Natural recombination between *Tomato yellow leaf curl virus*-Is and *Tomato leaf curl virus*


1 Estación Experimental ‘La Mayora’, Consejo Superior de Investigaciones Científicas, 29750 Algarrobo-Costa, Málaga, Spain
2 Istituto di Fitovirologia Applicata, CNR, Strada delle Cacce 73, 10135 Torino, Italy
3 Direción Geral de Protecção das Culturas, Quinta do Marquês, 2780 Oeiras, Portugal

The complete genome sequences (2791 and 2793 nt) of isolates of *Tomato yellow leaf curl virus*-Is (TYLCV-Is) from Spain (SP72/97) and Portugal (Port2/95) were determined. These isolates are closely related to TYLCV-Is isolates reported in Japan (Japan-A and Japan-S) and Israel (Israel/Mild). Comparison of all sequenced isolates of TYLCV-Is showed that part of the genome comprising the intergenic region and the 5' end of the *rep* gene of the Iran and Israel isolates was not closely related to that of other isolates. Phylogenetic analyses suggest that the Israel and Iran isolates may have chimeric genomes that have arisen by recombination between TYLCV-Is-like and tomato leaf curl virus (ToLCV)-like ancestors. The TYLCV-Is donors of the Iran and the Israel genomes were closely related to each other and to other known TYLCV-Is isolates. However, the ToLCV donors differed from each other, although both were related to ToLCV isolates from India (Bangalore-2 and Bangalore-4).

Tomato yellow leaf curl is one of the most devastating virus diseases of tomato crops in tropical and subtropical regions (Czosnek & Laterrot, 1997). The causal agents are a group of virus species of the genus *Begomovirus*, family *Geminiviridae*, all of them named *Tomato yellow leaf curl virus* (TYLCV), that are transmitted by the whitefly *Bemisia tabaci* Gen. (Briddon & Markham, 1995). The TYLCV species have a circular, single-stranded DNA genome that, except in one case in which two components were isolated (Rochester et al., 1994), is composed of a single molecule. Monopartite TYLCVs have a genome (about 2.8 kb) that contains six partially overlapping open reading frames (ORFs), two (V1 and V2) on the virion-sense strand and four (C1 to C4) on the complementary-sense strand, separated by an intergenic region (IR) of about 300 nt. Biological, molecular, epidemiological and control aspects of the TYLCV complex have been reviewed in Picó et al. (1996) and, more recently, in Moriones & Navas-Castillo (2000).

TYLCV has been known in the Iberian Peninsula since 1992, when the species TYLCV-Sar was reported as the causal agent of yellow leaf curl epidemics in tomatoes in southern Spain (Noris et al., 1994). The TYLCV-Is species was also found associated with epidemics in tomato in 1995 in southern Portugal and 1997 in southern Spain, either alone or in mixed infections with TYLCV-Sar (Louro et al., 1996; Navas-Castillo et al., 1997). In this paper, we report the genetic characterization of infectious clones from TYLCV-Is isolates collected in Spain and Portugal and we describe evidence of natural recombination between TYLCV-Is and the begomovirus *Tomato leaf curl virus* (ToLCV).

Infectious clones that produced *Bemisia tabaci*-transmissible progeny were obtained as described in Navas-Castillo et al. (1999) from the TYLCV-Is-infected tomato samples SP72/97, collected in Spain in June 1997, and Port2/95, collected in Portugal in September 1995. The complete genome sequences of these clones were determined from both strands of the full-length inserts cloned in pSP72/97 and pPort2 [GenBank accession numbers AF071228 (2791 nt) and AF105975 (2793 nt)] by automated sequencing with pUC/M13 universal primers and specific primers based on partial sequences. The DNA sequences of SP72/97 and Port2/95 were compared with those of TYLCV-Sar isolate Sardinia (GenBank accession no. X61153) and TYLCV-Is isolates for which complete genome information was available: Japan Shizuoka (Japan-S), GenBank accession no. AB014346; Japan Aichi (Japan-A), AB014347; Israel, X15656; Israel/Mild, X76319; Cuba, AJ223505; Dominican Republic, AF024715; and Iran, AJ132711. Isolates SP72/97 and Port2/95 appeared to be closely related variants of TYLCV-Is and their genomes strongly resembled those of the Japan-A and Japan-S isolates.
(Kato et al., 1998) (Table 1). In addition, the isolates from the Iberian Peninsula and Japan were closely related to the Israel/Mild isolate and more distantly related to the Israel, Cuba, Dominican Republic and Iran isolates. Sequence identities among the IRs or complete genomes followed a similar pattern. Nucleotide sequence comparisons of the different ORFs provided further support for the observed relationships between the TYLCV-Is isolates; interestingly, ORFs C1 and C4 were less conserved than V1, V2, C2 and C3 (Table 1). The presence of such closely related variants in geographical regions so distantly separated can only be explained by movement of infected plant material or viruliferous B. tabaci individuals via plant trade or human movement. In fact, this was argued as the cause of the primary spread of Middle East Mild isolate via plant trade or human movement. In fact, this was argued as the cause of the primary spread of Middle East Mild isolate and more distantly related to the Israel, Cuba, Dominican Republic and Iran isolates. Sequence identities among the IRs or complete genomes followed a similar pattern. Nucleotide sequence comparisons of the different ORFs provided further support for the observed relationships between the TYLCV-Is isolates; interestingly, ORFs C1 and C4 were less conserved than V1, V2, C2 and C3 (Table 1). The presence of such closely related variants in geographical regions so distantly separated can only be explained by movement of infected plant material or viruliferous B. tabaci individuals via plant trade or human movement. In fact, this was argued as the cause of the primary spread of Middle East Mild isolate to the Caribbean and then to USA (Polston & Anderson, 1997; Polston et al., 1999).

The IRs of SP72/97 and Port2/95 show features characteristic of begomoviruses, including the stem–loop structure with the conserved nonanucleotide sequence 5’-TAATATT-AC-3’ in the loop, where the breaking and joining site (ori) for rolling-circle replication occurs (for a review see Gutiérrez, 1999). In addition, a perfect conservation of the stem–loop structure with high thermodynamic stability (ΔG = −16.1 kcal/mol or −3.8 kJ/mol) (analyses performed using the DNA mfold server: M. Zuker, Washington University, St Louis, MO, USA; http://www.ibc.wustl.edu/~zuker/) was observed in the IRs of SP72/97, Port2/95 and all TYLCV-Is isolates listed above, except for Israel/Mild. In the latter, a less stable stem–loop structure was predicted (ΔG = −6.8 kcal/mol or −1.6 kJ/mol), owing to incomplete base-pairing in the stem. As point mutations that destroy base-pairing in the stem have been shown to be deleterious for virus replication (Orozco & Hanley-Bowdoin, 1966), the mildness reported for the Israel/Mild isolate in certain tomato cultivars (Antignus & Cohen, 1994) might be due to an altered stem–loop structure.

Phylogenetic relationships of SP72/97 and Port2/95 to other TYLCV-Is isolates (listed above) were studied on the basis of the nucleotide sequences of the IRs and ORFs V1, V2, and C1 to C4. As changes in the topological position of certain isolates occurred depending on which part of the genome was compared (not shown), a more detailed comparison throughout the genome was done with the aid of plotsimilarity diagrams (Wisconsin GCG software package; Devereux et al., 1984). In these comparisons, the nucleotide sequences of the SP72/97, Port2/95, Japan-A and Japan-S isolates on the one hand and the Israel, Cuba and Dominican Republic isolates on the other proved to be almost identical throughout the genome (not shown). Therefore, in Fig. 1, the first group of isolates is represented by SP72/97 while the second is represented by Israel. Comparisons revealed that a large portion of the genome, hereafter referred to as region I (see below), was quite well conserved among all TYLCV-Is isolates (Fig. 1). The nucleotide sequence of the Israel isolate differed from that of SP72/97 in the 5′-proximal two-thirds of the C1 ORF and the 5′-half of the IR (regions II and III; Fig. 1A). The Israel and Iran sequences also differed in the same part of the genome (Fig. 1B). However, when the sequences of SP72/97 and Iran were compared, differences were only observed in region III (Fig. 1C). The Israel/Mild sequence was highly similar to that of SP72/97 except for the central part of the IR (region IV; Fig. 1D). Nucleotide sequences in regions II to IV of TYLCV-Is isolates were then compared with other available geminivirus sequences, the limits of each region established by visual inspection of sequence alignments (indicated in Fig. 1). The sequence in region III of the Iran isolate was more closely related (91% identity) to the sequence of the equivalent region of the Bangalore-4 isolate of ToLCV (GenBank accession no. AF165098) than to that of any other TYLCV-Is isolate (below 75% identity). Sequences in regions II and III of the Israel, Cuba and Dominican Republic isolates were 97–100% identical, their closest relationships (between 73 and 83% identity) being to sequences in the equivalent regions of the Bangalore-2

<table>
<thead>
<tr>
<th>Domain</th>
<th>Port2/95</th>
<th>Japan-A</th>
<th>Japan-S</th>
<th>Israel/Mild</th>
<th>Israel</th>
<th>Cuba</th>
<th>Dom. Rep.</th>
<th>Iran</th>
<th>TYLCV-Sar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total DNA</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>80</td>
</tr>
<tr>
<td>IR</td>
<td>98</td>
<td>93</td>
<td>97</td>
<td>92</td>
<td>80</td>
<td>81</td>
<td>82</td>
<td>75</td>
<td>57</td>
</tr>
<tr>
<td>V1 (CP)</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>96</td>
</tr>
<tr>
<td>V2</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
</tr>
<tr>
<td>C1 (Rep)</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>96</td>
</tr>
<tr>
<td>C3</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>97</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>97</td>
</tr>
<tr>
<td>C4</td>
<td>99</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>97</td>
<td>73</td>
<td>74</td>
<td>75</td>
<td>66</td>
</tr>
</tbody>
</table>

Table 1. Percentage identity between the nucleotide sequence of TYLCV-Is SP72/97 and those of other TYLCV-Is isolates and TYLCV-Sar isolate Sardinia

Dom. Rep., Dominican Republic. Other isolate names are abbreviated as indicated in the text.
Recombination between TYLCV-Is and ToLCV

Fig. 1. PLOTSIMILARITY diagrams (scanning window = 50) comparing the nucleotide sequences of TYLCV-Is isolates SP72/97 and Israel (A), Israel and Iran (B), SP72/97 and Iran (C) and SP72/97 and Israel/Mild (D). Separation between regions I to IV, for which differential distribution of nucleotide identity is observed, is indicated by vertical dotted lines. The first nucleotide of the region in parentheses (numbers refer to nucleotide positions in the sequence of SP72/97 TYLCV-Is reported in this work). Positions of the ORFs and of the IR are indicated at the top of the figure. Horizontal broken lines indicate the mean similarities between the sequences compared.

Regions I to IV were then considered separately to infer the phylogenetic relationships among the TYLCV-Is isolates and the two ToLCV isolates indicated above. TYLCV-Sar and the A component of a New World begomovirus species, Tomato golden mosaic virus (TGMV) (GenBank accession no. K02029), were also used in the analyses, the latter as outgroup. The software packages PHYLIP 3.5 (J. Felsenstein, University of Washington, Seattle, WA, USA) and/or CLUSTAL W (Thompson et al., 1994) were used for these analyses. Both the neighbour-joining (Saitou & Nei, 1987) and unrooted parsimony (Fitch, 1971) methods yielded similar tree topologies in all cases. Representative phylogenetic trees obtained are shown in Fig. 2. Comparisons of sequences in regions I and IV produced similar trees except that a greater genetic divergence was observed for region IV, contained in the IR, a highly variable area between geminiviruses (Padidam et al., 1995). In both cases, all TYLCV-Is isolates formed a tight cluster related to TYLCV-Sar isolate Sardinia, all of them well separated from the Bangalore-2 and Bangalore-4 isolates of ToLCV [shown in Fig. 2(A) for region I], supporting the admitted species separation (Briddon & Markham, 1995). However, when comparing sequences in regions II and III, significant topological changes occurred in the position of certain TYLCV-Is isolates. For region II, most TYLCV isolates could be grouped together, but the isolates from Israel, Cuba and the Dominican Republic grouped best with Bangalore-2 ToLCV (not shown). When region III was considered, a similar tree topology was obtained, except that, in this case, the Iran isolate also grouped with ToLCV isolates, closely related to Bangalore-4 (Fig. 2B).

The analysis described above indicated that sequences in regions II and III of the Israel isolate and region III of the Iran isolates were most closely related to the corresponding sequences of the Bangalore-2 and Bangalore-4 isolates of ToLCV (Fig. 2B). Harrison & Robinson (1999) have already suggested that ORF C1 and the IR of the Israel and Israel/Mild isolates may have been acquired from different (unidentified) parents, based on the low sequence similarity between them. In addition, Padidam et al. (1999) claimed to have statistical support for recombination events having occurred between TYLCV-Is and ToLCV. Our results suggest that the Israel and Iran isolates are hybrids between TYLCV-Is-like and ToLCV-like ancestors, the exchange having occurred between the 5’ ‘half of ORF C1 and the 5’ ‘half of the IR. The TYLCV-Is donors of the Iran and Israel chimeric genomes were closely related to each other and to other known TYLCV-Is isolates. However, the ToLCV donors differed from each other, although both were related to ToLCV isolates from India (Bangalore-2 and Bangalore-4). In nature, TYLCV-Is and ToLCV have common hosts such as tomato, so mixed infections can occur and they may replicate simultaneously in the same cell, which is a prerequisite for recombination.

The retention of the 5’-proximal part of the IR with the
breeding is the best way to control this virus. To establish effective control measures, particularly because given geographical region is essential for the correct establishment of effective control measures, particularly because the best way to control this virus.

Fig. 2. Neighbour-joining phylogenetic trees depicting topology and distance relationships of sequences in regions I (A) and III (B) defined in Fig. 1. Trees were inferred using programs in the CLUSTAL W software package. The significance of the nodes in a bootstrap analysis with 1000 replicates is shown. Vertical lengths are arbitrary, while horizontal lengths reflect genetic distance between branch nodes. Scale bars indicate the horizontal distance equivalent to 0.1 replacements per position. TGMV was used as an outgroup. Isolates reported as members of the TYLCV-Is species are shaded.

We wish to thank E. R. Bejarano for providing some of the primers used for sequencing and R. G. Milne for critical reading of the manuscript. This work was financed by the Spain–Italy Integrated Action Programme (HI1997-0159), the CICYT, Spain (AGF98-0439-C05-02), the European Union (1FD97-0271) and Caja Rural de Málaga. S. Sánchez-Campos is the recipient of a predoctoral fellowship from the Ministerio de Educación y Cultura, Spain.

References


Received 12 May 2000; Accepted 28 July 2000