORFs, but is has not been established that both ORFs are expressed. The genera differ in the number of RNA molecules that make up the genome.

Family Reoviridae

The genomes of plant reoviruses comprise either 10 or 12 RNA segments. These segments are numbered according to their electrophoretic mobility in gels, the slowest being segment 1. However, sequencing of the genome segments has shown that the true size is not necessarily the size deduced from its electrophoretic mobility. There are several examples below of higher numbered RNA segments being larger than smaller numbered ones. Four types of gene products are recognized, those that make up the capsid (structural proteins), those involved with RNA replication, non-structural proteins, and those for which no function is known. In several cases, the function is attributed because of sequence similarities or common motifs with better characterized vertebrate-infecting reoviruses.

Genus Oryzavirus

Oryzaviruses have 10 RNA segments. The genome properties of the type member of this genus, RRSV, have been most studied and are summarized in Table 6.3. Unlike the other two genera, the RNA polymerase is not encoded on the largest RNA segment but on the second largest one. However, like the other two genera it is the largest gene product. As with most other plant reoviruses, the particles of oryzaviruses are double-shelled with A-type spikes on the outer shell and B-type spikes on the inner shell. The B-type spikes are encoded on segment 1 and the A-type spikes, which are involved in vector transmission by delphacid planthoppers, are encoded by segment 9.

NEGATIVE SENSE SINGLE STRANDED RNA GENOMES

There are two families (Rhabdoviridae and Bunyaviridae) and two unassigned genera (Tenuivirus and Ophiovirus) of plant viruses with (-)-sense ssRNA genomes. In each of the viral genomic RNA is closely associated with a protein, termed the nucleocapsid pro- tein. In members of the two families this is encapsidated in membrane-bound particles. The genomes of tenuiviruses and ophioviruses do not appear to be contained within membranous particles though, from the relationship that tenuiviruses have with Bunyaviruses, membrane bound particles might be expected. All the (-)-sense RNA viruses have the virus-coded RNA polymerase associated with the virion as this is required for initial transcription as an early step in the infection cycle

A. Family Rhabdoviridae

There are two genera of plant-infecting rhabdoviruses, the Cytorhabdovirus and the Nucleorhabdovirus. In many properties these resemble animal-infecting rhabdoviruses, the genome organizations of several of which are well characterized. Thus, plant-infecting rhabdoviruses, like those infecting vertebrates, possess a genome consisting of a single piece of single-stranded negative-sense RNA, with a length in the range of 11 000-13 000 nucleotides. The genomes of the cytorhabdovirus, LNYV, and of the nucleorhabdovirus, SYNV have been sequenced and shown to have very similar organizations. From the 3' end (this is a (-)-strand genome) there is a 144 nucleotide leader sequence followed by six genes, with short intergenic regions between them. Each gene is transcribed separately.

POSITIVE SENSE SINGLE STRANDED RNA GENOMES

The majority of plant viruses have genomes of positive-sense ssRNA. This can be either as a single molecule or the genome can be divided into several molecules.

A. Family Bromoviridae

1. Genus Bromovirus

a. Genome structure

BMV has a tripartite genome totaling 8243 nucleotides. In addition, a subgenomic RNA containing the coat protein gene is found in infected plants and in virus particles. This coat protein gene is encoded in the sequence toward the 3' end of RNA3. Each of the four RNAs has a 5' cap and a highly con- served 3'-terminal sequence of about 200 nucleotides. The terminal 135 nucleotides of this sequence can be folded into a three- dimensional tRNA-like structure (Perret et al., 1989), which accepts tyrosine, in a reaction similar to the aminoacylation of tRNAs, but this reaction probably needs the 3'-terminal 155 nucleotides. The genome of BMV has been completely sequenced. RNAs 1 and 2 each encode a single protein on what are termed ORFs la and 2a. RNA3 encodes a protein of 35 kDa on ORF 3a at its 5' end. Between this cistron and the coat protein cistron there is an intercistronic non-coding region approximately 250 nucleotides

long. An internal poly (A) sequence of heterogeneous length (16-22 nucleotides) occurs in this intercistronic region ending 20 bases 5' to the start of the coat protein gene (ORF 3b). The first nine bases of the RNA4 consist of the last nine bases of the intercistronic region. The internal poly (A) sequence is not essential for RNA3 replication. However, during replication, in vivo, of constructs lacking this sequence, restoration of the poly (A) took place.