Contribution of Mutation and RNA Recombination to the Evolution of a Plant Pathogenic RNA

Miguel A. Aranda,[1] Aurora Fraile,[1] Joaquín Dopazo,[2] José M. Malpica,[3] Fernando García-Arenal[1]

[1]Departamento de Biotecnología, Escuela Técnica Superior de Ingenieros Agrónomos, 28040 Madrid, Spain
[2]Unidad de Bioinformática, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain
[3]Area de Biología Molecular y Virología Vegetal, Centro de Investigación y Tecnología-Instituto Nacional de Investigacines Agrarias, 28040, Madrid, Spain

Abstract. The nucleotide sequence of 17 variants of the satellite RNA of cucumber mosaic virus (CMV-satRNA) isolated from field-infected tomato plants in the springs of 1989, 1990, and 1991 was determined. The sequence of each of the 17 satRNAs was unique and was between 334 and 340 nucleotides in length; 57 positions were polymorphic. There was much genetic divergence, ranging from 0.006 to 0.141 nucleotide substitutions per site for pairwise comparisons, and averaging 0.074 for any pair. When the polymorphic positions were analyzed relative to a secondary structure model proposed for CMV-satRNAs, it was found that there were significantly different numbers of changes in base-paired and non–base-paired positions, and that mutations that did not disrupt base pairing were preferred at the putatively paired sites. This supports the concept that the need to maintain a functional structure may limit genetic divergence of CMV-satRNA. Phylogenetic analyses showed that the 17 CMV-satRNA variants clustered into two subgroups, I and II, and evolutionary lines proceeding by the sequential accumulation of mutations were apparent. Three satRNA variants were outliers for these two phylogenetic groups. They were shown to be recombinants of subgroup I and II satRNAs by calculating phylogenies for different molecular regions and by using Sawyer's test for gene conversion. At least two recombination events were required to produce these three recombinant satRNAs. Thus, recombinants were found to be frequent (~17%) in natural populations of CMV-satRNA, and recombination may make an important contribution to the generation of new variants. To our knowledge this is the first report of data allowing the frequency of recombinant isolates in natural populations of an RNA replicon to be estimated.

Key words: Nucleotide sequence — Cucumber mosaic virus — Plant pathogenic RNA

Introduction

High rates of spontaneous mutation have often been reported for RNA genomes. These result from frequent errors in RNA synthesis, and the accumulation of these mutations is commonly held to be the main mechanism in the evolution of RNA viruses (for recent reviews cf. Domingo 1989; Domingo and Holland 1988; Drake 1993). However, major changes in viral genotpyes may also involve genetic exchange through the reassortment of genomic segments, as exemplified by influenza virus (Simon and Bujarski 1994; Webster et al. 1992). For RNA viruses with nonsegmented genomes, genetic exchange is only possible through RNA recombination. Evidence for RNA recombination in several groups of animal and plant RNA viruses has accumulated during the last decade (Lai 1992, 1995; Robinson et al. 1987;

Simon and Bujarski 1994). Most data come from experimental systems designed to favor selection of recombinants. In natural populations, though, recombination has been observed rarely, and the frequency of recombinants has not been estimated.

In this paper we analyze the contribution of mutation accumulation and RNA recombination to the evolution of natural populations of the satellite RNA (satRNA) of cucumber mosaic virus (CMV). CMV is an isometric plant virus with a tripartite, single-stranded RNA genome of messenger polarity. Some CMV isolates also contain a small (312–390 nucleotides long), linear, non-coding, satRNA that depends on CMV as a helper virus for its replication, dispersion within the infected plant, encapsidation, and transmission. Much information has accumulated on the molecular biology of CMV and its satRNA (cf. Collmer and Howell 1992; Palukaitis et al. 1992; Roossinck et al. 1992). Much less has been reported on the genetic structure and evolution of CMV-satRNA populations under natural, field conditions. The recent occurrence in Spain of an epidemic of tomato necrosis caused by CMV + satRNA (Jordá et al. 1992) gave us the opportunity to address these issues. Natural satRNA populations, representing different yearly episodes of this epidemic, were shown to be very variable. The amount of genetic variation was maintained through time, and satRNA haplotypes persisted between epidemic episodes. The genetic differences observed could be explained by a sequential accumulation of mutations that reached an apparent threshold to genetic divergence (see Aranda et al. 1993). Genetic divergence in CMV-satRNA has been shown to be checked by selective constraints related to the maintenance of a molecular structure (Fraile and García-Arenal 1991). To get a deeper insight into the mechanisms sustaining the observed genetic differences, the nucleotide sequences of a set of field isolates of CMV-satRNA were determined and analyzed. The data presented here show that, in addition to point mutation, RNA recombination may be important in shaping the genetic structure of CMV-satRNA populations.

Materials and Methods

CMV-satRNA Isolates. Seventeen CMV-satRNA isolates were randomly sampled out of 62 field isolates collected during the springs of 1989, 1990, and 1991 and described in Aranda et al. (1993). The isolates (from where these 17 were sampled) each represented an electrophoretic variant in semidenaturing polyacrylamide electrophoresis (as in Aranda et al. 1993; García-Luque et al. 1984) present in virion RNA directly purified from field-infected tomato plants. CMV-satRNA variants were named N/i.j, meaning electrophoretic variant j from field sample i obtained in year N (Aranda et al. 1993).

Nucleotide Sequence Determination and Analyses. Full-length cDNA for these CMV-satRNAs was obtained by reverse transcription with AMV reverse transcriptase (Boehringer Manheim) and PCR am-

plified with Taq DNA polymerase (Promega) (Saiki et al. 1988) using the primers 5'GGAATTCCCGGGGTCCTG 3', with restriction sites for endonucleases *Eco*RI and *Sma*I, and having eight bases (underlined) complementary to the 3' end of all CMV-satRNA reported to date, and 5'GGAATTCTAATACGACTCACTATAVGTTTTGTTTG 3', with an *Eco*RI site, a modified T7-RNA polymerase promoter, and 10 nucleotides (underlined) identical to those at the 5' end of all CMV-satRNA reported to date. cDNA was hydrolyzed using *Eco*RI and *Sma*I and cloned in the vector pBS+. The sequence of the cDNA inserts was determined by the dideoxynucleotide chain termination method (as in Sambrook et al. 1989). At least two independent clones were sequenced for each CMV-satRNA.

The obtained nucleotide sequences were aligned and values for nucleotide substitution per site were calculated by the Jukes and Cantor (1969) method. These values were used to infer phylogenetic relationships by the neighbor-joining method (Saitou and Nei 1987). Sequence data were also used to obtain phylogenies by the Wagner parsimony method (Kluge and Farris 1969). The mutation pattern was also analyzed as suggested by Li et al. (1984). The possibility of genetic exchange was analyzed using methods described by Dykhuizen and Green (1991) and by Sawyer (1989). Analyses were done by hand, or using the PHYLIP 3.5 package (J. Felsenstein, Seattle, WA). A computer program was developed to perform Sawyer's test; this program is available by anonymous FTP at the internet address FTP.cnb.uam.es in the directory software/molevol.

CMV-satRNA sequences are available under EMBL accession numbers Z75870–Z75886.

Results

The sequences of 17 CMV-satRNAs randomly sampled among those isolated during the epidemic episodes of 1989, 1990, and 1991 are shown in Fig. 1 aligned against that of satRNA 1989/12.1. The sequence of each of the 17 satRNAs was unique, but all of them belonged to the 330–340-nucleotide size class, ranging in length from 334 to 340 nucleotides.

Genetic divergences varied for pairwise comparisons from 0.006 (for pairs 1989/20.1–1991/2.2 and 1989/20.1–1989/16.1) to 0.141 (for pairs 1990/5.1–1990/15.1 and 1990/19.1–1990/15.1). The average value of nucleotide substitution per site between any two pairs (as in Nei 1987, p. 276) was 0.074. Of a total of 343 nucleotide positions in the alignment of Fig. 1, 57 (16.6%) were polymorphic. At 50 of these, only two bases, including point insertion and deletions (indels) as a possible base, were found. The bases at 30 of these dimorphic positions (shown in boldface in Fig. 1) did not occur at random, but defined two subgroups of sequences: subgroup I (satRNAs 1989/12.1, 1989/16.1, 1989/19.1, 1989/.20.1, 1989/20.2, 1990/15.1, 1990/20.1, 1991/2.2, and 1991/4.1) and subgroup II (satRNAs 1990/5.1, 1990/16.1, 1990/19.1, 1990/21.1, and 1991/3.1). SatRNAs 1989/3.4, 1989/24.1, and 1991/8.1 did not belong clearly to either of these two subgroups.

Genetic Divergence in CMV-satRNA by Sequence Mutation

Phylogenetic relationships among the 17 satRNAs in Fig. 1 were calculated by the Wagner parsimony and neigh-

```
                  10        20        30        40        50        60        70        80
1989/12.1  GUUUUGUUUGAUAGAGAAUUGCGUAGAGGGGUUAUAUCUACGUGAGGAUCUAUCACUCGGCGGUGUGGGUUACCUCCCUG
1989/19.1  ..........U.......................................G...........................
1989/16.1  ..........U.................................................................
1989/20.1  ..........U.................................................................
1991/2.2   ..........U.................................................................
1989/20.2  ..........U.................................................................
1991/4.1   ..........U.................................................................
1990/15.1  ..........U...........................................................A.....
1990/20.1  ..........U...........................................................A.....
1989/3.4   ..........G.......C.......G.......G...................................A.....
1990/19.1  ..........G.......C.......G.......G...................................A.....
1990/21.1  ..........G.......C.......G.......G...................................A.....
1990/16.1  ..........G.......C.......G.......G...................................A.....
1991/3.1   ..........G.......C.......G.......G...................................A.....
1990/5.1   ..........G.......C.......G.......G...................................A.....
1989/24.1  ..........U.................................................................
1991/8.1   ..........U.................................................................

                  90       100       110       120       130       140       150       160
1989/12.1  CUACGGCGGGUUGAGUUGACGCA-CCUCGGACUGGGGACCGCUGGCUUGCGAGCUAUGUCCGCUACUCUCAGCACUACGC
1989/19.1  ..............................................................................
1989/16.1  ..............................................................................
1989/20.1  ..............................................................................
1991/2.2   ..............................................................................
1989/20.2  .................G............................................................
1991/4.1   ..............................................................................
1990/15.1  ...A........G.-...............................................................
1990/20.1  ...A........G.-...U.UUC.......................................................
1989/3.4   ..............-...U.UUC.......................................................
1990/19.1  ............C.-...U.UUC.............................................U.....A.
1990/21.1  ...A........-...U.UUC...............................................U.....A.
1990/16.1  ...A........-...U.UUC...............................................U.....A.
1991/3.1   ...A........-...U.UUC...............................................U.....A.
1990/5.1   ...........-CC..U.UUC...............................................U.....A.
1989/24.1  ..............................................................................
1991/8.1   ..............................................................................

                 170       180       190       200       210       220       230       240
1989/12.1  ACUCAUUUGAGCCCCCGCUCAGUUUGCUAGCAAAACCCGGCACAUGGUUCGCCGUUACCAUGG-A-UU-UCGAAAGAAAC
1989/19.1  ..................................................-.U..........A.A...........
1989/16.1  ....................................................................A...........
1989/20.1  ..............................................................U...............
1991/2.2   ..........................................................U.....U...............
1989/20.2  ....................................................A..U.....U...............
1991/4.1   ..........................................................CAA.................
1990/15.1  ..........................................................U......CAAA.........
1990/20.1  ..........................................................U......CAAA.........
1989/3.4   ....................................................................A.A.........
1990/19.1  U..................A..G.....C..............G.....U.........
1990/21.1  U..................A..G.....C..............G...............
1990/16.1  U..................A..G.....C..............G...............
1991/3.1   U....................G.....C......U....G...............
1990/5.1   U..................A..G............U....G...............
1989/24.1  ..................................................................A...........
1991/8.1   .........................................................ACAC...............

                 250       260       270       280       290       300       310       320
1989/12.1  ACUCUGUUAGGUGGUAUCGUG-GAUGACGCACGCAGGGAGAGGCUAAAACCUAUAUGGUCAUGCUGAUCUCCGUGAAUGU
1989/19.1  .............................................................A...............
1989/16.1  ..............................A..........................A...............
1989/20.1  ..............................A..........................A...............
1991/2.2   .........................................................A...............
1989/20.2  ..............................A..........................A...............
1991/4.1   ..............................A..........................A...............
1990/15.1  .........................................................A...............
1990/20.1  ..............................A..........................A...............
1989/3.4   ..............................A.........................
1990/19.1  .......AU........-.A.UC.........GG.UA.---U.CU..............
1990/21.1  .......CAU.......-.A.UC.........GG.U.---..CU..............
1990/16.1  .......A........-.A.UC.........GG.U.---U.CU.C.............
1991/3.1   ................-.A.UC...............A..GG.U.---U.CU..............
1990/5.1   ................-.A.UC...............A..GG.GA.---U.CU..............
1989/24.1  .......A........-.A.UC.........-G.U.---..CU..............
1991/8.1   .......A........-.A.UC.........GG.U.---U.CU..............

                 330       340
1989/12.1  CUACACAUUCCUCUACAGGACCC
1989/19.1  ..............A........
1989/16.1  ..............C........
1989/20.1  .......................
1991/2.2   .......................
1989/20.2  .......................
1991/4.1   .......................
1990/15.1  ............A.C........
1990/20.1  ............C........
1989/3.4   ............C........
1990/19.1  ...U-.....AU.........
1990/21.1  ...U-.....AU.........
1990/16.1  ...U.........
1991/3.1   ...U-.........
1990/5.1   ...U-.....AU.........
1989/24.1  ...U-.......C.........
1991/8.1   ...U-.......C.........
```

Fig. 1. Alignment of the nucleotide sequence of 17 field isolates of CMV-satRNA. Only differences respective to the sequence of satRNA 1989/12.1 are presented; (–) indicates a deletion. Dimorphic positions dividing satellite RNA into two subgroups (I and II) are in **bold type.** Positions that would be base-paired according to Gordon and Symons (1983) secondary structure model are *underlined.*

bor-joining methods. E-satRNA (Hidaka et al. 1988), which is clearly distinct from most 330–340-nucleotide-long CMV-satRNAs (Fraile and García-Arenal 1991), was used as an outgroup. Both methods yielded similar tree topologies (shown in Fig. 2 for the Wagner parsimony method): two main branches, corresponding to subgroups I and II as defined above on the basis of dimorphic sites, were found. CMV-satRNAs 1989/3.4, and 1989/24.1 and 1991/8.1, were outliers, respectively, of subgroups I and II (branches Ib and IIb in Fig. 2). When these three satRNAs were excluded from the analysis, the significance of the clustering in a bootstrap test increased considerably; subgroup I was found in 98% of the trees, vs 72%, and subgroup II was found in 100% of the trees, vs 84%. Phylogenetic relationships among the satRNAs did not correspond to the date of isolation. As shown by branch Ia in the tree of Fig. 2, evolutionary lines may be detected with intermediary types found at or near the nodes, indicating divergence from a parental satRNA by the sequential accumulation of point mutations (see also the polymorphic sites in the corresponding satRNAs in Fig. 1).

Analyzing the sequences of 23 CMV-satRNAs isolated at different parts of the world from different host plants showed that the need to maintain a structure may limit the genetic divergence of CMV-satRNAs (Fraile and García-Arenal 1991). To check whether structure-related constraints may be detected in a population of CMV-satRNA from a single host plant, and from a restricted area, a similar analysis was done. A putative ancestral sequence was derived by maximum parsimony (with E-satRNA as an outgroup). That sequence was folded into a putative secondary structure like that proposed by Gordon and Symons (1983) for Q-satRNA, with 50% of its bases paired (in 44 G:C, 24 A:U, and 16 G:U pairs). Then, mutations in each of the 17 satRNAs in Fig. 1 respective to this ancestral satRNA were analyzed separately for base-paired (underlined in Fig. 1) and non–base-paired positions (Gojobori et al. 1982). Based on a total of 359 differences for the 57 polymorphic positions it was found that the number of point mutations per site (single base substitutions + point indels) is significantly greater for nonpaired than for paired positions (0.0875 ± 0.0045 vs 0.0641 ± 0.0035, $P \leqslant 0.0001$, all comparisons by chi-square and Wilcoxon's nonparametric tests). Indels are also significantly ($P \leqslant 10^{-6}$) more frequent at unpaired (0.0247 ± 0.0022 per site) than at paired (0.0028 ± 0.0007) positions. The frequency of transitions calculated as in Li et al. (1984) was significantly ($P \leqslant 0.003$) greater for paired (58.19%) than for unpaired positions (41.92%). Also, replacement tendencies for G and C were lower in paired (9.97% for both G and C) than in unpaired (20.57% for G and 21.52% for C) positions. Thus, the pattern of mutation accumulation differed significantly according to positions, so muta-

tions that disrupt base pairing were less preferred at the sites likely to be parts of secondary structure elements.

Genetic Divergence by RNA Recombination

The bases at the dimorphic positions for the satRNAs that are outliers in subgroups I and II (Fig. 2) show that satRNA 1989/3.4 resembles subgroup II from its 5′ end to some point between nucleotides 105 and 152, and it has a sequence similar to subgroup I from this point to its 3′ end. On the other hand, the sequences of satRNAs 1989/24.1 and 1991/8.1 are similar to those of subgroup I from the 5′ and to some point between nucleotides 230 and 250, and from here to the 3′ end they are similar to subgroup II (see Fig. 1). This suggests that these three satRNAs might have originated by recombination between satRNAs in subgroups I and II. To test this hypothesis two different approaches were used. First, phylogenetic analyses of different regions of the satRNAs (Dykhuizen and Green 1991) were done, involving satRNAs from each of the subgroups Ia and IIa together with the putative recombinant(s). The significance of the tree topologies obtained by the Wagner parsimony (Fig. 3) and neighbor-joining methods was analyzed by bootstrap (Felsenstein 1985) with 1,000 replicates. Results were the same for both phylogenetic methods. SatRNA 1989/3.4 clustered with subgroup II when nucleotides 1–150 were considered and with subgroup I when nucleotides 150–3′ end were considered (Fig. 3B). Conversely, satRNAs 1989/24.1 and 1991/8.1 clustered with subgroup I when nucleotides 1–250 were considered and with subgroup II when nucleotides 250–3′ end were considered (Fig. 3A). For all analyses, and for both methods of phylogenetic inference, these clusterings were statistically significant.

Second, Sawyer's test for gene conversion (Sawyer 1989) was applied to the set of 17 satRNAs. This test is based on the search for stretches of sequences identical for different isolates that would be significantly longer than expected based on the random distribution of polymorphic sites. Table 1 shows that the value of both statistics related to the sum of the square length of fragments conserved between sequences departed very significantly from what would be expected by random, $P \leqslant 0.00001$ for both SSCF (see table for abbreviation key), when only the condensed 57 polymorphic positions among the 17 satRNAs were considered, and SSUF when all (uncondensed) positions were considered. It could be that these results of Sawyer's test were an artefact resulting from the sequence conservation of some regions of the satRNA caused by base pairing (see above). Therefore, Sawyer's test was repeated just for the putatively unpaired positions, and the results were the same as when all nucleotide positions were considered. On the other hand, both when all positions were consid-
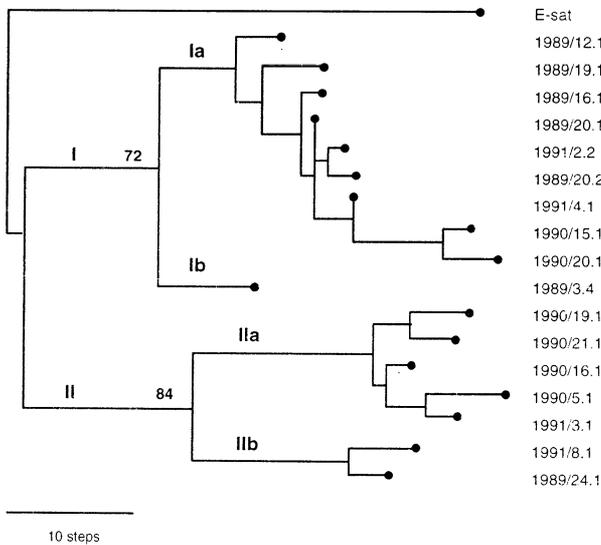
Fig. 2. Most parsimonious phylogenetic tree for 17 field isolates of CMV-satRNA. E-satRNA (Hidaka et al. 1988) was used as an outgroup. The percentage of trees in a bootstrap analysis, after 1,000 replicates, that support this topology is indicated.

ered and when only the putatively unpaired positions were considered, the value of statistics related to the maximum length of fragments shared between two sequences (MCF and MUF) did not depart from what was expected randomly.

The ability of Sawyer's test to locate the putative recombinant fragment(s) is very dependent on the set of sequences used. This is because the test compares the largest common fragment, condensed (MCF) or uncondensed (MUR), found in any pair of sequences with their distributions obtained from random permutations of the data. If very similar sequences are included in the comparison, the MCF (or MUF) values will be determined by those sequences: In the simulation, large fragments will be found, increasing the P value—that is, the number of simulation sets showing a MCF (or MUF) score higher than the fragment. To avoid this effect, for the location of putatively exchanged fragments, we performed Sawyer's test on the 136 possible pairs of sequences. With this modification the test becomes more powerful, and many putative recombinant regions were found using uncondensed sequences. The fragments with a length that deviated significantly from what was expected from random ($P \leq 0.00001$) were shared by satRNAs 1989/24.1 and 1991/8.1 and the satRNAs in subgroup Ia, involving sequences 5′ to position 225, or by satRNA 1989/3.4 and satRNAs in subgroup IIa, involving sequences 5′ to position 152. These significances remained and were statistically significant except for two out of 23 tests, when considered as part of a multiple test with discontinuous distribution of significances (Malpica and Briscoe 1982), and the combined probabilities (Sokal and Rohlf 1981, p. 780) of any of the three sets of tests was again highly

significant. When condensed sequences were used in addition to the region above, the test also detected regions shared between satRNA 1991/8.1 and satRNAs in subgroup IIa 3′ to position 250. With a lesser significance ($P \leq 0.02$) the test also detected small regions in which some sequences present an unusual accumulation of mutations (i.e., fragments shared between satRNAs in subgroups I and II between positions 223 and 250, and between 248 and 300 in Fig. 1) (data not shown).

Thus, all evidence taken together supports the conclusion that satRNAs 1989/3.4, 1989/24.1, and 1991/8.1 originated from a recombination between satRNAs in subgroup I and in subgroup II. The frequency of recombinant satRNAs in the analyzed population is high (17.6%). These recombinant satRNAs have been generated by at least two different recombination events: One would have generated satRNA 1989/3.4 and a second one satRNAs 1989/24.1 and 1991/8.1. The sequence data are compatible with the possibility that satRNAs 1989/24.1 and 1991/8.1 evolved by mutation accumulation from the same parental recombinant sequence.

Discussion

We have previously reported, from the analysis of 62 isolates of CMV-satRNA by the ribonuclease protection assay method (RPA, Palukaitis et al. 1994), that populations of CMV-satRNA representing three episodes of a field epidemic on tomato are very heterogeneous. The diversity of the population was maintained through time, with little replacement of haplotypes between epidemic episodes, and genetic diversity seemed to be near an upper permitted threshold (Aranda et al. 1993). The data presented here, derived from the analysis of 17 CMV-satRNAs randomly sampled from those 62, confirm our previous conclusions: All the sequenced satRNA isolates belong to the 330–340-nucleotide size class, each determined sequence is unique, phylogenetic analyses show that relationships among satRNA isolates do not correlate with the year in which they were isolated, and nucleotide substitution values per site in the population are very large even as compared to those reported for other plant or animal RNA viruses (Moya and García-Arenal 1995; Rodríguez-Cerezo et al. 1989, 1991). The average and extreme values of diversity, estimated as number of nucleotide substitutions per site, also agree with those estimated from RPA data (Aranda et al. 1993) and are similar to those reported for a set of 23 CMV-satRNAs isolated from different parts of the world and from different host plants (Fraile and García-Arenal 1991). Point mutation appears to be the main cause for the observed genetic diversity, as has frequently been reported for RNA genomes (Domingo 1989; Drake 1993). Our data support the notion that divergence may
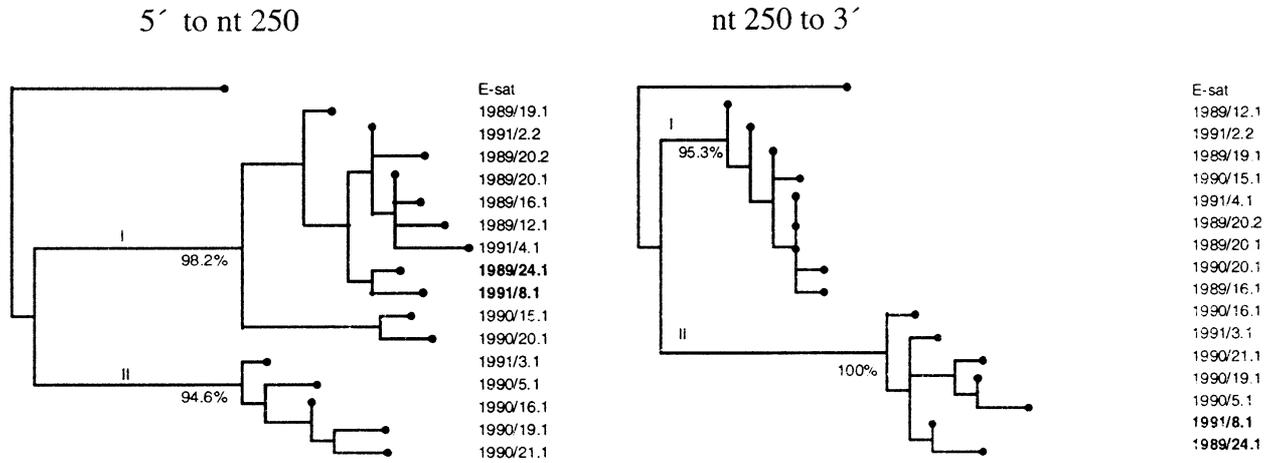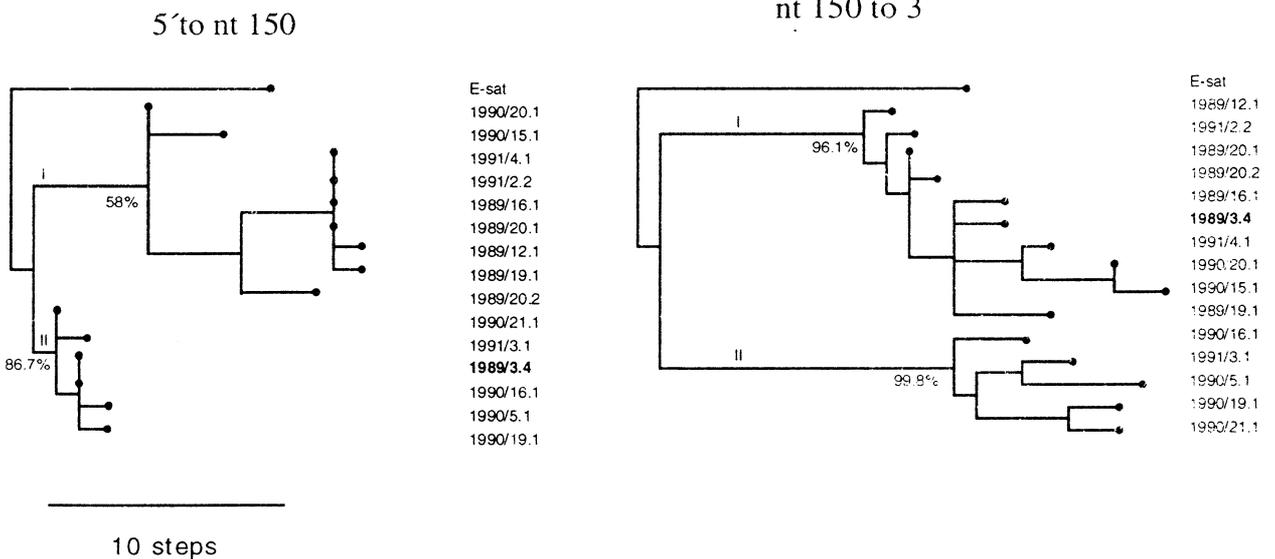
Fig. 3. Most parsimonious phylogenetic trees for different molecular regions of CMV-satRNA. CMV-satRNAs in subgroups Ia and IIa (as in Fig. 2) plus satRNAs 1989/24.1 and 1991/8.1 (**A**) or 1989/3.4 (**B**) were considered. Molecular regions were from 5′ end to nucleotide 250 and from nucleotide 250 to 3′ end (in **A**) or 5′ end to nucleotide 150 and from nucleotide 150 to 3′ end (in **B**). The percentage of trees in a bootstrap analysis with 1,000 replicates that support the shown topologies is indicated.

occur along different evolutionary liens by the sequential accumulation of mutations to reach what may be an upper threshold: It has been shown for CMV-satRNA (Fraile and García-Arenal 1991) that negative selection to maintain a functional secondary structure may be important in limiting genetic divergence. This is also apparent from the sequences of 17 field satRNAs: The analysis of the observed mutation pattern shows it to be different for nucleotide positions that might be paired, or unpaired, as predicted from a secondary structure of the satRNA molecule proposed initially for Q-satRNA (Gordon and Symons 1983) and extended later to many other satRNAs (García-Arenal et al. 1987; Hidaka et al. 1988; our unpublished results). Mutations that are less liable to disrupt base pairing are found at the putatively paired

positions. It could be that constraints required to maintain a molecular structure would be less restrictive than those related to the maintenance of an encoded protein sequence, and this could explain the high values of genetic divergence found for CMV-satRNA as compared to RNA viruses, including its helper virus, CMV (our unpublished results). It could also be that its dependence from functions encoded by a virus (CMV) allow a wider range of divergence.

Interestingly, RNA recombination may play an important part in the evolution of CMV-satRNA: Three out of the 17 analyzed satRNA isolates probably arose through recombination between satRNAs in subgroups I and II, according to the significance of analyses based on the congruency of phylogenies for different molecular

Table 1. Values of the statistics of Sawyer's test for gene conversion[a]

Statistics[b]	Observed score	P value[c]	SD above mean[d]	SD of scores[e]
SSCF	46,499	0.0000	4.14	3,801.46
MCF	29	0.9412	−1.49	6.63
SSUF	2,925,600	0.0000	7.08	187,559.62
MUF	225	0.6170	−0.33	41.04

[a] Permuting 57 polymorphic positions in 17 CMV-satRNA sequence variants

[b] SSCF = sum of the squares of the condensed fragment lengths, MCF = maximum condensed fragment length, SSUF = sum of the squares of the uncondensed fragment lengths, MUF = maximum uncondensed fragment length

[c] Relative number of permuted data sets with scores greater than or equal to the observed score

[d] For 10,000 random permutations of the data set

[e] (observed score − mean)/SD

regions (Dykhuizen and Green 1991), and on Sawyer's test for gene conversion (Sawyer 1989). CMV-satRNA has a highly structured molecule, and this might favor the stop and template-switch of the polymerase complex, resulting in recombination (Carpenter and Simon 1994; Nagy and Bujarski 1993; Simon and Bujarski 1994). By Sawyer's test we have been able to identify putative recombinant regions in the satRNAs, but the exact points of crossing-over cannot be identified by the analysis reported here. Thus, no specific structural or sequence element can be linked with the generation of satRNAs 1989/3.4 and 1989/24.1, or 1991/8.1. Nevertheless, it is likely that there are long stretches of paired positions (i.e., nt ~ 85–120, nt ~ 240–260) near the regions where crossing-over must have occurred.

RNA recombination has frequently been shown to occur for both animal and plant RNA viruses under experimental conditions (Lai 1992, 1995; Simon and Bujarski 1994), and it may be particularly frequent for non-coding RNA replicons such as DI RNAs or satRNAs (Carpenter and Simon 1994; Cascone et al. 1990). It has been argued that RNA recombination may gain selective advantages by introducing major changes to the genetic and biological properties of the virus and, thus, increasing its genetic plasticity. Also, RNA recombination could diminish the negative effects of high mutation rates due to the error frequencies of RNA polymerases (Chao 1990; Lai 1992, 1995). Nevertheless, recombinants are rare, or unreported, in natural populations, even for virus groups like picornaviruses or coronaviruses, which recombine easily under experimental conditions (Lai 1995). It is thought that most viral recombinants may have selective disadvantages in nature due to requirements for perfect sequence and structural complementarity of the encoded proteins and/or RNAs for their optimal functionality (Lai 1995). The dependence of CMV-satRNA upon its helper virus CMV would diminish these constraints, what may be related to the high

frequency of recombinants in natural populations. The maintenance of a secondary structure is, so far, the only identified constraint for the genetic divergence of CMV-satRNA. SatRNAs 1989/3.4, 1989/24.1, and 1991/8.1 can be folded into structures that have a stability closely similar to that of the isolates in subgroups I and II.

Although natural isolates of several RNA viruses have been reported for which a recombinant origin has been proposed (cf. Simon and Bujarski 1994 for plant viruses; or Robertson et al. 1995 for AIDS viruses), this is the first report, to our knowledge, of recombinant isolates being identified in the same natural population with isolates that could be parental to them. Also, no data have been reported for the frequency of recombinant isolates in natural populations. Our data show that RNA recombinants may be present at high frequencies in natural populations of RNA replicons, and that recombination may make an important contribution to their evolution. For CMV-satRNA, RNA recombination may bridge the constraints that appear to limit mutation accumulation into two main genetic types (subgroups I and II), thus increasing the genetic plasticity of this highly variable RNA replicon.

References

Aranda MA, Fraile A, García-Arenal F (1993) Genetic variability and evolution of the satellite RNA of cucumber mosaic virus during natural epidemics. J Virol 67:5896–5901

Carpenter CD, Simon AE (1994) Recombination between plus and minus strands of turnip crinkle virus. Virology 201:419–423

Cascone PJ, Carpenter CD, Li XH, Simon AE (1990) Recombination between satellite RNAs of turnip crinkle virus. EMBO J 9:1709–1715

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

Collmer CW, Howell SH (1992) Role of satellite RNA in the expression of symptoms caused by plant viruses. Annu Rev Phytopathol 30:419–442

Domingo E (1989) RNA virus evolution and the control of viral disease. Prog Drug Res 33:93–133

Domingo E, Holland JJ (1988) High error rates, population equilibrium, and evolution of RNA replication systems. In: Domingo E, Holland JJ, Ahlquist P (eds) RNA Genetics, vol 3. CRC Press, Boca Raton, FL

Drake JW (1993) Role of spontaneous mutation among RNA viruses. Proc Natl Acad Sci USA 90:4171–4175

Dykhuizen DE, Green L (1991) Recombination in *Escherichia coli* and the definition of biological species. J Bacteriol 173:7257–7268

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fraile A, García-Arenal F (1991) Secondary structure as a constraint to the evolution of a plant viral satellite RNA. J Mol Biol 221:1065–1069

García-Arenal F, Zaitlin M, Palukaitis P (1987) Nucleotide sequence analysis of six satellite RNAs of cucumber mosaic virus: primary sequence and secondary structure alterations do not correlate with differences in pathogenicity. Virology 158:339–347

García-Luque I, Kaper JM, Díaz-Ruiz JR, Rubio-Huertos M (1984) Emergence and characterization of satellite RNAs associated with Spanish cucumber mosaic virus isolates. J Gen Virol 65:539–547

Gojobori T, Li WH, Graur D (1982) Patterns of nucleotide substitutions in pseudogenes and functional genes. J Mol Evol 18:360–369

Gordon KHJ, Symons RH (1983) Satellite RNA of cucumber mosaic virus forms a secondary structure with partial 3′-terminal homology to genomal RNAs. Nucleic Acids Res 11:947–960

Hidaka S, Hanada K, Ishikawa K, Miura K (1988) Complete nucleotide sequence of two new satellite RNA associated with cucumber mosaic virus. Virology 164:326–333

Jordá C, Alfaro A, Aranda MA, Moriones E, García-Arenal F (1992) An epidemic of cucumber mosaic virus plus satellite RNA in tomatoes in Eastern Spain. Plant Dis 76:363–366

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York

Kluge AG, Farris JS (1969) Quantitative phyletics and the evolution of anurans. Syst Zool 18:1–32

Lai MM (1992) RNA recombination in animal and plant viruses. Microbiol Rev 56:61–79

Lai MM (1995) Recombination and its evolutionary effect on viruses with RNA genomes. In: Gibbs A, Calisher CH, García-Arenal F (eds) Molecular basis of virus evolution. Cambridge University Press, Oxford, Great Britain

Li W-H, Wu C-I, Luo C-C (1984) Nonrandomness of point mutation as reflected in nucleotide substitution in pseudogenes and its evolutionary implications. J Mol Evol 21:58–71

Malpica JM, Briscoe DA (1982) Multilocus nonrandom associations in *Drosophila melanogaster.* Genetical Res 39:41–61

Moya A, García-Arenal F (1995) Population genetics of viruses: an introduction. In: Gibbs A, Calisher CH, García Arenal F (eds) Molecular basis of virus evolution. Cambridge University Press, Great Britain

Nagy PD, Bujarski JJ (1993) Targeting the site RNA-RNA recombi-nation in brome mosaic virus with antisense sequences. Proc Natl Acad Sci USA 90:6390–6394

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Palukaitis P, Roossinck MJ, Dietzgen RG, Francki RIB (1992) Cucumber mosaic virus. Adv Virus Res 41:281–348

Palukaitis P, Roossinck MJ, García-Arenal F (1994) Applications of ribonuclease protection assay in plant virology. Methods Mol Genet 4:237–250

Robertson DL, Hahn BH, Sharp PM (1985) Recombination in AIDS viruses. J Mol Evol 40:249–259

Robinson DJ, Hamilton WDO, Harrison BD, Baulcombe DC (1987) The anomalous tobravirus isolates: evidence for RNA recombination in nature. J Gen Virol 68:2551–2561

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

Rodríguez-Cerezo E, Elena SF, Moya A, García-Arenal F (1991) High genetic stability in natural populations of the plant RNA virus tobacco mild green mosaic virus. J Mol Evol 32:328–332

Roossinck MJ, Sleat D, Palukaitis P (1992) Satellite RNAs of plant viruses: structure and biological effects. Microbiol Rev 56:265–279

Saiki RH, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487–491

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 3:57–74

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Sawyer S (1989) Statistical test for detecting gene conversion. Mol Biol Evol 6:526–538

Simon AE, Bujarski JJ (1994) RNA-RNA recombination and evolution in virus-infected plants. Annu Rev Phytopathol 32:337–362

Sokal RR, Rohlf FJ (1981) Biometry. 2nd ed. WH Freeman, New York

Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y (1992) Evolution and ecology of influenza A viruses. Microbiol Rev 56;152–179