Nucleotide sequences from tomato leaf curl viruses from different countries: evidence for three geographically separate branches in evolution of the coat protein of whitefly-transmitted geminiviruses

Y. G. Hong† and B. D. Harrison★

Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA and ★Department of Biological Sciences, University of Dundee, Dundee DD1 4HN, UK

The coat protein (CP) gene-containing circular DNA molecule of an isolate of tomato leaf curl geminivirus (ITmLCV; 2749 nt) obtained from southern India, and the CP genes of tomato yellow leaf curl geminivirus isolates from Nigeria and two regions of Saudi Arabia were sequenced. ITmLCV DNA had the same arrangement of ORFs, and the same pattern of repeats in the large intergenic region as is found in DNA-A of other whitefly-transmitted geminiviruses (WTGs) from the Old World. However, the sequence of ITmLCV DNA and the sequences of its predicted translation products differed substantially from those of other WTGs, including one isolate obtained from a tomato plant in northern India. Comparison of the four CP sequences deduced here with those of 18 WTGs previously studied indicated that their relationships can be represented by a tree with three branches that are unrelated to plant host species but which contain viruses from the Americas, Africa to the Middle East, and Asia to Australia, respectively. It is suggested that WTG CP evolution has proceeded along different paths in these three main regions, and that WTGs have adapted freely to new hosts in each region. Indeed, the virus isolates causing similar diseases of tomato plants in the different continents are, with few exceptions, not closely related and warrant recognition as separate species.

Introduction

Diseases described as tomato leaf curl and tomato yellow leaf curl occur in many parts of the tropics and subtropics, ranging from Central America to the Mediterranean region, Africa, Asia and Australia. They are important causes of loss in yield of tomato crops (Cohen & Harpaz, 1964; Makkouk & Laterrot, 1983; Saikia & Muniyappa, 1989) and have become increasingly prevalent in recent years. The diseases are caused by whitefly-transmitted geminiviruses (WTGs) but the genomes of different virus isolates differ in complexity and nucleotide sequence. Virus isolates from Thailand (Rochester et al., 1994) and northern India (Padidam et al., 1995a) have genomes consisting of two molecules of circular ssDNA (DNA-A and DNA-B), whereas isolates from Israel (Navot et al., 1991), Sardinia (Kheyr-Pour et al., 1992) and Australia (Dry et al., 1993) appear to have only one, which closely resembles DNA-A. Moreover, the nucleotide sequences of the different isolates, although related, are no more similar to one another than to the sequences of WTGs from other plant species.

Evidence of antigenic differences, which parallel these genomic differences, was provided by the results of tests in which extracts of naturally infected tomato leaves from many countries were allowed to react with panels of MAbs raised against the particles of two other WTGs, African cassava mosaic and Indian cassava mosaic viruses. The WTGs from tomato plants in different geographical regions proved to have consistently different epitope profiles, whereas those from the same region had similar profiles (Harrison et al., 1991; Muniyappa et al., 1991; Macintosh et al., 1992). To obtain further information on genomic and antigenic differences among tomato-infecting WTGs, we have determined the nucleotide sequences of a DNA molecule (equivalent to DNA-A of other WTGs) of a tomato leaf curl virus isolate from southern India (ITmLCV; Muniyappa et al., 1991) and of the coat protein (CP) genes of tomato yellow leaf curl virus isolates from Nigeria and two regions of Saudi Arabia. Comparison of the nucleotide sequence of DNA-A of ITmLCV with the equivalent sequence of a virus isolate from northern
India shows that there are many differences between the two isolates, although the deduced amino acid sequences of their CPs are similar. Comparison of the four new CP sequences with other published sequences identifies a consistent pattern of geographical variation among WTGs from tomato plants.

Methods

Virus isolates. Four virus isolates were studied. ITmLCV was originally obtained from a leaf curl-affected tomato plant from Bangalore, southern India; it was transmitted initially by a single whitefly (Bemisia tabaci) and subsequently was cultured in graft-inoculated tomato plants (Muniyappa et al., 1991). The other three isolates were obtained from yellow leaf curl-affected tomato plants growing at Ibadan, Nigeria (TYLCV-NIG), Al Soudieri, central Saudi Arabia (TYLCV-NSA) and Najran, southern Saudi Arabia (TYLCV-SSA) (I. Al-Shawan, B. D. Harrison & P. F. McGrath, unpublished), and were cultured in graft-inoculated tomato cv. Moneymaker plants. All inoculated plants were kept in containment conditions under licence from the Scottish Office Agriculture and Fisheries Department.

DNA extraction and PCR. Total DNA was extracted as described by Hong et al. (1993) from young leaves of tomato plants that had recently developed systemic symptoms. Viral CP genes were obtained by PCR using degenerate primers P1 and P2, and these genes were then cloned in pUC19, as previously described (Hong et al., 1993). Full-length DNA-A of ITmLCV was amplified by PCR using total DNA from infected leaf tissue as the template, and the following two primers based on sequences in the CP gene. In primer Pa [5’ d(aggagacct ACAGGCCTTCTAGGAAACATCATG) Y], the upper-case letters represent nucleotides complementary to residues 505-523. The lower-case octanucleotides in each primer contain a BamHI site. Reaction mixtures (100 μl) contained 1-2 μg total DNA, 75 pmol each of Pa and Pb, 100 μM of each dNTP, 2.5 mM-MgCl2, 50 mM-KCl, 0.1 mg/ml gelatin, 10 mM-Tris–HCl (pH 8.0) and 2.5 units of DNA polymerase (Cambio). The reaction mixtures were overlaid with 50 μl of light oil. PCR was initiated by one cycle at 94 °C for 2 min, 52 °C for 1.5 min, 72 °C for 3 min, followed by five cycles at 94 °C for 30 s, 55 °C for 1 min, 72 °C for 3 min, then by 30 cycles at 94 °C for 30 s, 58 °C for 1 min, 72 °C for 3 min, and finally by one cycle at 72 °C for 5 min, in a Cambio Intelligent Heating Block. After electrophoresis of the PCR products in a 1% LMP agarose gel (Gibco BRL) gel, the required fragment was recovered and cloned into the BamHI site of pUC19.

Sequence analysis. Nucleotide sequences were determined by the dideoxynucleotide chain termination method (Sanger et al., 1977) in both directions using plasmid DNA templates and the Klenow fragment of Escherichia coli DNA polymerase I (Pharmacia) or Sequenase Version 2.0 (United States Biochemical) with 7-deaza-dGTP in the sequencing reaction. Sequences of ITmLCV DNA by using the abutting primers Pa and Pb. Two subfragments produced by treatment with BamHI were cloned in pUC19, and the whole 2749 nt molecule (27.2% A, 19.9% C, 22.4% G and 30.5% T) was sequenced in both orientations from these two contiguous clones. Sequences at the junctions between the two fragments were confirmed by sequencing the CP gene (see below) and a fragment produced by the PCR that spanned nucleotides 125-1245.

Results and Discussion

Sequences of ITmLCV

The complete ITmLCV sequence forms a circular molecule containing, as in other Old World WTGs such as ACMV, ICMV and TYLCV-ISR, two ORFs in the virus sense and four in the complementary sense (Fig. 1). In accord with evidence indicating the site of the nick made during replication of the DNA of other WTGs (Stanley, 1995), nucleotide number 1 was assigned to the A underlined in the sequence TAATATTAC, which also occurs in the intergenic region of all other geminiviruses (Fig. 1).

Comparison of sequences of the large intergenic regions of five WTGs from tomato showed that ITmLCV
differed not only from TYLCV-ISR, TYLCV-SAR and TLCV-A (61%, 65% and 53% identity, respectively), but also from TLCV-IN (60%; Fig. 2). However, despite these differences, the 34 nucleotide sequence GCGGC-CATCCGTATA ATATTACCGG ATGGCCGCG, which occur in other WTGs, such as ICMV, EACMV and tomato. Dots represent spaces inserted to improve the match. The sixth nucleotide in the ITmLCV sequence is residue 2614 in the complete sequence.

### Table 1. Percentage amino acid sequence identity of putative products of open reading frames of 11 WTGs compared with ITmLCV

<table>
<thead>
<tr>
<th>Virus</th>
<th>V1</th>
<th>V2</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYLCV-ISR</td>
<td>74</td>
<td>69</td>
<td>78</td>
<td>67</td>
<td>68</td>
<td>72</td>
</tr>
<tr>
<td>TYLCV-SAR</td>
<td>69</td>
<td>64</td>
<td>74</td>
<td>62</td>
<td>61</td>
<td>48</td>
</tr>
<tr>
<td>TLCV-A</td>
<td>74</td>
<td>59</td>
<td>80</td>
<td>63</td>
<td>63</td>
<td>52</td>
</tr>
<tr>
<td>TLCV-IN</td>
<td>86</td>
<td>61</td>
<td>77</td>
<td>58</td>
<td>64</td>
<td>58</td>
</tr>
<tr>
<td>ICMV</td>
<td>87</td>
<td>69</td>
<td>74</td>
<td>62</td>
<td>67</td>
<td>51</td>
</tr>
<tr>
<td>ACMV-K</td>
<td>71</td>
<td>61</td>
<td>73</td>
<td>66</td>
<td>64</td>
<td>31</td>
</tr>
<tr>
<td>AbMV</td>
<td>76</td>
<td>62</td>
<td>50</td>
<td>53</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>BGMV</td>
<td>72</td>
<td>13</td>
<td>63</td>
<td>54</td>
<td>50</td>
<td>53</td>
</tr>
<tr>
<td>PYMV</td>
<td>74</td>
<td>–</td>
<td>69</td>
<td>53</td>
<td>54</td>
<td>–</td>
</tr>
<tr>
<td>SqLCV-E</td>
<td>73</td>
<td>–</td>
<td>56</td>
<td>49</td>
<td>48</td>
<td>23</td>
</tr>
<tr>
<td>TGMV</td>
<td>75</td>
<td>–</td>
<td>67</td>
<td>50</td>
<td>53</td>
<td>54</td>
</tr>
</tbody>
</table>

Fig. 2. Comparison of the sequences of the intergenic regions of five WTGs derived from tomato. Dots represent spaces inserted to improve the match. The sixth nucleotide in the ITmLCV sequence is residue 2614 in the complete sequence.

Acid sequences of the gene products of ITmLCV and 11 other WTGs provides further support for this proposition (Table 1). The products of ORFs C2, C3 and C4 are most similar to those of TYLCV-ISR whereas the C1 product is most like that of TLCV-A; the V1 product (CP) is most like that of ICMV and the V2 product is similar to the same extent to those of ICMV and TYLCV-ISR. With the possible exception of CP, the gene products of ITmLCV differ from those of other WTGs to about the same extent that other WTGs differ from one another. However, the C1, C2 and C3 products
Fig. 3. Comparison of the amino acid sequences of the CPs of nine WTGs derived from tomato. Differences from TYLCV-NSA are indicated and conserved residues are underlined; an asterisk indicates that the residue is absent. Virus acronyms are given in the text.
Tomato geminivirus CP relationships

Fig. 4. A dendrogram produced by using the UWGCG program PILEUP, showing relationships among the CPs of 22 WTGs. The three main branches contain the proteins of viruses from the Americas (upper branch), Africa/Mediterranean region/Middle East (middle branch) and Asia/Australia (lower branch). Virus acronyms are given in the text.

Coat proteins of tomato geminiviruses

To obtain a more complete picture of the relationships among tomato WTGs from different parts of the world, nucleotide sequences were determined for the CP genes of TYLCV-NIG, TYLCV-NSA and TYLCV-SSA, and the deduced amino acid sequences were compared with those of ITmLCV, TLCV-A, TLCV-IN, TYLCV-THI, TYLCV-ISR and TYLCV-SAR (Fig. 3). The proteins all consist of 256–260 residues, about 45% of which are identical in all nine sequences. Of the amino acid substitutions found, many occur between residues with similar properties, such as V/I and R/K, and most are clustered in specific parts of the sequence, notably residues 28–47, 78–93, 129–130, 147–157 and 174–221. In addition, TYLCV-SAR lacks residue 163, and TYLCV-ISR has two residues (FI) inserted at positions 214–215. Other regions, notably residues 222–256 in Fig. 3, are strongly conserved.

Comparisons of pairs of sequences show that those of TYLCV-NSA and TYLCV-ISR are 96% identical, suggesting that these two isolates are strains of the same virus. In contrast, the CP of TYLCV-ISR is distinct from that of TYLCV-NSA and of the other isolates from Mediterranean countries or West Africa (75–78% identity), and slightly more like those of the two Indian isolates, which are even more similar (86%) to one another. TYLCV-NIG CP is most similar to those of TYLCV-ISR and TYLCV-NSA (84–86%). TLCV-A CP does not closely resemble CP of any other isolate but is most similar to the proteins of the Indian and Thailand isolates (74–77%), whereas TYLCV-SAR CP most closely resembles the CPs of TYLCV-ISR and TYLCV-NSA (84–85%).

Geographical variation in geminivirus coat proteins

The similarities and differences among these tomato WTGs are put in perspective by including in the analysis the CPs of 13 other WTGs from a variety of hosts. It can then be seen that sequence divergence is independent of the natural host of each virus (Fig. 4). Thus on the basis of their CPs, WTGs from tomato in the New World resemble WTGs from other hosts in New World countries more closely than they resemble Old World WTGs from tomato. Comparable similarities in epitope profiles of the particles of WTGs from the Americas, and differences between these and Old World WTGs, were found by Swanson et al. (1992). Indeed, inspection of Fig. 4 shows that the relationships among the CPs reflect the geographical sources of the viruses. The dendrogram has three main branches, representing viruses from the Americas, Africa/Mediterranean region/Middle East and Asia/Australia, respectively. The CPs of WTGs from different geographical regions typically have 67–79% sequence identity, whereas the values for WTGs from the same region range from 80–97% mostly.

This pattern of geographically associated variation in WTG CP is suggestive of (i) an effect of spatial isolation, coupled with non-adaptive genetic drift and/or adaptive change in response to regionally specific selection pressures, and (ii) an ability of the virus lineage in a specific region to adapt to different hosts. Evidence for a difference associated with spatial isolation is provided by comparison of TYLCV-NSA and TYLCV-SSA, the
Saudi Arabian sources of which are separated by a large expanse of desert. Where no such barrier exists, as in India, the CPs of ITmLCV and TLCV-IN are considerably more alike. Evidence that a geographically related selection pressure operating on CP structure may be exerted by vector whiteflies comes from three findings: the key role of geminivirus CP specificity in effecting transmission by insect vectors (Briddon et al., 1990), the occurrence of different biotypes of B. tabaci in different countries (Costa & Brown, 1991), and the greater frequency of transmission of a WTG by a sympatric than by an allopatric biotype (McGrath & Harrison, 1995).

The association of different WTGs with similar diseases of the same plant species in different countries, and the possibility that WTGs can adapt more readily than other plant viruses to different host species, raises the questions of how WTGs should be distinguished and what criteria should be used to justify naming individual virus species. For WTGs with bipartite genomes, the ability of two viruses to form pseudo-recombinants is a strong indication that they are best considered as strains of the same virus and such findings appear to correlate with an identity of the iterated sequences in the large intergenic region of their DNA (Argüello-Astorga et al., 1994). However, tests for pseudo-recombinant production require considerable effort and are not applicable to WTGs with monopartite genomes. For WTGs in general, a convenient rule of thumb for distinguishing and naming different viruses is that their large intergenic regions should differ in nucleotide sequence and their putative gene products, other than CP, should in general have sequence identities of less than 80%. By these criteria TYLCV-ISR, TYLCV-SAR, TYLCV-THI, TLCV-A, TLCV-IN and ITmLCV are different virus species. Further data are needed to enable the status of TYLCV-NIG and TYLCV-SSA to be assessed but the difference between the CP of TYLCV-SSA and CPs of other WTGs suggests that TYLCV-SSA is likely to represent a seventh species.

The comparisons made in this paper and elsewhere (Padidam et al., 1995b) imply that WTGs are evolving rapidly and may be able to adapt more readily to new host species than plant viruses with RNA genomes. This is despite the general conclusion that the mutation rate caused by copying errors during replication is several orders of magnitude less for viral DNA than viral RNA, probably because of the host proof-reading mechanism that exists for DNA replication (Reanney, 1984). However, variants were found to occur more frequently than expected in isolates of some vertebrate-infecting viruses with ssDNA genomes (Cotmore & Tattersall, 1987; Gottschalk et al., 1991) and their occurrence may perhaps be characteristic of other eukaryote-infecting viruses with genomes of this type. If so, their existence could be a factor underlying the occurrence of swarms of variant WTG isolates that differ to greater or lesser extents in nucleotide sequence and so provide raw material on which selection pressures can act.

We thank R. N. Beachy, C. M. Fauquet, R. L. Gilbertson and M. Padidam for providing sequences before publication and our colleague David Robinson for helpful comments. We are indebted to the EEC (Contract TS2A-0137-C(CD)) for financial support.

References


gene regulation in geminiviruses. Proceedings of the National Academy of Sciences, USA 82, 3572-3576.


(Received 14 February 1995; Accepted 20 April 1995)