ABSTRACT. Citrus tristeza closterovirus (CTV) causes economically important diseases of citrus throughout the world. CTV has a genome size of about 20 kb, the largest known for a plant RNA virus. The virus particles are long and flexuous and consist of two capsid proteins. The major capsid protein (CP), p25, encapsidates about 95% of the virion while the minor CP (CPm), p27, represents the other 5% and is present at one end of the virus particle. The complexity of the virus and the biology of the diseases it causes have been well studied through symptomatology, graft and aphid transmissions, and double-stranded RNA patterns. The work on the molecular biology of the virus has progressed rapidly in the last few years. Sequences of four CTV isolates are now available and partial sequences of several other isolates have been obtained. The viral genome contains 11 open reading frames, and the virus employs several strategies for expression of its mature proteins including the polyprotein processing by proteases, a +1 ribosomal frameshift and subgenomic RNAs. Functions of many of the viral proteins are not known at present. Many CTV isolates have also been found to be associated with one or several defective RNAs, ranging in size from 1.5 to 10 kb. The sequences of CTV genomes appear to be relatively conserved at the 3’ half with about 10% variability among the isolates. However, the 5’ half shows about 30% sequence divergence among some isolates. The maximum sequence variability of up to 45% was observed in 5’ untranslated regions of some isolates, but RNA secondary structure appears to be conserved in all the isolates. The complexity of virus populations is analyzed by molecular techniques such as sequencing, hybridization and single-stranded conformational polymorphism analysis. Their use in understanding the virus strains, the quasispecies nature of the virus, and strain differentiation are discussed. Development of new experimental systems for replication of CTV in protoplasts and also in citrus plants provides powerful tools for the development of a better understanding of the biology of the virus.

Citrus tristeza virus (CTV), a member of the family Closteroviridae, causes some of the most economically important diseases of citrus (5, 50). Introduction of new strains of the virus and its efficient vector, Toxoptera citricida Kirkaldy, has resulted in severe losses in many countries (51). The most important diseases caused by CTV are: i) decline of most citrus cultivars on sour orange rootstock, ii) stem pitting of different scions irrespective of the rootstocks and iii) seedling yellows symptoms on sour orange and grapefruit seedlings. CTV is associated mainly with the phloem tissue and produces a unique cytopathology of membrane-associated vesicles in infected cells (11).

The flexuous virus particles of CTV are about 2000 × 11 nm in size and consist of an unsegmented, positive sense, single stranded RNA genome of about 20 kb (5) encapsidated by two capsid proteins (CP) (17). The major CP species, p25, encapsidates about 95% of the virion while the minor capsid protein (CPm), p27, is present at one end of the virion (17).

The biology and characterization of CTV isolates has been well documented (5, 50). Mixtures of several genotypes of CTV and the complex nature of viral RNAs which include single-stranded (ssRNA), double-stranded (dsRNA) and defective (dRNA) RNAs in infected plants present problems in understanding the basis of pathogenesis, “strain” differentiation and in devising control measures. Variability in the virus populations has been studied by analysis of dsRNA patterns from infected plants (10, 20, 25, 36) and
characterization of aphid transmitted subisolates (6). The genetic analysis of CTV has progressed rapidly since the first report of the sequence of the complete genome of CTV isolate T36 from Florida (24, 32, 55). The current understanding of the molecular biology of CTV with reference to the genome sequences, “strain” mixtures and strain interactions is discussed.

The CTV Genome. The unusually large genomes of closteroviruses are proposed to have evolved from ancestral tobamo-like viruses via complex shuffling of genome elements, gene duplication and divergence, acquisition of genes from host cells and/or other viruses and development of new mechanisms of gene expression (11). The genomes of closteroviruses can be tentatively divided into four modules. These include: i) domains of methyl transferase, helicase and polymerase which are conserved throughout the alphavirus subgroup; ii) a leader papain-like protease domain, iii) a chaperone module of a small membrane-associated protein, a heat shock protein homolog (HSP70h) and another protein of about 60 kDa; and iv) a fourth module consisting of CPs and several additional 3’ proteins. A protein of about 30 kDa is found between the first two modules of CTV and some other closteroviruses.

The genome of a quick decline causing isolate of CTV (isolate T36 from Florida) was sequenced first (24, 43) and contains 19,296 bases with 11 open reading frames (ORF) (Fig. 1). It is predicted to produce at least 17 proteins ranging from 6 kDa to 401 kDa (24). Untranslatable regions (UTR) of 107 and 275 nucleotides are present at the 5’ and 3’ termini, respectively (24, 43). Computer analysis of the first ORF (1a) encoding a 349 kDa protein revealed the presence of two protease domains at the amino-terminal end, followed by methyl transferase and helicase domains. The latter two domains are separated by a region of unknown function (24). A +1 ribosomal frameshift at the end of ORF 1a...
is proposed to be involved in the translation of ORF 1b encoding a 57 kDa protein, with sequence homology to RNA dependent RNA polymerase (RdRp) (24). ORF 2 encodes a 33 kDa protein of unknown function. This is followed by a five-gene block unique to closteroviruses: a 6 kDa membrane-associated protein, a 65 kDa protein with a HSP70-like domain, a 61 kDa protein of unknown function, and two CPs. In beet yellows virus (BYV), another closterovirus, this gene quintet has been proposed to be involved in virus movement (48). At the 3' end there are four ORFs potentially encoding p18, p13, p20 and p23, the functions of which are not clear (24).

Development of a protoplast system for CTV replication (38) and a full length cDNA clone of CTV which replicates in *Nicotiana benthamiana* protoplasts (54) have facilitated the genetic analysis of CTV genome. Experiments using deletions of different regions of the CTV genome showed that only the first ORF and the UTRs at the genome termini are required for its replication in protoplasts (54).

Besides T36, the genomes of three other CTV isolates, varying in size from 19,226 to 19,296 nucleotides, have also been sequenced: VT, a stem pitting and seedling yellows isolate from Israel (30); T385, a mild isolate from Spain (63) and SY 568, a stem pitting and seedling yellows isolate from California (66). The sequences were most conserved among all the isolates at the 3' UTRs (above 97% identity) and least conserved at the 5' UTRs (below 50% identity among some isolates) (26). Three of the CTV isolates, VT, T385 and SY 568, exhibit only about 10% or lower diversity throughout the genome. Their relationship with T36 was asynchronous with the 3'-half of the genome of T36 showing only about 10% diversity from other isolates, while the 5'-half exhibited about 30% diversity from the other three isolates. The sequences in the 5'-half of the genome of two populations, designated T2K and T38K, from a grapefruit stem pitting isolate (UF3800), showed an interesting relationship with others (Fig. 2). The sequence of T2K showed about 10% diversity from T36, while T38K showed about 30% diversity from both the T36-like group and the VT-like group, thus forming a third group (29). Based on the comparison of 5' UTR sequences from several isolates, Lopez et al. (26) classified the isolates into three groups: T36, VT and T317-8-like. Although the primary sequence varies significantly in the 5' UTR region, a secondary structure consisting of two stem loops that may have functional significance is predicted to be present in all the isolates (26).

**The viral proteins.** Expression of CTV proteins occurs through several mechanisms including subgenomic RNAs (sgRNAs), ribosomal frameshifting and proteolysis by viral and possibly host-encoded proteases resulting in at least 17 proteins (Fig. 1A). A series of 3' co-terminal sgRNAs (Fig. 1B) are present in infected plants (21) and

![Fig. 2. An unrooted tree based on the sequences of the ORF 1a of different citrus tristeza virus isolates. The tree was generated by using Clustal W (61) and TreeView (42) sequence analysis programs. The scale at the bottom represents a unit of branch length representing 0.1 nucleotide substitutions per site.](image)
allow translation of their respective 5'-proximal ORFs. The sgRNAs vary in infected plants with those of p20 and p23 being expressed at higher levels, followed by the two CP genes (21, 45). The aforementioned sgRNAs were expressed at higher levels in infected protoplasts as well (38). The sgRNAs of p20 and p23 were shown to have a 5' UTR of 48 and 38 nucleotides respectively, and were colinear with the respective upstream sequences of the genome (23). The sgRNAs are found as dsRNAs in abundant quantities in plants, and have also been detected from virion preparations indicating their encapsidation (21).

Defective RNAs. CTV infected plants were found to contain a series of dRNAs ranging in size from ca. 1.5 kb to over 10 kb (3, 30, 31, 33, 64). Available dRNA sequences suggest that they originate by a single internal deletion of the genomic RNA and contain varying lengths of the 5' and 3' termini of the genome. Since these dRNAs lack complete replicase genes, their replication in cells must be assisted in trans by the replication-associated proteins translated from the full length genomic RNA (helper). Usually one or a few species of defective RNAs are predominant in an infected tissue. The seedling yellows symptoms caused by the VT isolate from Israel were found to be associated with a smaller dRNA of either 2.4 or 2.7 kb, while the non-symptomatic and recovered plants showed a dRNA of 4.5 kb (65). Even though, their 5' termini differ widely, it has been demonstrated that helper RNAs (from partially purified virus) can replicate dRNAs of heterologous origin (33). A minus strand jumping model and a role for sgRNAs was proposed by Yang et al. (64) since most of the dRNAs analyzed in this study had varying lengths of 5' terminal sequences from the helper and the complete sequence from the sgRNA of ORF 11 at the 3' termini. Ayllón et al. (3) found a direct repeat of four to five nucleotides in the flanking region or close vicinity of the junction regions of the dRNAs. Sequence analysis of encapsidated dRNAs suggests that the origin of CP assembly must reside within about 1,200 bases in the 5' and/or 450 bases in the 3' termini of the genome (31).

A short 0.8 kb positive sense ssRNA, corresponding to the 5' terminal part of the genome, designated LMT 5'-RNA, of unknown function was found in several CTV isolates of diverse geographical origin (31, 33).

Virus populations in infected plants. The perennial nature of Citrus and multiple infections by
aphids over the years results in a mixture of several CTV genotypes in plants. Infected plants contain both ssRNAs and dsRNAs of gRNAs, sgRNAs and dRNAs of several CTV genotypes. In addition, each of these genotypes may be present as quasispecies, and recombination between the genotypes resulting in either functional or dRNAs is possible. The virus population also is known to change with host, temperature, aphid transmission and several other factors. The dynamic composition of CTV populations in plants presents problems in taxonomy, in understanding the mechanisms of disease, strain differentiation, strain interactions, and in designing effective control measures.

Considering the complexity of viral RNA populations in infected plants, a brief review of the current understanding of the concept of virus species, genotypes and quasispecies is appropriate. Genotype and quasispecies refer to relatedness at the molecular and/or genetic level. Differences between genotypes are relatively large, but quasispecies represent minor molecular variants, usually with only 1% or 2% nucleotide heterogeneity (9). The relatedness of genotypes is studied by phylogenetic analysis, which uses a variety of mathematical and statistical models to define genotype differences based on predicted changes within nucleotides. For example, hepatitis C virus (HCV) is classified into six genotypes and at least 80 subtypes, with some genotypes showing less than 70% nucleotide identity among themselves (56). The concept of species in viruses is much different than for other living groups. A virus species is defined as ‘a polythetic class of viruses that constitutes a replicating lineage and occupies a particular ecological niche’ (62). The criteria considered for grouping a virus into species and strains are not the same for all groups (34). The range of variability in nucleotide sequence of the genome within a virus species ranges from about 10% in potyviruses, 15% in geminiviruses (62) to over 30% in HCV (57). The variability in CTV sequences is comparable to that described in HCV. Despite such vast differences among different CTV isolates, similarities in the genome organization, host range, serology and aphid transmission indicate relatedness of these isolates and justify their grouping as a single species.

Virus quasispecies are closely related, but not identical, mutant and recombinant viral genomes subject to continuous genetic variation, competition and selection (12). This concept was proposed to describe populations that replicate with high error rate. Description of a plant virus as a single sequence (e.g., from a cDNA clone) is useful for laboratory evaluations. Infectious cDNA clones are powerful tools for genetic analysis of viruses. However, in nature, the virus population of an infected plant is not a single sequence, but a swarm of mutants that vary around a consensus sequence (52). The consensus sequence of a virus can be obtained by direct sequencing of the viral RNA and their variability by comparing the sequences of several clones (41). This variation in the viral sequences within an individual is generated by RNA polymerase errors estimated to range from less than $2.1 \times 10^{-6}$/nt/replication cycle (46) to more than $10^4$ in vesicular stomatitis virus (59). Not all mutants of a quasispecies population are infectious, and they exhibit a range of replication efficacy among quasispecies (60). The relationship between mutation rate and mutation frequency depends on the viral polymerase and the direction and strength of selection (12). Most mutations are probably lethal or possess lower fitness because of the sequence and structural requirements in the virus life cycle in a particular environment. The swarm of
mutants in an infected plant will center around one or a few fitness peaks. Fitness is defined as the relative reproductive ability in a defined environment. A major shift in the swarm would be predicted to occur when the environment changes. Such environmental changes may include vector populations, host varieties, temperature, etc. Changes in quasispecies population have been recorded with environmental changes (12, 27). There is increasing evidence to show that quasispecies evolution may lead to the emergence of new virulent genotypes and new viral pathogens (13). Confusingly, ‘quasispecies’ has sometimes been used to distinguish an isolate from that present in other infected individuals or to describe the variants in an individual infected from multiple sources (see references in 58). Since CTV-infected plants often contain a mixture of several genotypes and probably each of these exists as quasispecies, careful analysis is necessary to differentiate between genotypes and quasispecies.

The presence of mixtures of CTV genotypes in infected plants has been demonstrated earlier by several methods including single aphid transmission studies (6), symptomatology and dsRNA patterns of graft-subinoculations (37). Sequencing (29), hybridization with sequence specific probes (7, 22), single-stranded conformational polymorphism SSCP (4, 53), PCR colony hybridization techniques (18) and other techniques have shown the presence of this variability in infected plants. Dramatic shifts in complexity of stably maintained CTV populations upon transfer to a different host or aphid transmission have been observed (2).

How does CTV show sequence conservation of about 90% at one half of the genome and diversity up to about 30% or higher at the other half? Did these diverse genotypes evolve from a common ancestor under different selection pressures? What is the role of recombination in generating either defective or chimeric full-length replication-competent RNAs? These are important considerations in understanding the biology of the virus, designing molecular tools for strain differentiation and understanding the phenomenon of cross protection.

The relationship of the sequence of T36 with three other isolates showing about 90% identity at the 3'-half and less than 70% at the 5'-half suggested that the T36 isolate probably originated by a recombination event between the 3'-half of the major genotype and the 5'-half of an unknown virus type (30). Two distinct consensus sequences (T2K and T38K) from isolate UF3800 showed about 10% and 30% divergence, respectively, from T36 in the 5'-half of the genome. T38K showed about 30% sequence divergence from both T36 and the VT groups (29). These observations indicate that the 3'-half of the genome is probably more conserved and that the 5'-half of the genome has evolved into highly divergent sequences, probably under different selection pressures. About 10% or higher diversity, as seen among most of the genomic sequences, would be sufficient to classify them as distinct genotypes. It is possible that recombination occurs between different genotypes in the 3'-half of the genome resulting in replication-competent new isolates. Because of higher levels of identity in the 3'-half of the genome, analysis of possible recombination in this region should be done with genotype specific cDNAs. The presence of dRNAs (23, 33) suggests strand jumping ability of the CTV replication complex. It has also been demonstrated that certain CTV isolates can recognize heterologous termini for generation of dRNAs (33). Whether such recombination can also result in full length chimeric and functional sequences remain to be investigated.
Initially, molecular characterization of CTV isolates was done using mostly CP gene sequences from clones of different isolates, (28, 32, 44). Analysis of populations as opposed to individual clones of CTV has been emphasized in later studies. In addition to the CP gene, several other genes at the 3'-end of the genome and partial sequences from the 5'-end have been studied. Sequencing, restriction analysis of PCR products of CP gene (19), SSCP of several 3'-end genes (4, 53) and hybridization with group or genotype-specific probes (7, 18, 22) have been used in such studies.

Phylogenetic analysis of CP sequences from several CTV isolates placed the mild, quick decline and stem pitting isolates into distinct groups (Fig. 3). Cevik (7) used short biotinylated and group specific oligonucleotide probes to detect mixtures of CTV by hybridization to PCR amplified CP genes. A serological assay has been developed to detect grapefruit stem pitting isolates (40). Gillings et al. (19) used PCR, restriction analysis and sequencing of CP genes to demonstrate mixtures of CTV genotypes and their segregation by aphid transmissions. Restriction analysis of the PCR products of CP gene and sequencing have been used to demonstrate mixtures of two genotypes, probably belonging to mild and severe (Capão Bonito) isolates of CTV in Brazil (M. L. P. N. Targon, pers. comm.). Genc (18) screened several clones of CP genes from biologically distinct isolates by hybridization with group-specific oligo-nucleotide probes, followed by sequencing and detected heterogeneous genotypes occurring in low levels in plants that are otherwise not easily detectable. For example, "PCR-cloning-colony hybridization" (PCCH) of the B23

Fig. 3. An unrooted tree based on the coat protein gene sequences of several CTV isolates showing the association of different groups with biological activity. The tree was constructed using the neighbor-joining method from the PHYLIP package (evolution.genetics.washington.edu/phylip.html) and TreeView (42) computer programs.
isolate of CTV from Israel showed a mixture of three genotypes including T36 and T30-like populations (18). The presence of these genotypes was confirmed by sequencing. In an earlier study, aphid transmissions of B23, a severe isolate (reacts strongly to MCA-13, a monoclonal antibody) yielded subisolates that were mild and MCA-13 negative, characteristics similar to T30 isolate (S. Garnsey, pers. comm.).

Sequencing, hybridization with genotype specific probes (7) and PCCH (18) have shown the presence of Florida T30 and T36-like sequences (CP gene) in samples from Mexico, Colombia, Spain, Israel, and South Africa (19, 47, L. Marais, pers. comm.). Two of the mild strains, T385 and T30 exhibiting similar genomic sequences (63) were obtained from two geographically different regions of Spain and the USA respectively, and are separated from each other by at least several decades. Widespread occurrence of these CTV genotypes indicates that probably they have spread along with the movement of Citrus. The dominance or suppression of a genotype in a virus population in any region is probably determined by the complexity of local CTV population, hosts (rootstocks and scion varieties), vector population and environmental factors. Any change in one of these factors may lead an altered situation. As an example, introduction of *T. citricida* has led to several tristeza epidemics (50).

A great deal of information on the molecular biology of CTV was generated in the last 5 to 6 yr. Progress has been made in our understanding of genome structure, organization and expression. Potential breakthroughs in the areas of reverse genetics using the infectious cDNA clone, development of highly sensitive strain specific probes and control methods by either improved cross protection or development of resistant transgenic plants are likely to occur in the near future.

**ACKNOWLEDGEMENTS**

This paper is Florida Agricultural Experiment Station Journal Series No. R-07484.

**LITERATURE CITED**


30. Mawassi, M., E., Mietkiewska, R. Gofman, G. Yang, and M. Bar-Joseph


32. Mawassi, M., R. Gafny, and M. Bar-Joseph


34. Mayo M. A. and C. R. Pringle

35. Medina, V., V. V. Peremyslov, Y. Hagiwara, and V. V. Dolja

36. Moreno, P., J. Guerri, and N. Muñoz

37. Moreno, P., J. Guerri, J. F. Ballester-Olmos, and M. E. Martinez


41. Odeberg, J., Z. Yun, A. Sonnerborg, M. Uhlen, and J. Lundeberg

42. Page, R. D.


44. Pappu, H. R., S. S. Pappu, C. L. Niblett, R. F. Lee, and E. Civerolo


46. Parvin, J. D., A. Moscona, W. T. Pan, J. M. Leider, and P. Palese


48. Peremyslov, V. V., Y. Hagiwara, and V. V. Dolja


