Genetic Variation of *Citrus Tristeza Virus* Isolates from California and Spain: Evidence for Mixed Infections and Recombination

Luis Rubio,[1] María Angeles Ayllón,[2]† Ping Kong,[1] Andres Fernández,[2]
Marylou Polek,[3] José Guerri,[2] Pedro Moreno,[2] and Bryce W. Falk[1]*

*Plant Pathology Department, University of California, Davis, California 95616[1]; Instituto Valenciano de Investigaciones
Agrarias, Valencia, Spain[2]; and Central California Tristeza Eradication Agency, Tulare, California 93224[3]*

We examined the population structure and genetic variation of four genomic regions within and between 30 *Citrus tristeza virus* (CTV) isolates from Spain and California. Our analyses showed that most isolates contained a population of sequence variants, with one being predominant. Four isolates showed two major sequence variants in some genomic regions. The two major variants of three of these isolates showed very low nucleotide identity to each other but were very similar to those of other isolates, suggesting the possibility of mixed infections with two divergent isolates. Incongruencies of phylogenetic relationships in the different genomic regions and statistical analyses suggested that the genomes of some CTV sequence variants originated by recombination events between diverged sequence variants. No correlation was observed between geographic origin and nucleotide distance, and thus from a genetic view, the Spanish and Californian isolates analyzed here could be considered members of the same population.

Citrus tristeza virus (CTV) is distributed worldwide and is the causal agent of one of the most economically important diseases of citrus. CTV, a member of the genus *Closterovirus* within the family *Closteroviridae*, is phloem limited and is transmitted by aphids in a semipersistent manner. CTV virions are filamentous flexuous particles about 2,000 nm long, with two coat proteins (CP and CPm) covering 95 and 5% of the particle length, respectively (8). The CTV genome is a single-stranded, positive-sense RNA of 19,226 to 19,296 nucleotides (nt) (18, 27, 48, 51) organized in 12 open reading frames encoding at least 19 proteins. These include two papain-like proteases, replication-associated proteins (RNA polymerase, helicase, and methyltransferase), a homologue of the HSP70 protein, two coat proteins (CP and CPm), RNA-binding protein p23 (23), a p20 protein that accumulates in the amorphous inclusion bodies (14), and other proteins of so far unknown function (p61, p13, and p18) (Fig. 1). CTV-infected plants contain, in addition to the genomic RNA, 3′-coterminal subgenomic RNAs (15) and defective RNAs (D RNAs), the latter resulting from extensive internal deletions of the genomic RNA (2, 26, 28, 50).

CTV isolates differing in the type and intensity of symptoms induced in different citrus species and cultivars and in their aphid transmissibility have been reported worldwide (38). In the last two decades, efforts have been taken to develop molecular techniques for rapid differentiation of CTV isolates and identification of molecular markers related to CTV-induced symptoms. Variation in serological reactivity, peptide maps of the coat protein, double-stranded RNA (dsRNA) patterns, hybridization with cDNA probes, restriction fragment length polymorphism, and single-strand conformation polymorphism

(SSCP) have been described in attempts to differentiate CTV isolates (29).

Nucleotide sequence analysis is the most accurate procedure for CTV differentiation and estimation of molecular or genetic variation. To date, the complete genome nucleotide sequences of the five CTV isolates T36 and T30 from Florida (1, 18, 34), VT from Israel (27), SY568 from California (51), and T385 from Spain (48) have been reported. Also the partial nucleotide sequences of several CTV isolates have been reported (1, 17, 22, 25, 35, 36). Recently, it has been shown that individual CTV isolates are composed of a population of sequence variants (3, 19, 22). These reports showed the genetic differences between CTV isolates but did not estimate the genetic diversity of natural populations of CTV. Only Moya and García-Arenal (31) estimated the genetic diversity of CTV in Spain based on the number and position of D RNAs associated with Spanish CTV isolates (14). However, for CTV, nucleotide identity seems not to be correlated with the similarity of D RNA patterns (2).

In this study, we assessed the structure and genetic diversity of four genomic regions from two natural CTV populations located in two important citrus-growing areas, Spain and California. By SSCP analysis we estimated the population structure of sequence variants within individual isolates. Nucleotide analysis was used to estimate the genetic distance between sequence variants. These analyses open new insights about mixed infections, recombination, and spatial population structure in the effort to understand CTV complexity.

MATERIALS AND METHODS

Virus isolates. Nineteen Californian and 11 Spanish CTV isolates were obtained from field trees and maintained in sweet orange (*Citrus sinensis*) plants in insect-proof greenhouses. Symptom evaluation was performed for these isolates in sweet orange plants and in graft-inoculated Mexican lime (*Citrus aurantiifolia*) and grapefruit (*Citrus paradisi*) plants. Geographic origin and symptoms induced by these CTV isolates are summarized in Table 1. More detailed information about the biological characteristics of CTV Spanish isolates can be found in Ballester-Olmos et al. (4).

---

* Corresponding author. Mailing address: Department of Plant Pathology, One Shields Ave., University of California–Davis, Davis, CA 95616. Phone: (530) 752-0302. Fax: (530) 752-5674. E-mail: bwfalk @ucdavis.edu.
† Present address: Citrus Research and Education Center, Department of Plant Pathology, University of Florida, Lake Alfred, FL 33850.
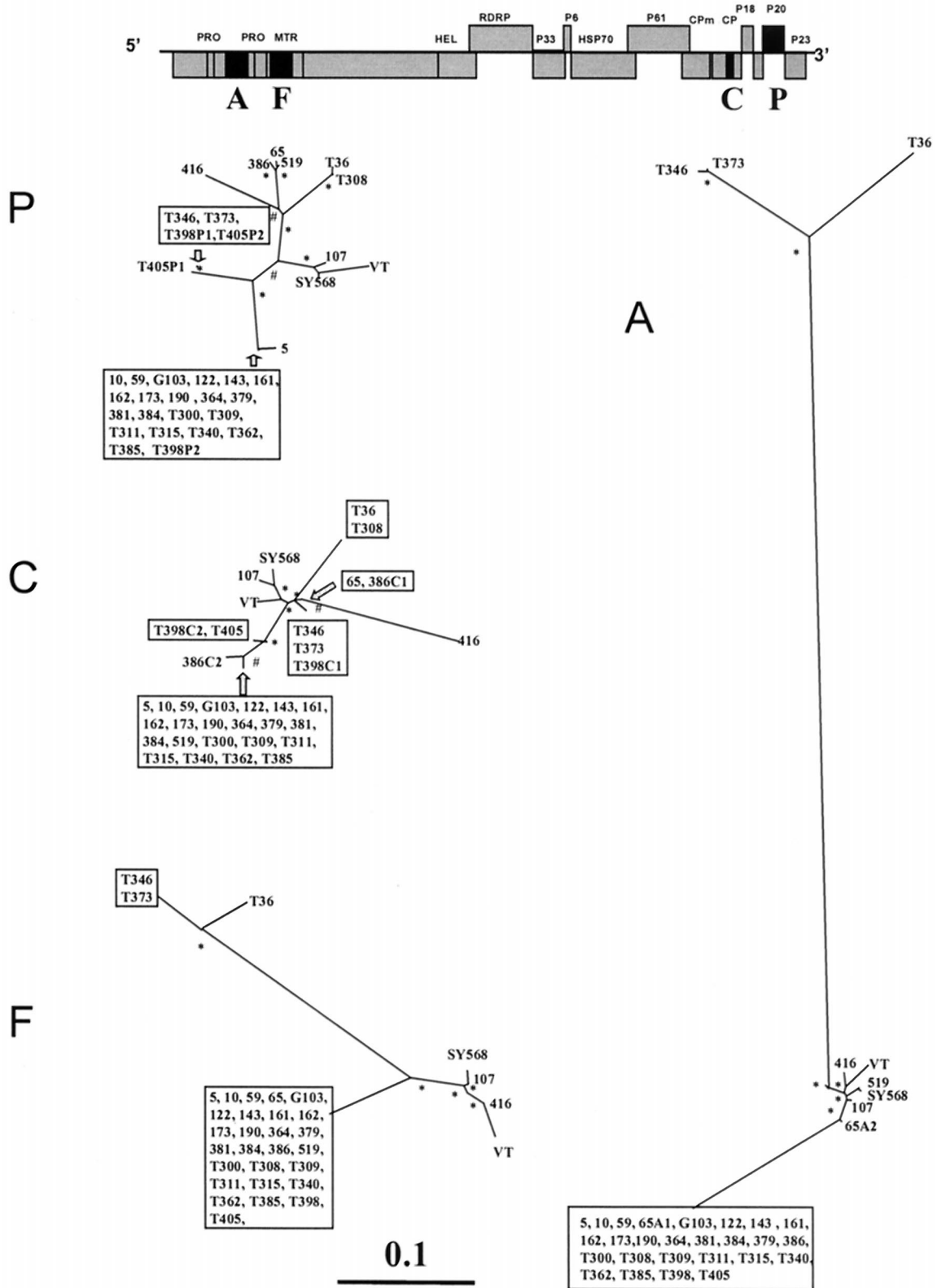
FIG. 1. Unrooted maximum-likelihood phylogenetic trees of genomic regions A, F, C, and P (see also Table 2) of 34 CTV isolates (Table 1), constructed using the PHYLIP program DNAML. Bootstrap values of between 600 and 800 for 1,000 replicates are indicated by #, and values greater than 800 are indicated by *. Branch lengths are proportional to the genetic distances. Boxes include sequences with a nucleotide identity of greater than 99%. Above is a layout of the CTV genome, with the regions analyzed in black boxes. When an isolate contained more than one major sequence, these are indicated by a letter (corresponding to the genomic region) and a number (corresponding to the sequence variant), e.g., 65A1 and 65A2 occurred in genomic region A of isolate 65.

TABLE 1. Origin and biological characterization of CTV isolates

Origin	Isolate no.	Specific origin (city, county or province)	Symptoms induced[a]		
			Mexican lime	Sweet orange	SY
California	5	Lindsay, Tulare	2	0	−
	10	Exeter, Tulare	3	0	−
	59	Edison, Kern	2	0	+
	65	Exeter, Tulare	0	0	−
	G103	Ventura	N	2	+
	107	North Exeter, Tulare	2	2	−
	122	Exeter, Tulare	2	1	−
	143	Mc Farland, Kern	N	0	+
	161	Fresno	N	0	+
	162	McFarland, Kern	N	0	+
	173	Delano, Kern	N	0	+
	190	McFarland, Kern	1	0	−
	364	Delano, Kern	N	0	+
	379	Lindsay, Tulare	N	0	+
	381	Lindsay, Tulare	N	0	−
	384	Hemet, Riverside	N	0	+
	386	Riverside City, Riverside	N	0	+
	416	Fresno	N	N	N
	519	Riverside	N	0	−
Spain	T300	Puebla Larga, Valencia	1	0	−
	T308	Burjasot, Valencia	3	0	+
	T309	Almenara, Castellon	3	0	−
	T311	Xeresa, Valencia	1	0	−
	T315	Huercal, Almeria	1	0	−
	T340	Godella, Valencia	2	0	−
	T346	Brenes, Sevilla	2	0	−
	T362	Tortosa, Tarragona	3	0	−
	T373	Alhama, Murcia	2	0	−
	T398	Alcira, Valencia	2	2	−
	T405	Alcira, Valencia	2	0	−

[a] Symptoms induced by CTV isolates in different hosts: Mexican lime (vein clearing and stem pitting) and sweet orange (stunting and stem pitting). Symptom intensity is scored from 1 to 3, with 1 being mild and 3 very severe. SY, seedling yellows reaction in sour orange and grapefruit; − and +, absence and presence of symptoms, respectively; N, characterization not done.

Purification of CTV dsRNA. CTV-infected sweet orange bark was pulverized with nitrogen liquid, and total nucleic acids were extracted with phenol-detergent buffer. The dsRNAs were then purified by column chromatography on nonionic cellulose and precipitated as previously described (30).

cDNA synthesis and PCR amplification. Four pairs of primers (A, F, C, and P) were designed for performing reverse transcription (RT)-PCR of different regions of the CTV genomic RNA (Table 2). Approximately 200 ng of dsRNA was denatured at 95°C for 5 min and reverse transcribed by incubation at 42°C for 45 min in a reaction mixture (20 μl) containing 1× avian myeloblastosis virus (AMV) buffer, 200 μM each of the four deoxynucleoside triphosphates (dNTPs), 40 ng of reverse primer, 2 U of RNasin (Promega Corp.), and 0.3 U of AMV reverse transcriptase. An aliquot (1/10) of cDNA product was PCR amplified in

a 20-μl reaction mixture containing 1× PCR buffer, 1.5 mM MgCl$_2$, 200 μM each dNTPs, 20 ng of each primer, and *Taq* DNA polymerase (Promega). The following PCR conditions were used: 94°C for 2 min; 30 cycles each of 94°C for 30 s, 50°C for 30 s, and 72°C 40 s; and 72°C for 5 min. The resulting RT-PCR products were separated by electrophoresis in a 2% agarose gel and detected by ethidium bromide staining.

Cloning and SSCP analysis. The cDNAs obtained by RT-PCR were cloned into PGEM-T (Promega) using T4 DNA ligase (Promega) according to the manufacturer's instructions, followed by transformation into *Eschericia coli* DH5α (44). Ten clones obtained from each cDNA product were selected and PCR amplified using the same conditions as the PCR described above. SSCP analysis was performed on the resulting PCR products as previously described (19, 40). Denatured samples were electrophoresed in nondenaturing 8% polyacrylamide gels at 200 V for 3 h (genomic regions A and F), 2 h (region P), or 1 h (region C).

Nucleotide sequences and statistical analysis. Nucleotide sequences of cDNAs were determined in both directions by means of an ABI Prism DNA sequencer 377 (Perkin-Elmer, Foster City, Calif.). Multiple alignments of the nucleotide sequences were realized with the program CLUSTAL W (47). Numbers of synonymous and nonsynonymous nucleotide substitutions were determined using the program DIVERGE in the Genetics Computer Group (GCG) package (7), which is based on the method described by Pamilo and Bianchi (33) and Li (21) (PBL method). Nucleotide identity of pairs of sequences over the entire sequence length was plotted using GCG's program PLOTSIMILARITY (7). Nucleotide distances were estimated by using the DNADIST program of PHYLIP package version 3.573 (9) using the Jukes and Cantor method for correction of superimposed substitutions. Phylogenetic relationships were inferred using the PHYLIP programs DNAML, based on the maximum-likelihood method, and NEIGHBOR, which implements the neighbor-joining method from a nucleotide distance matrix. SEQBOOT (1,000 repetitions) and CONSENSE were used for bootstrap analysis. Genetic diversity (average number and standard errors of the number of nucleotide substitutions for each pair of sequence variants) and the degree of population subdivision were calculated by the method of Lynch and Crease (24). Recombination events between diverged nucleotide sequences were explored with the programs GENECONV (45) and PHYLPRO (49).

Nucleotide sequence accession numbers. The nucleotide sequence data reported in this paper have been deposited in the GenBank database under accession numbers AF356226 to AF356329. Other CTV nucleotide sequences used in our analyses were obtained from the indicated GenBank entries: isolate T36 (U16034), isolate VT (U56902), isolate T385 (Y18420), isolate SY568 (AF001623), isolate T30 (AF260651), Japanese isolates (AB011185 to AB011197), Portuguese isolates (AF184113 to 184118), and Californian isolates 65, 107, 122, 173, and 190 (AF203035 to AF203037, AF203040, AF203043, AF203047, AF203048, AF203051, AF203053, AF203055, AF203057, AF203060, AF203062, AF203063, AF203066, AF203070, AF203073, AF203075, AF203077, AF203079, and AF203081).

RESULTS AND DISCUSSION

Genetic variation within CTV isolates: evidence of mixed infections. RNA viruses have a great potential for genetic variation due to their error-prone RNA replication, large populations, and short replication times. As a consequence, each

TABLE 2. Primers designed for PCR amplification of four regions of CTV genomic RNA

Primer[a]	Nucleotide sequence	Genomic region	Positions[b]	Size[c] (nt)
A-forward	5′-ACGTGTTCGTGAAACGCGG-3′	A	2021–2039	528
A-reverse	5′-GTCGATAACTCGACAAACGAGC-3′		2527–2548	
F-forward	5′-GTGTTATCATGCGTCTGAAGCG-3′	F	3561–3582	438
F-reverse	5′-GGAATCTTAATCCTAATCAAG-3′		3978–3998	
CP-forward	5′-AACGCCCTTCGAGTCTGGGGTAGGA-3′	C	16515–16539	273
CP-reverse	5′-TCAACGTGTGTTGAATTTCCCAAGC-3′		16763–16787	
P20-forward	5′-ACAATATGCGAGCTTACTTTA-3′	P	17720–17740	561
P20-reverse	5′-AACCTACACGCAAGATGGA-3′		18262–18280	

[a] Oligonucleotide primers used. Forward indicates primers complementary to the negative strand, and reverse indicates those complementary to the positive strand.
[b] Nucleotide positions of the primers in the genomic RNA of CTV isolate T385 (48).
[c] Expected size of the RT-PCR products.

isolate of RNA virus is expected to consist of a population of genetically related variants, termed quasispecies (16). To estimate the within-isolate population structure, we performed RT-PCR of four genomic regions (Table 2) of 11 Spanish and 19 Californian CTV isolates and analyzed the SSCP patterns of 10 clones obtained from each genomic region and isolate. Previously we assessed the accuracy of these SSCP analyses (19). The average number of nucleotide differences per site between pairs of CTV sequence variant clones was only 0.001346 ± 0.002126, supporting SSCP analysis as a precise tool for population studies. To minimize primer-directed selection of sequence variants within a isolate, the primers used here (Table 2) were designed from nucleotide sequences conserved for the five CTV isolates whose complete genome sequences are known (1, 18, 27, 34, 48, 51). Previously, we also found that the intensities of the DNA bands in SSCP profiles reflected the relative proportion of CTV RNA variants within an isolate, suggesting that these primers do not bind preferentially to some CTV sequence variants (42). Furthermore, to minimize the possibility that nucleotide incorporation errors in the initial phases of the RT-PCR give rise to a detectable subpopulation of false mutants (20), we used as a template a large amount of dsRNAs (~200 ng). Nonetheless, because we analyzed only the major sequence variants, RT-PCR-induced errors, if present, did not affect our analysis. Finally, to minimize possible variations due to irregular distribution of sequence variants in infected tissues, the dsRNAs were extracted from different branches of a large plant and pooled.

Our results and the results reported previously (19) showed that 26 of 30 field CTV isolates, in the four genomic regions analyzed, had a within-isolate population structure consisting of one major variant (frequency greater than 0.7) and other sequence variants of lower frequencies (minor variants), a typical quasispecies structure (16). In contrast, genomic region A of isolate 65 (19), genomic region C of isolates T398 and 386, and genomic region P of isolates T398 and T405 had two major variants with frequencies of 0.4 or greater. SSCP frequencies and nucleotide distances for all variants from Californian isolates 65, 107, 122, 173, and 190 have been described in detail by Kong et al. (19). We determined the nucleotide sequences of all CTV major variants and estimated the nucleotide distances between them. The nucleotide distances between sequences are represented in Fig. 1 as branch lengths of phylogenetic trees (see next section). Interestingly, for 65 A, T398 and 386 C, and T398 P, the nucleotide distances between the two major variants within the same isolate were high (Fig. 1). For example, the nucleotide distance of genomic region P between the two major variants of isolate T398 (T398P1 and T398P2) was 0.0795. Furthermore, within these isolate regions, each major variant was significantly more similar to major variants of other isolates than to the other major variant from the same isolate (Fig. 1). For example, the major variant T398P2 and that of isolate T300 showed a nucleotide distance of 0.0019, whereas T398P1 and T346 had a nucleotide distance of 0.0039. These data suggest that 65 A, T398 and 386 C, and T398 P could have originated from mixed infections of two CTV isolates with diverged sequence variants. Also, the fact that some genomic regions in the same isolate have two diverged major variants whereas other genomic regions have only one suggests the

TABLE 3. Average number of nucleotide substitutions between CTV isolates in different genomic regions (A, F, CP, and P20)[a]

Region	D	dN	dS	dN/dS
A	0.13656 (0.04209)	0.11125	0.20472	0.56323
F	0.05993 (0.01601)	0.03455	0.29849	0.12302
C	0.03792 (0.00626)	0.00511	0.25128	0.02739
P	0.06029 (0.00763)	0.03221	0.25093	0.13424

[a] D, nucleotide diversity: average number of nucleotide substitutions per site between pairs of sequences. Standard errors are indicated in parentheses. dN, average number of nonsynonymous substitutions per nonsynonymous site. dS, average of synonymous substitutions per synonymous site. dN/dS, average of the ratio between nonsynonymous and synonymous substitutions for each pair of comparisons. dN, dS, and dN/dS were estimated by the PBL method (21, 33).

possibility of recombination events between sequence variants of the original coinfecting CTV isolates.

Mixed infections are possible, as CTV hosts are long-lived perennial plants (some living 100 years or more), allowing the possibility of repeated inoculations of CTV by viruliferous aphids. Although we have evidence for mixed infections in a small proportion of CTV isolates here analyzed (3 of 30), it is probable that mixed CTV infections occur more frequently in nature. With the approaches used here, mixed infections could only be detected between two isolates with diverged sequence variants, and both sequences should be in relatively high proportion when they were sampled. The coexistence of two quasispecies in the same host is probably not common, as differences in fitness would likely cause the displacement of one quasispecies.

Genetic variation in different genomic regions: evidence for natural recombination events. The major variant sequences from the CTV isolates analyzed here and the sequences of CTV isolates T36, VT, T385, and SY568 (18, 27, 34, 48, 51) were used to estimate the nucleotide diversity (average number of nucleotide substitutions per site in each pair of sequence variants) for each genomic region. These analyses showed appreciable differences between the different genomic regions. Genomic region A showed the greatest diversity, twice that for F and P and three times that for the C genomic region (Table 3). To estimate the degree of selective constraints on each genomic region, nonsynonymous and synonymous substitutions were computed separately. The number of synonymous substitutions per synonymous site (dS) was similar over the four genomic regions analyzed (Table 3). However, the number of nonsynonymous substitutions per nonsynonymous site (dN) was smaller than dS and varied considerably between genomic regions (Table 3). This suggests a negative selective pressure for most amino acid changes (functional constraints) and that the degree of the functional constraints varies for the different genomic regions. Genomic regions F and P had a ratio dN/dS in the range of most DNA and plant RNA virus protein-coding sequences (12, 32). F corresponds to part of the methyltransferase domain, involved in virus replication, and P corresponds to gene p20, coding for a protein of unknown function that accumulates in amorphous inclusion bodies (13). Genomic region A, which so far is not known to be part of any known functional domain, showed the greatest dN/dS ratio. Finally, C, corresponding to a portion of the coat protein gene, showed the greatest functional constraints. The CTV coat protein, in addition to constraints related to virion structure and

stability, might play a critical role in interactions with its aphid vector and/or plant host. The *dN*/*dS* ratio of CTV genomic region C was very low in comparison with that of other plant virus coat protein genes (12).

Phylogenetic relationships between the CTV isolates were inferred using neighbor-joining and maximum-likelihood methods (9). Both gave basically the same relative phylogenetic grouping and showed bootstrap values greater than 80% in the main nodes (Fig. 1). Most CTV isolate major variants showed the same grouping in the phylogenetic trees obtained from the four genomic regions analyzed here. However, some isolate major variants showed sharp differences in their genetic relationships with other CTV sequences in different genomic regions. For example, Spanish isolate T308 was genetically very close to Florida isolate T36 and very different from isolate T385 in the 3′-terminal genomic regions (C and P), but in the 5′-terminal regions (A and F) the genetic relationships between these three isolates were opposite (Fig. 1). Incongruencies in the phylogenetic relationships of different genomic regions were also observed in CTV isolates 65, 386, 519, T398, and T405 (Fig. 1). These phylogenetic incongruencies between different genomic regions for the same CTV isolate major variants suggest that some sequence variants might have originated by recombination events between diverged sequence variants.

To attempt to identify possible recombination events located within each genomic region, we used the program PHYLPRO (49). PHYLPRO displays graphically the coherence of sequence relationships (phylogenetic correlation) over the entire length of a set of aligned homologous sequences. Recombination signals appear as areas of low phylogenetic correlation, visualized by single sharp-pointed downward peaks in the graph. In Fig. 2, the phylogenetic correlation profiles of all individual CTV sequences corresponding to the four genomic regions are shown. Genomic regions A and F showed several weak recombination signals, suggesting possible recombination events that might be blurred by mutations and other recombination events that have accumulated over time. Genomic regions C and P, however, gave very complex phylogenetic correlation profiles, suggesting the possibility of multiple recombination events in different sequence positions, hindering the identification of specific recombination points between CTV sequences (Fig. 2). Because strong recombination signals might obscure the weaker ones, to improve the detection of sequences with low recombination signals, new analyses excluding sequences with regions of low phylogenetic correlation were performed. Subsequently, the removed sequences were reintroduced individually into the set of sequences. By using this procedure and by modifying the program parameters, we could locate more likely recombination events for some CTV sequences (not shown). For example, in genomic region P, several possible recombination points between an ancestor of SY568, 107, and VT and an ancestor of 519, 65, 386, and 416 (see Fig. 1) were detected.

In cases with clear recombination signals, probably reflecting recent recombination events, the display of nucleotide identity between pairs of sequences over their length provided more precise information. For example, genomic region A of the major variant 65A2 seems to have originated by recombination between SY568-like and T385-like isolates in a position between nt 40 and 80 (Fig. 3). Another clear example would be

genomic region C of the major variant T398C2. When the identity profiles of T398C2 with respect to T398C1 and T385 were displayed, it was observed that from positions 110 to 170, the three sequences show 100% nucleotide identity; T398a and T398b have 100% nucleotide identity in the segment between positions 110 and 273; and T398C2 and T385 have 100% identity between positions 1 and 170. All this suggests that the T398C2 sequence might have resulted from recombination between T398C1-like and T385-like sequences at some point between positions 110 and 170. It is possible that the complete identity between nucleotides 110 and 170 resulted from a previous recombination of the ancestor of T385 and T398C1 with another sequence variant (Fig. 3). In all cases, the recombination events inferred from the nucleotide identity profiles (Fig. 3) confirmed those obtained from the phylogenetic correlation profiles (Fig. 2). It seems unlikely that the possible recombination sites could have arisen in vitro during RT-PCR, as in our previous work no recombinant variants were observed when dsRNAs from two isolates, each with one major variant, were mixed and amplified by RT-PCR (42).

We also performed conversion analyses using the program GENECONV (45) on the complete genome nucleotide sequences of isolates T36, VT, T385, and SY568. This program is based on the analysis of whether some regions of a pair of sequences have more consecutive identical silent polymorphic sites in common than would be expected by chance. GENECONV finds and ranks fragments with the highest score (number of matches for pairs of sequences). To avoid the selection of high-scoring fragments by chance, the polymorphic sites are permuted randomly among themselves 10,000 times and scored each time. Statistical significance is evaluated by the parameter *P*, the proportion of permuted alignments for which the maximal fragment score for that pair of sequences is greater than or equal to the original fragment score. Because mutations could accumulate after the recombination event, we used the program option to allow mismatches (penalty was set for a gscale of 2). Our GENECONV analyses suggested recombination between isolates T385 and SY568, which showed a nucleotide identity of greater than 99% between positions 9304 and 16107 ($P = 0.0000$), whereas the rest was below 93%, which corroborates the results obtained by Vives et al. (48). Also, we found a possible double recombination between VT and SY568: SY568 and VT showed a nucleotide identity of about 96% in positions 1309 to 2781 ($P = 0.0027$) and 6546 to 9302 ($P = 0.0000$), whereas the rest was below 92%.

Recombination in natural populations has been reported for other plant viruses (5, 10). It has been proposed that recombination can be advantageous for RNA viruses. The high mutation rates of RNA viruses cause accumulation of deleterious mutations, limiting the RNA genome size. In large populations, fitness can be maintained or increased by natural selection, but in small populations, genetic drift will lead to progressive loss of replicative fitness. Recombination of viral genomes with deleterious mutations can regenerate functional genomes (6, 39, 46). For CTV, having the largest genome among the known single-stranded plus-sense plant RNA viruses, recombination could act as a compensatory mechanism to offset accumulation of deleterious mutations in bottleneck episodes, such as aphid transmission. Also, coinfection and recombination of different genomic regions between diverged

Genomic region A

Genomic region F
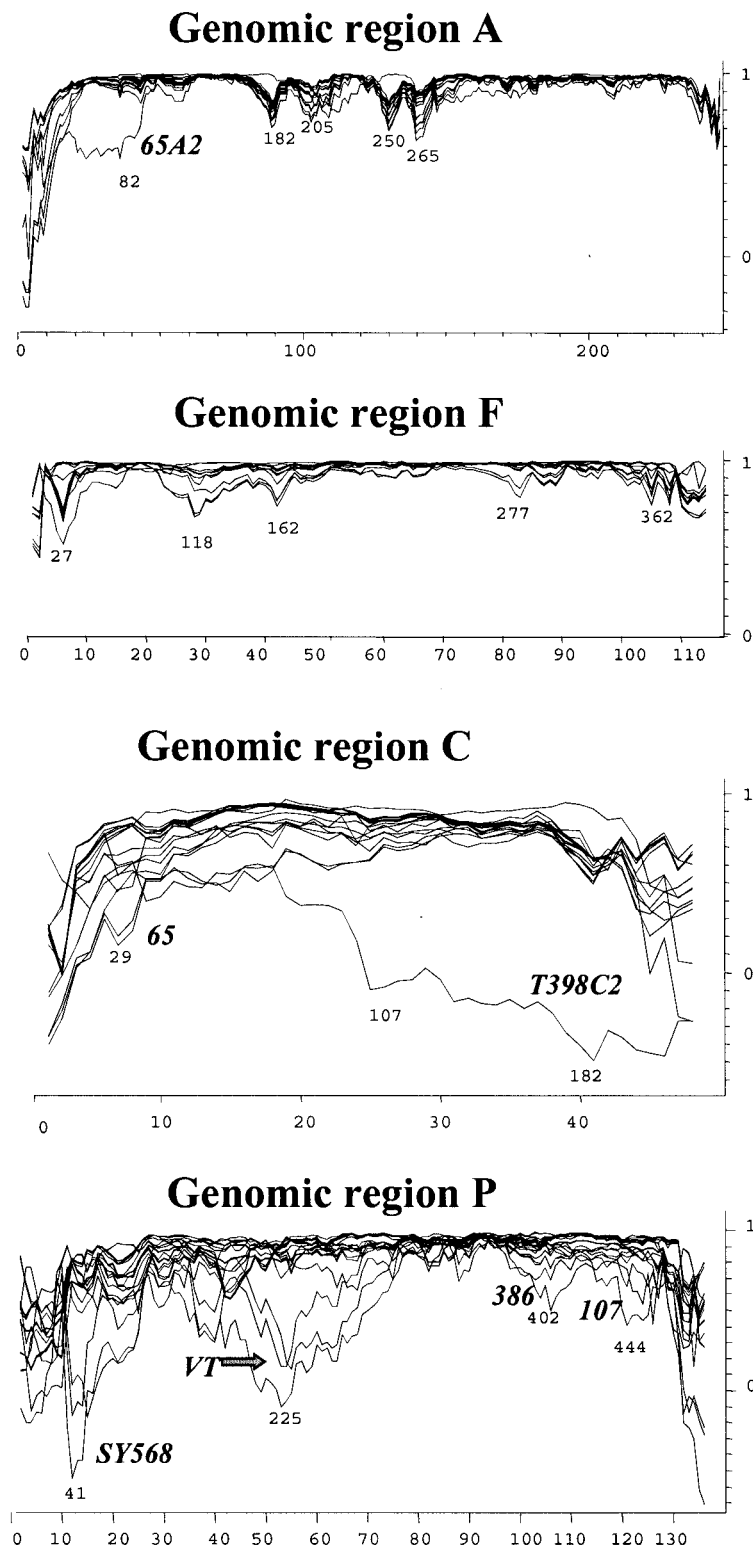
Genomic region C

Genomic region P

FIG. 2. Phylogenetic correlation profiles (graphic representation of the coherence of sequence relationships) of 34 CTV isolates (Table 1) for genomic regions A, F, C, and P (Table 2). Only the variable sites (*x* axis) are represented in the graph. Phylogenetic correlation (*y* axis) was obtained at each variable site from pairwise distance analysis of all aligned sequences by using the program PHYLPRO, with a fixed window of 40 bp. Numbers under low phylogenetic correlation areas (possible recombination signals) indicate nucleotide positions. Some CTV sequences are indicated near their individual correlation profiles.
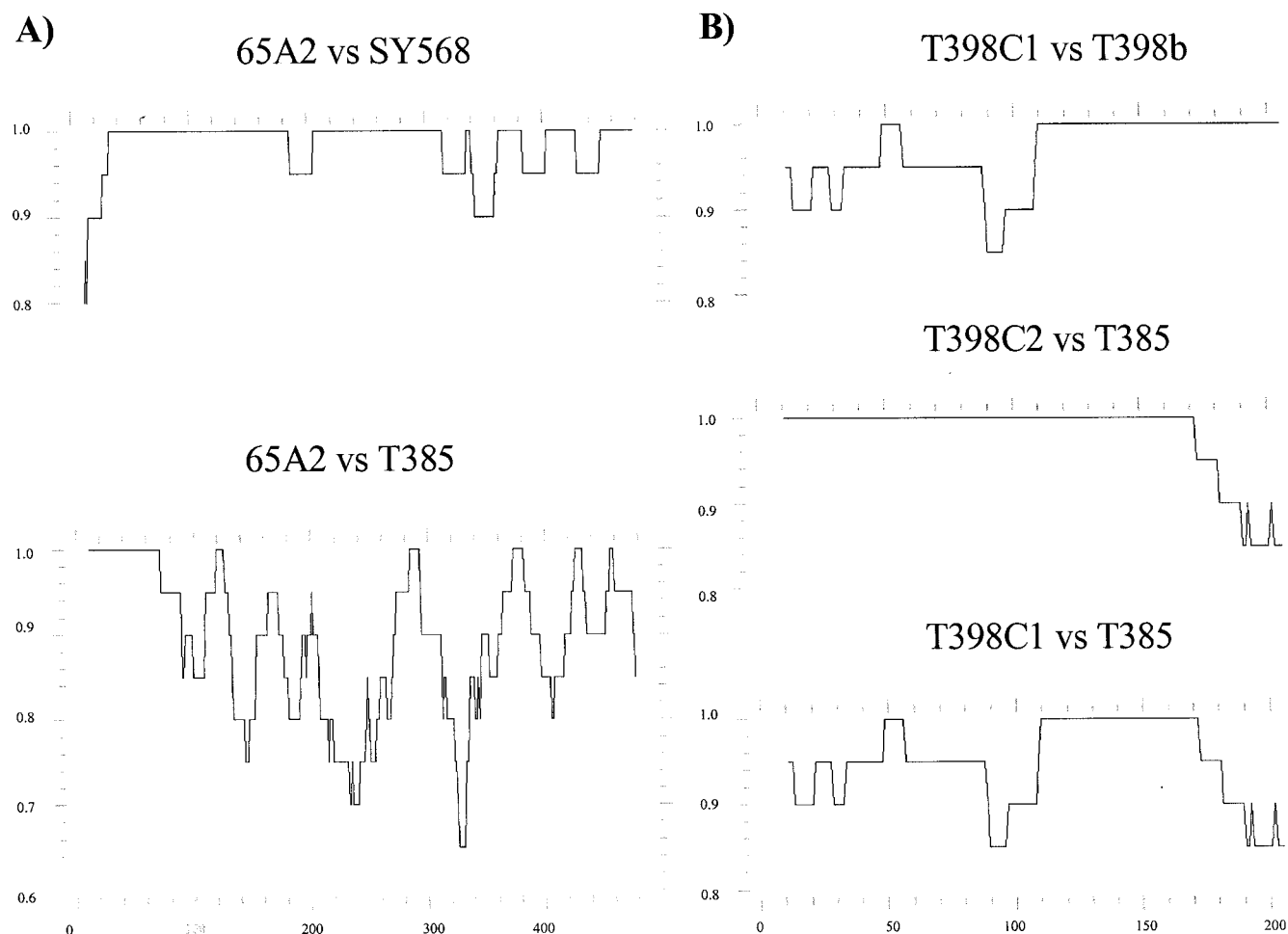
FIG. 3. Nucleotide identity profiles between pairs of CTV sequences. (A) Genomic region A, sequence variants 65A2, SY568, and T385; (B) genomic region C, sequence variants T398C1, T398C2, and T385. The profiles were constructed using the PLOTSIMILARITY program, with a window of 20 bp. The *y* axis corresponds to nucleotide identity and the *x* axis to nucleotide positions.

virus genomes allow greater genome diversity and adaptability to new environments (39, 46). Thus, recombination might explain in part the great number of CTV isolates with different biological characteristics described worldwide (38). Recombination among CTV isolates also has important practical implications. For example, for application of disease control measures, such as cross-protection or transgenic plant resistance, caution must be taken to avoid the introduction of exotic CTV sequences that might recombine and give rise to CTV isolates with new biological properties.

Genetic variation with respect to geographic distribution: absence of correlation between genetic and geographic proximity. According to CTV genomic RNA 5′ untranslated region nucleotide identity, Lopez et al. (22) classified clones from 11 CTV isolates in three groups: I, represented by isolate T36; II, represented by VT; and III, represented by T385. According to our analyses, genomic regions A and F (both located in the 5′ half of the genome; Fig. 1) of all the Spanish and Californian isolates analyzed here can be included in these three groups. The major variants of Spanish isolates T346 and T373 are included in group I, three to five Californian isolates are in group II, and about 80% of the Spanish and California isolates

belonged to group III (Fig. 1). For genomic regions C and P, the CTV isolates could not be readily assigned to these three discrete groups. Curiously, group III isolates showed a nucleotide identity of greater than 99% in the four genomic regions (Fig. 1). Group III isolates, with a very high nucleotide identity, has been also found in Florida, Taiwan, and Colombia (1).

For a better comparison of CTV populations, we estimated the nucleotide diversity (average nucleotide distance between two pairs of CTV major sequences chosen randomly) of Spanish and Californian CTV isolates (including T385 and SY568) for each genomic region (Table 4). We found that in each genomic region, the nucleotide diversity of CTV isolates from the same geographic population was remarkably higher than that assessed between the two geographic populations (Table 4). For example, the nucleotide diversity of CTV genomic region P within California and within Spain was about 0.050 for both, whereas the nucleotide diversity considering only nucleotide distances between Spanish and Californian isolates was less than 0.004 (Table 4). Application of the D statistic (24) showed that the Spanish and Californian CTV populations were not significantly genetically different ($D = 0.0000$), and hence, from a genetic view, the Spanish and Californian CTV

TABLE 4. Nucleotide diversity of geographically different CTV populations[a]

Group	Region	Nucleotide diversity				V_b				V_w	N_{st}
		Sp	Cal			Sp-Cal					
1	A	0.17627 (0.08425)	0.05809 (0.01306)			0.00819 (0.01924)				0.11718 (0.05909)	0.06496
	F	0.06825 (0.02866)	0.03328 (0.01094)			0.00182 (0.00764)				0.04846 (0.01979)	0.03620
	C	0.03518 (0.00749)	0.03264 (0.01025)			0.00059 (0.00380)				0.03391 (0.00582)	0.01710
	P	0.05390 (0.01083)	0.05349 (0.01174)			0.00351 (0.00537)				0.05369 (0.00784)	0.06136
		Sp	Cal	Por	Sp-Cal	Cal-Por	Sp-Por	Sp-Cal-Por		V_w	N_{st}
2	C	0.03518 (0.00749)	0.03264 (0.01025)	0.06946 (0.01535)	0.00059 (0.00380)	0.00432 (0.01103)	0.00209 (0.00965)	0.00233 (0.00722)		0.04576 (0.01187)	0.04845
		Ja	Por			Ja-Por				V_W	N_{st}
3	C2	0.08262 (0.00652)	0.07798 (0.01236)			0.00544 (0.00798)				0.0803 (0.00608)	0.06344

[a] Genomic regions A, F, C (nt 424 to 647 of the coat protein gene), P (Table 1 and Fig. 1), and C2 (first 462 nt of coat protein gene) were analyzed. The first group shows diversities of regions A, F, C, and P using the sequences determined by us here. The second includes comparisons only of region C, using our sequences and additional nucleotide sequences obtained from GenBank (see Materials and Methods). The last shows nucleotide diversities of region C2 estimated from sequences obtained from GenBank. The CTV isolates analyzed were from Spain (Sp), California (Cal), Portugal (Por), and Japan (Ja). V_b, average nucleotide diversity between geographic populations (24); V_w, average nucleotide diversity within geographic populations (24); $N_{st} = V_b/(V_w + V_b)$, proportion of the diversity between geographic populations with respect to total nucleotide diversity (24). Nucleotide diversity is the average number of nucleotide substitutions per site in each pair of sequence variants. Standard errors are indicated in parentheses.

isolates can be considered part of the same population. We wanted to know if this was also true for other CTV geographic populations. Unfortunately, not many CTV nucleotide sequences from the same geographic region are available. When genomic region C of six Portuguese isolates (GenBank accession nos. AF184113-AF84118) was included in the analysis, we found that the Portuguese CTV isolates were part of the same genetic population as the Spanish and Californian CTV isolates (Table 4). Analysis of the first 462 nt of the coat protein gene (C2) of 6 Portuguese and 13 Japanese CTV isolates (17) also showed higher diversity within a geographic population than between the two geographic populations (Table 4). Low genetic variation among geographically distant isolates has also been observed for *Curcubit yellow stunting disorder virus* and *Beet pseudo-yellows virus* in the genus *Crinivirus*, the other genus of the family *Closteroviridae* (41, 43), and for *Tobacco mild green mosaic virus* and *Pepper mild mottle virus* in the genus *Tobamovirus* (11, 37).

The genetic structure observed here suggests migration of CTV isolates among geographically isolated CTV populations. It is well known that an intense traffic of CTV-infected propagative citrus material has occurred between distant regions in the world (38). Within each geographic region, local dispersal is then effected by aphids.

REFERENCES

1. **Albiach-Martí, M. R., M. Mawassi, S. Gowda, T. Satyanarayana, M. E. Hilf, S. Shanker, E. C. Almira, M. C. Vives, C. López, J. Guerri, R. Flores, P. Moreno, S. M. Garnsey, and W. O. Dawson.** 2000. Sequences of citrus tristeza virus separated in time and space are essentially identical. J. Virol. **74:**6856–6865.

2. **Ayllón, M. A., C. López, J. Navas-Castillo, M. Mawassi, W. O. Dawson, J. Guerri, R. Flores, and P. Moreno.** 1999a. New defective RNAs from citrus tristeza virus: evidence for a replicase-driven template switching mechanism in their generation. J. Gen. Virol. **80:**817–821.

3. **Ayllon, M. A., L. Rubio, A. Moya, J. Guerri, and P. Moreno.** 1999b. The haplotype distribution of two genes of citrus tristeza virus is altered after host change or aphid transmission. Virology **255:**32–39.

4. **Ballester-Olmos, J. F., J. A. Pina, E. A. Carbonell, P. Moreno, A. Hermoso de Mendoza, M. Cambra, and L. Navarro.** 1988. Biological diversity of citrus tristeza virus (CTV) isolates in Spain. Plant Pathol. **42:**219–229.

5. **Bousalem, M., E. J. P. Dizery, and D. Fargette.** 2000. High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of yam mosaic virus: a contribution to understanding potyvirus evolution. J. Gen. Virol. **81:**243–255.

6. **Chao, L.** 1990. Fitness of RNA viruses decreased by Muller's ratchet. Nature **348:**454–455.

7. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.

8. **Febres, V. J., L. Ashoulin, M. Mawasi, A. Frank, M. Bar-Joseph, K. L. Manjunath, R. F. Lee, and C. L. Niblett.** 1996. The p27 protein is present at one end of citrus tristeza particles. Phytopathology **86:**1331–1335.

9. **Felsenstein, J.** 1989. PHYLIP—Phylogenetic Inference Package (version 3.2). Cladistics **5:**164–166.

10. **Fraile, A., J. L. Alonso-Prados, M. A. Aranda, J. J. Bernal, J. M. Malpica, and F. Garcia-Arenal.** 1997. Reports about recombination in virus natural populations. J. Virol. **71:**934–940.

11. **Fraile, A., J. M. Malpica, M. A. Aranda, E. Rodríguez-Cerezo, and F. García-Arenal.** 1996. Genetic diversity in tobacco mild green mosaic tobamovirus infecting the wild plant *Nicotiana glauca*. Virology **223:**148–155.

12. **Garcia-Arenal, F., A. Fraile, and J. M. Malpica.** 1999. Genetic variability and evolution, p. 143–159. *In* C. L. Mandahar (ed.), Molecular biology of plant viruses. Kluwer Academic Publishers, Boston, Mass.

13. **Gowda, S., T. Satyanarayana, C. L. Davis, J. Navas-Castillo, M. R. Albiach-Martí, M. Mawassi, N. Valkov, M. Bar-Joseph, P. Moreno, and W. O. Dawson.** 2000. The p20 gene product of *Citrus tristeza closterovirus* accumulates in the amorphous inclusion bodies. Virology **274:**246–254.

14. **Guerri, J., P. Moreno, N. Muñoz, and E. Martinez.** 1991. Variability among Spanish citrus tristeza virus isolates revealed by double-stranded RNA analysis. Plant Pathol. **40:**38–44.

15. **Hilf, M. E., A. V. Karasev, H. R. Pappu, D. J. Gumpf, C. L. Niblett, and S. M. Garnsey.** 1995. Characterization of citrus tristeza virus subgenomic RNAs in infected tissue. Virology **208:**576–582.

16. **Holland, J. J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol.** 1982. Rapid evolution of RNA genomes. Science **215:**1577–1582.

17. **Kano, T., T. Hiyama, T. Natsuaki, N. Imanishi, S. Okuda, and H. Ieki.** 1998. Comparative sequence analysis of biologically distinct isolates of citrus tristeza virus in Japan. Ann. Phytopathol. Soc. Jpn. **64:**270–275.

18. **Karasev, A. V., V. P. Boyko, S. Gowda, O. V. Nikolaeva, M. E. Hilf, E. V. Koonin, C. L. Nibblet, K. Cline, D. J. Gumpf, R. F. Lee, S. M. Garnsey, and W. O. Dawson.** 1995. Complete sequence of the citrus tristeza virus RNA genome. Virology **208:**511–520.

19. **Kong, P., L. Rubio, M. Polek, and B. W. Falk.** 2000. Population structure and genetic diversity of California citrus tristeza virus (CTV) field isolates. Virus Genes **21:**139–145.

20. **Krawczak, M., J. Reiss, J. Schmidtke, and U. Rösler.** 1989. Polymerase chain reaction: replication errors and reliability of gene diagnosis. Nucleic Acids Res. **17:**2197–2201.

21. **Li, W.-H.** 1993. Unbiased estimation of the rates of synonymous and non-synonymous substitutions. J. Mol. Evol. **36:**96–99.

22. **López, C., M. A. Ayllón, J. Navas, J. Guerri, P. Moreno, and R. Flores.** 1998. Molecular variability of the 5′ and 3′ terminal regions of the citrus tristeza virus RNA. Phytopathology **88:**685–691.

23. **López, C., J. Navas-Castillo, S. Gowda, P. Moreno, and R. Flores.** 2000. The 23-kDa protein coded by the 3′-terminal gene of citrus tristeza virus is an RNA-binding protein. Virology **269:**462–470.

24. **Lynch, M., and T. J. Crease.** 1990. The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol. **7:**377–394.

25. **Mawassi, M., R. Gafny, and M. Bar-Joseph.** 1993. Nucleotide sequence of the coat protein gene of citrus tristeza virus: comparison of biologically diverse isolates collected in Israel. Virus Genes **7:**265–275.

26. **Mawassi, M., A. V. Karasev, E. Mietkiewska, R. Gafny, R. F. Lee, W. O. Dawson, and M. Bar-Joseph.** 1995. Defective RNA molecules associated with citrus tristeza virus. Virology **208:**383–387.

27. **Mawassi, M., E. Mietkiewska, R. Gofman, G. Yang, and M. Bar-Joseph.** 1996. Unusual sequence relationship between two isolates of citrus tristeza virus. J. Gen. Virol. **77:**2359–2364.

28. **Mawassi, M., E. Mietkiewska, M. E. Hilf, L. Ashoulin, A. V. Karasev, R. Gafny, R. F. Lee, S. M. Garnsey, W. O. Dawson, and M. Bar-Joseph.** 1995. Multiple species of defective RNAs in plants infected with citrus tristeza virus. Virology **214:**264–268.

29. **Moreno, P., and J. Guerri.** 1997. Variability of citrus tristeza closterovirus (CTV): methods to differentiate isolates, p. 97–107. *In* P. Monette (ed.), Filamentous viruses of woody plants. Research Signpost, Trivandrum, India.

30. **Moreno, P., J. Guerri, and N. Muñoz.** 1990. Identification of Spanish strains of citrus tristeza virus (CTV) by analysis of double-stranded RNAs (dsRNA). Phytopathology **80:**477–482.

31. **Moya, A., and F. García-Arenal.** 1995. Populations genetics of viruses: an introduction, p. 213–223. *In* A. Gibbs, C. H. Calisher, and F. García-Arenal (ed.), Molecular basis of virus evolution. Cambridge University Press, Cambridge, England.

32. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York, N.Y.

33. **Pamilo, P., and N. O. Bianchi.** 1993. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol. Biol. Evol. **10:**271–281.

34. **Pappu, H. R., A. V. Karasev, E. J. Anderson, S. S. Pappu, M. E. Hilf, V. J. Febres, R. M. G. Eckloff, M. McCaffery, V. Boyko, S. Gowda, V. V. Dolja, E. V. Koonin, D. J. Gumpf, K. C. Cline, S. M. Garnsey, W. O. Dawson, R. F. Lee, and C. L. Niblett.** 1994. Nucleotide sequence and organization of eight open reading frames of the citrus tristeza closterovirus genome. Virology **199:**35–46.

35. **Pappu, H. R., S. Pappu, C. L. Niblett, R. F. Lee, and E. L. Civerolo.** 1993. Comparative sequence analysis of the coat protein of biologically distinct citrus tristeza closterovirus isolates. Virus Genes **7:**255–264.

36. **Pappu, S. S., V. J. Febres, H. R. Pappu, R. F. Lee, and C. L. Niblett.** 1997. Characterization of the 3′ proximal gene of the citrus tristeza closterovirus genome. Virus Res. **47:**51–57.

37. **Rodríguez-Cerezo, E., A. Moya, and F. García-Arenal.** 1989. Variability and evolution of the plant RNA virus pepper mild mottle virus. J. Virol. **63:**2198–2203.

38. **Roistacher, R. N., and P. Moreno.** 1991. The worldwide threat from destructive isolates of citrus tristeza virus—a review, p. 7–19. *In* R. H. Brlansky, R. F. Lee, and L. W. Timmer (ed.), Proceedings of the 11th Conference of the International Organization of Citrus Virologists. Department of Plant Pathology, University of California, Riverside, Calif.

39. **Roossinck, M. J.** 1997. Mechanisms of plant virus evolution. Annu. Rev. Phytopathol. **35:**1953–1965.

40. **Rubio, L., M. A. Ayllón, J. Guerri, H. Pappu, C. Niblett, and P. Moreno.** 1996. Differentiation of citrus tristeza closterovirus (CTV) isolates by single-strand conformation polymorphism analysis of the coat protein gene. Ann. Appl. Biol. **129:**479–489.

41. **Rubio, L., J. Soong, J. Kao, and B. W. Falk.** 1999. Geographic and molecular variation of isolates of three whitefly-borne closteroviruses: lettuce infectious yellows virus, cucurbit yellow stunting disorder virus and beet pseudo-yellows virus. Phytopathology **89:**707–711.

42. **Rubio, L., J. Guerri, and P. Moreno.** 2000. Characterization of *Citrus tristeza virus* isolates by single-strand conformation polymorphism analysis of DNA complementary to their RNA population, p. 12–17. *In* R. H. Yokomi, J. V. da Graca, and R. F. Lee (ed.), Proceedings of the 14th Conference of the International Organization of Citrus Virologists. Department of Plant Pathology, University of California, Riverside, Calif.

43. **Rubio, L., Y. Abou-Jawdah, H. Lin, and B. W. Falk.** 2001. Geographically distant isolates of the *Crinivirus cucurbit yellow stunting disorder virus* (CYSDV) show very low genetic diversity in the coat protein gene. J. Gen. Virol. **82:**929–933.

44. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

45. **Sawyer, S.** 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6:**526–538.

46. **Simon, A. E., and J. J. Bujarski.** 1994. RNA-RNA recombination and evolution in virus-infected plants. Annu. Rev. Phytopathol. **32:**337–362.

47. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.

48. **Vives, M. C., L. Rubio, C. Lopez, J. Navas-Castillo, M. R. Albiach-Martí, W. O. Dawson, J. Guerri, R. Flores, and P. Moreno.** 1999. The complete genome sequence of the major component of a mild citrus tristeza isolate. J. Gen. Virol. **80:**811–816.

49. **Weiller, G. F.** 1998. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. Mol. Biol. Evol. **15:**326–335.

50. **Yang, G., M. Mawassi, R. Gofman, R. Gafny, and M. Bar-Joseph.** 1997. Involvement of a subgenomic mRNA in the generation of a variable population of defective citrus tristeza virus molecules. J. Virol. **71:**9800–9802.

51. **Yang, Z.-N., D. H. Mathews, J. A. Dodds, and T. E. Mirkov.** 1999. Molecular characterization of an isolate of citrus tristeza virus that causes severe symptoms in sweet orange. Virus Genes **19:**131–142.