Multiple infection, recombination and genome relationships among begomovirus isolates found in cotton and other plants in Pakistan

Ana I. Sanz,† Aurora Fraile,† Fernando García-Arenal,† Xueping Zhou,‡ David J. Robinson,‡ Saif Khalid,§ Tahir Butt§ and Bryan D. Harrison‡

1 Departmento de Biotecnologia, E.T.S. Ingenieros Agronomos, Universidad Politecnica, 28040 Madrid, Spain
2 Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK
3 National Agricultural Research Centre, Islamabad 45500, Pakistan

Begomoviruses occur in many plant species in Pakistan and are associated with an epidemic of cotton leaf curl disease that has developed since 1985. PCR analysis with primer pairs specific for each of four already sequenced types of DNA-A of cotton leaf curl virus (CLCuV-PK types a, 26, 72b and 804a), or for okra yellow vein mosaic virus (OYVMV), indicated that many individual naturally infected plants of cotton and other malvaceous species contained two or three begomovirus sequences. Similarly, sequence differences among overlapping fragments of begomovirus DNA-A, amplified from individual naturally infected plants, indicated much multiple infection in malvaceous and non-malvaceous species. Some cotton plants contained DNA-A sequences typical of begomoviruses from non-malvaceous species, and some non-malvaceous plants contained sequences typical of CLCuV-PK. Some DNA-A sequences were chimaeric; they each included elements typical of different types of CLCuV-PK, or of different malvaceous and/or non-malvaceous begomoviruses. Often an apparent recombination site occurred at the origin of replication. No complete CLCuV-PK DNA-A sequence was found in malvaceous or non-malvaceous species collected in Pakistan outside the area of the cotton leaf curl epidemic but chimaeric sequences, including a part that was typical of CLCuV-PK DNA-A, did occur there. We suggest that recombination among such pre-existing sequences was crucial for the emergence of CLCuV-PK. Recombination, following multiple infection, could also explain the network of relationships among many of the begomoviruses found in the Indian subcontinent, and their evolutionary divergence, as a group, from begomoviruses causing similar diseases in other geographical regions.

Introduction

Members of the genus Begomovirus, family Geminiviridae, are plant viruses that have serologically related geminate particles and genomes consisting of two molecules of circular single-stranded DNA (DNA-A and DNA-B) or, less commonly, one molecule (equivalent to DNA-A; Navot et al., 1991; Harrison & Robinson, 1999), each molecule containing about 2600–2800 nucleotides. Begomoviruses are transmitted by whiteflies of the Bemisia tabaci complex (including B. argentifolii; Brown et al., 1995), which occur mainly in tropical and warm temperate areas, where the viruses cause important diseases in many dicotyledonous crops (Harrison, 1985; Brown & Bird, 1992). In recent years, a particularly serious epidemic of
cotton leaf curl disease has developed in Pakistan (Ali et al., 1995). Affected cotton (Gossypium hirsutum) plants contain begomovirus DNA-A-like molecules (Nadeem, 1995; Zhou et al., 1998), which will be referred to subsequently as DNA-A, but no DNA-B-like component has been detected (Liu et al., 1998; Mansoor et al., 1999). However, some experimentally infected test plants (Liu et al., 1998) and some naturally infected cotton plants (Mansoor et al., 1999) contained, in addition to DNA-A, circular single-stranded DNA molecules about half the size of DNA-A and derived from it by various combinations of sequence deletion, inversion, duplication and rearrangement (Liu et al., 1998). These defective molecules could be transmitted, along with DNA-A, by grafting and by B. tabaci. It is not known whether they play any important part in disease aetiology.

Recent work has also revealed another type of circular single-stranded DNA of about 1-4 kb in all leaf curl-affected cotton plants that were tested from several locations in Pakistan; this molecule resembles a genome component of nanoviruses that encodes a replication-associated protein (Rep). However, this nanovirus-like DNA becomes packaged in begomovirus coat protein and is transmissible by B. tabaci (Mansoor et al., 1999). When cotton plants were inoculated with cloned infectious molecules of the nanovirus-like component, together with cloned molecules of a DNA-A from leaf curl-affected cotton, each of these agents infected the plants systemically but only very mild symptoms developed in co-infected plants, not those typical of cotton leaf curl disease (Mansoor et al., 1999). The aetiology of the disease is therefore still unclear.

A further complication is that four different types of Pakistan cotton leaf curl virus (CLCuV-PK) DNA-A (types a, 26, 72b and 804a) have been distinguished by comparing their complete sequences, which differ by 8-29% and, for three of these types, by as much as distinct begomovirus species (Zhou et al., 1998). However, the differences are not uniformly distributed along the DNA-A sequence, so that some parts differ much more than the mean value, whereas other parts are virtually the same. It was concluded that three of the types of CLCuV-PK DNA-A are probably recombinant molecules, and all four were found to share part, but not the whole, of their sequence with DNA-A from okra plants infected by okra yellow vein mosaic virus (OYVMV; Zhou et al., 1998). However, the other parent(s) of each of the three putative recombinants is not known and the direction of recombination is therefore unclear. These data, and other analyses of partial sequences of DNA-A, suggested that begomovirus isolates possessing recombinant DNA-A molecules might be widespread in cotton and other malvaceous species in Pakistan (Zhou et al., 1998; Harrison & Robinson, 1999; Sanz et al., 1999).

Two kinds of evidence suggest that begomoviruses infecting non-malvaceous plants in Pakistan may play a part in the evolution of Pakistani cotton-infecting begomoviruses. Firstly, tests with a panel of monoclonal antibodies showed that Pakistani begomovirus isolates from a range of naturally infected non-malvaceous species were antigenically strongly related to, although mostly distinguishable from, CLCuV-PK (Harrison et al., 1997a). Indeed, begomoviruses found in various hosts in India or Pakistan are antigenically more closely related to one another than to begomoviruses associated with similar diseases in other geographical areas, such as the Americas or the African/Mediterranean region (Swanson et al., 1992; Nateshan et al., 1996; Harrison et al., 1997a). Secondly, CLCuV-PK was transmitted experimentally by B. tabaci to bean, tobacco and tomato, as well as to the malvaceous species okra (Abelmoschus esculentus), which developed okra leaf curl disease (Harrison et al., 1997a). Individual plants of a variety of species in Pakistan might therefore be co-infected with two or more begomovirus isolates. Indeed, brief records of begomovirus co-infection in three plants (one cotton, two okra) from Pakistan (Zhou et al., 1998; Sanz et al., 1999) support this idea. Co-infection is presumably a precondition for recombination to occur, and earlier work has not detected any bar to begomovirus co-infection at either the strain or species level. For instance, Lazarowitz (1991) reported co-infection of squash plants in California with two strains of squash leaf curl virus, and Harrison et al. (1997b) detected the DNA-A of two distinct begomoviruses in many severely mosaic-affected cassava plants from Uganda.

In this paper, we describe work to assess the extent to which begomovirus co-infection occurs in cotton and other plants in Pakistan, and to explore the variety of begomovirus DNA-A sequences occurring in singly or multiply infected plants. The results show that co-infection and recombination are rife in naturally infected plants, and that they involve several begomoviruses, and hosts in several botanical families. Based on this evidence, we propose a hypothesis to explain the evolutionary divergence, as a group, of begomoviruses in the Indian subcontinent from those occurring in other geographical regions.

**Methods**

**Sources of viral DNA.** In 1993, and in 1995–1997, young leaves were collected from naturally infected plants of cotton and other species showing begomovirus-like symptoms and growing at widely separated locations in Pakistan. Many samples were collected in Punjab Province (see Harrison et al., 1997a; Zhou et al., 1998; Sanz et al., 1999), especially from cotton and okra. DNA-containing extracts were prepared from the leaf tissue as described by Harrison et al. (1997a). In some instances, viral dsDNA was further purified using a plasmid isolation kit (Boehringer Mannheim) (Sanz et al., 1999). Additional leaf samples, mostly collected in Sindh Province in 1997, were used as sources of DNA from begomovirus-infected plants of other species and are listed below. All DNA samples were stored at −20 °C.

(a) Samples from malvaceous species.

1. p12 Hollyhock (Althaea rosea) from Mirpur Mathelo, north Sindh
2. p17 China rose (Hibiscus rosa-sinensis) from Moen jo Daro, central Sindh
Begomovirus multiple infection and recombination

Fig. 1. Diagram showing regions of begomovirus DNA-A amplified in PCR with different combinations of primers. The grey circle represents DNA-A, grey arrows indicate the positions of open reading frames, and the black triangle marks the origin of replication (nt 1). Black lines indicate the regions amplified with the primers represented by black squares. Products shown outside the DNA-A circle are those analysed in the overlapping sequence method; those obtained with the virus type-specific primers are shown inside the circle.

### PCR and sequence determination

The first kind of analysis was used to test for mixed infections involving any of the four main types of CLCuV-PK (clc26, clca, clc72b and clc804a) and the isolates of OYVMV described by Zhou et al. (1998). Five pairs of primers were designed, four that were specific for DNA-A of each type of CLCuV-PK and one OYVMV-specific pair that detected DNA-A of isolates of both type oyvm201 and type oyvm301. Positions of the primers are shown in Fig. 1; the primer pairs were (sequences not listed in full are given by Zhou et al., 1998): (a) for CLCuV-PK type 26, primer pair CL-CR/R2 and CL-72/F (5’ ATTCCAGGGTGTTTGTGACC 3’, identical to nt 2469–2489); (b) for CLCuV-PK type a, primer pair CL-CR/R2 and CL-AR/F2 (5’ GCGTTTGGTTTTAAACAGCTG 3’, nt 2556–2578); (c) for OYVMV, primer pair CL-CR/R2 and OYF/6 (5’ TGGTGAGAAAAGCAATGCT 3’, nt 1579–1599); (d) for CLCuV-PK type 72b, primer pair CL72-AL/R and CL72-AL/F; (e) for CLCuV-PK type 804a, primer pair CL800-AL/R and CL11/F. Each primer pair was used in a separate reaction (Harrison et al., 1997a) to test samples from malvaceous plants.

The second kind of analysis used three other pairs of primers (Sanz et al., 1999) that represent relatively well conserved sequences in DNA-A of Pakistani begomoviruses, and amplify DNA fragments with overlapping sequences (Fig. 1). These primers were used to test samples from a wider range of plant species, both malvaceous and non-malvaceous, with the aim of detecting a greater variety of begomoviruses. Excluding the primer sequences, the fragment amplified with primers IRv and IRc (Fragment 2) overlaps with 202 nt of Fragment 1 (amplified with primers AC1v and AC1c), and with 79 nt of Fragment 3 (amplified with primers CPv and CPc). Fragment 3 includes the CP gene, except for its 3′-terminal 15 nt. However, primer CL3a (Harrison et al., 1997a) was used instead of primer CPc for seven DNA samples; the fragment thus amplified (Fragment 3a) lacks 73 nt at the 3′ end of the CP gene. Each primer pair was used in a separate reaction, in conditions described by Sanz et al. (1999). The amplified fragments were cloned, and their sequences were determined with an automatic sequencer (ABI Prism, Perkin Elmer) (Sanz et al., 1999).

### Sequence analysis

The new sequences (EMBL acc. nos AJ245495–AJ245501 and AJ270853–AJ270873) obtained as described above were compared with other partial DNA-A sequences obtained by Sanz et al. (1999) (EMBL acc. nos AJ228561–AJ228599) but not previously analysed in this manner. The comparisons also included the relevant parts of the published sequences of DNA-A of the four types of CLCuV-PK and two types of OYVMV (AJ002447–AJ002459), and of...
two types of tomato leaf curl virus from India (U15015 and U38239). Sequence data were analysed with the Wisconsin Package, version 8.1 (Anon., 1994). Multiple alignments were optimized manually. A phylogenetic tree representing the CP gene (768 nt) of 33 virus isolates was constructed by a Maximum Likelihood method using PUZZLE version 4.0, with 10000 puzzling steps. Where some nucleotides at the 3' end of the CP gene were not determined, as explained above, it was assumed that the open reading frames of all isolates had the same length as that of isolate clc26, and the missing nucleotides were considered to be unknown.

**Results**

Detection of co-infection in malvaceous plants by PCR

Twenty-four samples of leaf curl-affected cotton, 16 of okra and 3 of hollyhock, collected at widely separated locations in Pakistan in 1993 or 1995–1997, were tested by PCR with primers specific for DNA-A of each of the four main types of CLCuV-PK and that of OYVMV; fragments of the expected size were consistently amplified. In cotton, DNA typical of CLCuV-PK type a was found in 1993, 1995 and 1997 (in a total of 3 samples), that of types 26 and 72b in 1995–1997 (17 and 12 samples, respectively) but that of type 804a only in 1996 (4 samples). No amplification with the OYVMV-specific primers was obtained with samples from cotton. Only one type of CLCuV-PK DNA was detected in 16 of the cotton samples (data not shown) but, of the 23 samples collected in 1995–1997, 8 contained DNA typical of two or three types (Table 1). Multiple infections were not found in samples collected near the southern (e.g. north Sindh) or western borders of the area affected by the leaf-curl epidemic (0 of 5 samples), where the disease incidence was < 1%, but were common (8 of 19 samples) in the main epidemic area, where the disease incidence was 50–100%.

In okra, OYVMV-like DNA-A was detected in 5 of the 6 samples from plants showing yellow vein mosaic symptoms, and none contained that of CLCuV-PK. In contrast, 8 of the 10 okra plants with leaf curl symptoms contained DNA typical of CLCuV-PK (types a, 26 or 804a). Four of the plants that contained CLCuV-PK DNA also yielded a product in PCR with the OYVMV-specific primers (Table 1). Also, a hollyhock plant with leaf curl symptoms apparently contained DNA of both CLCuV-PK type 804a and OYVMV, and a second hollyhock sample (S12) contained both of these, plus DNA of CLCuV-PK type 26 (Table 1). In a further test, the 1486 bp fragment amplified with the OYVMV primers from leaf curl-affected okra T19 (Table 1) was cloned and sequenced. In the intergenic region (IR), the iterons (Arguello-Astorga et al., 1994) were very like those of OYVMV isolate 201, and the whole sequence had its greatest similarity to that of OYVMV isolate 301 (88% identical). This sequence was only 74% identical with the sequence of CLCuV-PK type 804a (Zhou et al., 1998) and, in particular, lacked a sequence corresponding to the type 804a primer, CL800-AL/R, and so was distinct from the fragment amplified with the CLCuV-PK 804a primers. Thus, sample T19 was apparently co-infected, as were okra samples 206, 208 and T14. The results therefore provide evidence that multiple infection with begomoviruses occurs in both okra and hollyhock as well as in cotton.

The two leaf curl-affected okra plants in which CLCuV-PK was not detected gave a product with the OYVMV-specific

---

**Table 1. Evidence of begomovirus co-infection in malvaceous plants with leaf curl symptoms sampled in 1995–1997**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Year</th>
<th>Plant</th>
<th>Location*</th>
<th>PCR with primers specific for:†</th>
</tr>
</thead>
<tbody>
<tr>
<td>72</td>
<td>1995</td>
<td>Cotton</td>
<td>Khokhran, CP</td>
<td>clc26 clca clc72b clc804a OYVMV</td>
</tr>
<tr>
<td>722</td>
<td>1996</td>
<td>Cotton</td>
<td>Bhakkar, WP</td>
<td>+ 0 + 0 0 0</td>
</tr>
<tr>
<td>803</td>
<td>1996</td>
<td>Cotton</td>
<td>Alabad, CP</td>
<td>+ 0 + 0 0 0</td>
</tr>
<tr>
<td>802</td>
<td>1996</td>
<td>Cotton</td>
<td>Kothiwala, CP</td>
<td>+ 0 + + 0</td>
</tr>
<tr>
<td>804</td>
<td>1996</td>
<td>Cotton</td>
<td>Alabad, CP</td>
<td>+ 0 + + 0</td>
</tr>
<tr>
<td>806</td>
<td>1996</td>
<td>Cotton</td>
<td>Manisal, CP</td>
<td>+ 0 + + 0</td>
</tr>
<tr>
<td>810</td>
<td>1996</td>
<td>Cotton</td>
<td>Multan, CP</td>
<td>+ 0 + + 0</td>
</tr>
<tr>
<td>T18</td>
<td>1997</td>
<td>Cotton</td>
<td>Multan, CP</td>
<td>+ 0 + 0 0 0</td>
</tr>
<tr>
<td>206</td>
<td>1995</td>
<td>Okra</td>
<td>Faisalabad, EP</td>
<td>0 + 0 0 +</td>
</tr>
<tr>
<td>208</td>
<td>1995</td>
<td>Okra</td>
<td>Faisalabad, EP</td>
<td>0 + 0 0 +</td>
</tr>
<tr>
<td>T14</td>
<td>1997</td>
<td>Okra</td>
<td>Multan, CP</td>
<td>0 0 0 + +</td>
</tr>
<tr>
<td>T19</td>
<td>1997</td>
<td>Okra</td>
<td>Multan, CP</td>
<td>0 0 0 + +</td>
</tr>
<tr>
<td>S26</td>
<td>1997</td>
<td>Hollyhock</td>
<td>Hyderabad, SS</td>
<td>0 0 0 + +</td>
</tr>
<tr>
<td>S12</td>
<td>1997</td>
<td>Hollyhock</td>
<td>Mirpur Mathelo, NS</td>
<td>+ 0 0 + +</td>
</tr>
</tbody>
</table>

* CP, WP: central, east and west Punjab. NS, SS: north and south Sindh.
† +, Fragment of expected size amplified; 0, fragment not amplified.
primers even though yellow vein mosaic symptoms were uncommon (< 1%) or apparently absent in the source fields. Further work is needed to clarify the aetiology of these two kinds of begomovirus-associated disease symptom in okra.

**Co-infection detected by sequence comparisons**

The three primer pairs, which amplify overlapping sequences in begomovirus DNA-A, were used in tests on samples from a range of begomovirus-infected plant species. Overlapping DNA fragments were obtained from 18 samples and, in 12 of these, the overlapping sequences differed (Table 2). Hence, begomovirus co-infection seems easy to find in Pakistan in non-malvaceous plants, such as bottlegourd, tobacco and tomato, as well as in malvaceous species. However, nearly all the sequences closely resembled those of reference virus isolates (sequence identity always at least 89%, usually > 93%), as indicated in Table 2. In most instances, affinities of the non-overlapping and overlapping parts of Fragment 1 were closest to the same reference isolate; the same was true for Fragments 2 and 3a.

In tests of this kind, a pair of co-infecting viruses would theoretically yield six distinct sequence fragments but the greatest number obtained in practice was three, probably because the primers could not amplify all the relevant sequences or because not all the kinds of fragment amplified were cloned and sequenced. Nevertheless, Table 2 brings out several points. With sample 49, for example, the sequences differed at both overlaps: Fragment 1 was novel, Fragment 2 was like CLCuV-PK type a and Fragment 3 was like type 804a. Thus this sample appeared to contain at least two viruses, and possibly three, depending on whether or not the novel AC1 sequence in Fragment 1 was connected to the CLCuV-PK type 804a-like Fragment 3. With other samples, such as sample 85, three sequence fragments were amplified but only one of the two overlaps provided evidence of multiple infection. With a few samples, all three sequence fragments were amplified and no sequence differences occurred at the overlaps.

**Relationships among isolates**

When comparing 13 CP genes of Pakistani begomovirus isolates from malvaceous plants, Zhou *et al.* (1998) found that the isolates from cotton fell into three clusters, with the isolates from okra being in, or close to, two of these clusters. Twenty-two additional begomovirus CP gene sequences from Pakistani malvaceous and non-malvaceous plants are now available, so making a more comprehensive analysis possible (Fig. 2). The CP gene sequences from cotton or okra now fall into four main clusters, three of which also contain CP genes of viruses from non-malvaceous plants. The new cluster, typified by isolate olc (okra leaf curl) 90, includes the CP genes of viruses from bottlegourd, hollyhock, tobacco and a weed, all of which are virtually indistinguishable from that of olc90. Among the cotton viruses, several examples were found of two additional minor CP gene variants [typified by clc49 (1-9% nucleotide sequence difference from clc804a) and clc82 (1-3% difference from olc311)] and, among the viruses from okra or saklai

### Table 2. Components of sequence mixtures revealed by comparison of overlapping sequences in two regions of DNA-A

<table>
<thead>
<tr>
<th>Sample</th>
<th>Source</th>
<th>Source location*</th>
<th>AC1 region: overlap of Fragments 1 and 2 (202 nt)†</th>
<th>AV2–CP region: overlap of Fragments 2 and 3 (79 nt)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Cotton</td>
<td>Faisalabad, EP</td>
<td>clca + clc26</td>
<td>clca</td>
</tr>
<tr>
<td>49</td>
<td>Cotton</td>
<td>Sahiwal, EP</td>
<td>clc49 + clca</td>
<td>clca + clc804a</td>
</tr>
<tr>
<td>75</td>
<td>Cotton</td>
<td>Khokhran, CP</td>
<td>clc26 + clc72b</td>
<td>clc72b + clc804a</td>
</tr>
<tr>
<td>83</td>
<td>Cotton</td>
<td>Multan, CP</td>
<td>clc72b + clc26</td>
<td>(clc26)</td>
</tr>
<tr>
<td>85</td>
<td>Cotton</td>
<td>Multan, CP</td>
<td>(clc26)</td>
<td>clc26 + clc804a</td>
</tr>
<tr>
<td>O1</td>
<td>Okra</td>
<td>Faisalabad, EP</td>
<td>oymv201 + ToLCV-U38239</td>
<td>(clc804a)</td>
</tr>
<tr>
<td>p26</td>
<td>Hollyhock</td>
<td>Hyderabad, SS</td>
<td>clc1 + clca</td>
<td>clca + olc90</td>
</tr>
<tr>
<td>p27</td>
<td>Tobacco</td>
<td>Hyderabad, SS</td>
<td>clc72b + oymv201</td>
<td>ToLCV-U38239 + olc90</td>
</tr>
<tr>
<td>p28</td>
<td>Tomato</td>
<td>Tando, Muhammad Khan, SS</td>
<td>ToLCV-U38239 + ToLCV-U15015</td>
<td>ToLCV-U15015 + clc804a</td>
</tr>
<tr>
<td>p29</td>
<td>Bottlegourd</td>
<td>Tando, Muhammad Khan, SS</td>
<td>(ToLCV-U15015)</td>
<td>ToLCV-U15015 + olc90</td>
</tr>
<tr>
<td>p24</td>
<td>Weed</td>
<td>Mirpur Khas, SS</td>
<td>ToLCV-U38239 + ToLCV-U15015</td>
<td>ToLCV-U15015 + olc90</td>
</tr>
<tr>
<td>p25</td>
<td>Weed</td>
<td>Mirpur Khas, SS</td>
<td>clc1 + ToLCV-U15015</td>
<td>ToLCV-U15015 + olc90</td>
</tr>
</tbody>
</table>

† Affinities of the overlapping parts of the sequences are given. The first of each pair of designations refers to the sequence that comes from the DNA fragment that extends in the 5' direction and the second to that from the fragment that extends in the 3' direction from the overlap. Where the overlapping sequences were the same, the designation is shown in parentheses. See text for explanation of sequence types.
A. I. Sanz and others

**Fig. 2.** Phylogenetic tree obtained, using the PUZZLE program, from an alignment of the sequences of 33 begomovirus CP genes from Pakistan or India. Numbers above each branch are the percentage support for that branch. The scale bar indicates the horizontal distance equivalent to 0–1 replacements per position. The sources of viral DNA are cotton with leaf curl (clc prefix), okra with leaf curl (olc) or yellow vein mosaic (oyvm), saklai with leaf curl (slc) and other plant species, as indicated.

**Table 3.** Examples of putative recombinant sequences in Pakistani begomoviruses

<table>
<thead>
<tr>
<th>Sample</th>
<th>Source</th>
<th>Region sequenced</th>
<th>Affinity of sequence elements*</th>
</tr>
</thead>
<tbody>
<tr>
<td>p10</td>
<td>Cotton</td>
<td>IR</td>
<td>clc26</td>
</tr>
<tr>
<td>p12</td>
<td>Hollyhock</td>
<td>IR</td>
<td>ToLCV-U38239</td>
</tr>
<tr>
<td>p31</td>
<td>China rose</td>
<td>IR</td>
<td>clc26</td>
</tr>
<tr>
<td>p27</td>
<td>Tobacco</td>
<td>IR</td>
<td>ToLCV-U38239</td>
</tr>
<tr>
<td>74</td>
<td>Cotton</td>
<td>AC1</td>
<td>clc26</td>
</tr>
<tr>
<td>p31</td>
<td>China rose</td>
<td>AC1</td>
<td>p30</td>
</tr>
<tr>
<td>slc60</td>
<td>Saklai</td>
<td>CP</td>
<td>clc26</td>
</tr>
<tr>
<td>O1</td>
<td>Okra</td>
<td>CP</td>
<td>clc26</td>
</tr>
</tbody>
</table>

*See text for explanation of isolate types.

(Hibiscus tiliaceus; slc60), three extra variants with up to 6% sequence difference from CLCuV-PK type 804a could be recognized. Among viruses from other species, the CP genes of p28 (tomato) and p30 (watermelon) had strong affinities (< 3.3% difference) to that of CLCuV-PK type 804a (Fig. 2).

The two virus isolates from leaf curl-affected tomato in India [U15015 (Padidam et al., 1995a) and U38239] have CP genes that differ substantially from all the others but the U38239 CP gene is closest to that of CLCuV-PK type 72b (Fig. 2).

Examination of the sequences of Fragment 2 of virus isolates from non-malvaceous species, including the overlaps with Fragments 1 and 3 (Table 2), revealed that, in contrast to
Fig. 3. For legend see p. 1846.
the CP genes, most of these sequences have affinities to those of one or other of the above two tomato leaf curl viruses (ToLCV) from India. For example, the sequence of ToLCV-U15015 corresponding to Fragment 2 (713 nt) was 95.5–95.7% identical with the Fragment 2 sequences of p19 (from Solanum nigrum), p20 (Cyamopsis tetragonoloba), p28 (tomato) and p29 (bottlegourd); the p19, p28 and p29 sequences were >99% identical with that of p20. Similarly, the sequences of the 3' region of Fragment 1 of p28 and p29 were each like those of one of the ToLCV isolates (Table 2). However, one of the overlapping sequences at the 3' region of the AC1 gene, obtained from tobacco (p27), is CLCuV-PK-like. Thus some viruses from non-malvaceous plants have sequence resemblance in parts of their DNA-A to isolates from malvaceous species. Conversely, the 3' region of the AC1 gene of one component of a mixed infection of okra (sample O1; Table 2) resembles that of a ToLCV isolate.

Another complexity revealed by our sequence comparisons is that different parts of some individual sequences have different affinities and appear to have been produced by recombination. Clear-cut changes in affinity occurred at points in the IR, or in the AC1 or CP genes (Table 3). For instance, in the AC1 sequence from sample 74, the sequences brought together are typical of those of the two most disparate types of CLCuV-PK described by Zhou et al. (1998), types 26 and 72b. Similarly, the CP gene of slc60 combines sequences typical of CLCuV-PK types 26 (5' half) and 804a (3' half). Of the equivalent sequences of slc1 (from sample O1) and slc16, the 5' 700 nt closely resemble those of CLCuV-PK type 804a but the rest are unlike any of the other sequences, although similar to one another. In other instances (p12, p31), putative recombinant viral sequences from malvaceous plants incorporate elements typical of ToLCV; and part of a viral sequence from tobacco (p27) is typical of an okra-infecting virus (oyvm201; Table 3). In these last three examples, the presumed recombination site is close to the origin of replication (ori), but it cannot be located precisely because almost all the reference isolates have the same 21 nt sequence at this point. To check the reliability of the sequences determined in this work, two independent clones of 12 PCR products were sequenced. The sequences included examples of Fragments 1, 2 and 3.

In further analyses, the sequences of Fragment 2 (Fig. 3) were divided into four parts: (a) extending in the 5' direction from the IR, in the AC1 gene (252–258 nt in length), (b) the part of the IR on the 5' side of ori (145–164 nt), (c) the part of the IR on the 3' side of ori (115–142 nt), and (d) extending in the 3' direction from the IR, in the AV2 gene (181 nt). For the p27 sequence (Fig. 3), the percentage identities of these four parts with those of ToLCV-U38239 are 80, 60, 92 and 97, respectively whereas, when compared with oyvm201, they are 97, 91, 44 and 72. The converse relationships are found in the p12 sequence, which has percentage identities of 91, 91, 45 and 71 with the tomato virus, and 79, 60, 94 and 95 with oyvm201. Fragment 2 of p31 (Fig. 3) apparently has one recombination site at ori and a second near the start of the AV2 gene. Its percentage sequence identities are 95, 91, 47 and 74 with ToLCV-U38239, and 79, 55, 97 and 78 with CLCuV-PK.
Table 4. Categories of intergenic region among 31 Pakistani begomovirus sequences

<table>
<thead>
<tr>
<th>5’ half</th>
<th>3’ half</th>
<th>Iteron</th>
<th>Sequence designation*</th>
<th>Source host</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 1</td>
<td>AATTGGAGAC</td>
<td>clc12, clc26, clc62, clc82, clc85</td>
<td>Cotton</td>
<td></td>
</tr>
<tr>
<td>A 2</td>
<td>AATTGGAGAC</td>
<td>clca, clc8, clc49, clc78, clc311, p26</td>
<td>Cotton, okra, Althaea rosea</td>
<td></td>
</tr>
<tr>
<td>B 1</td>
<td>GCAATCGGTGT</td>
<td>oym s</td>
<td>Okra</td>
<td></td>
</tr>
<tr>
<td>C 1</td>
<td>TGACCA</td>
<td>oym 301</td>
<td>Okra</td>
<td></td>
</tr>
<tr>
<td>C 2</td>
<td>TTGACCA</td>
<td>p27</td>
<td>Tobacco</td>
<td></td>
</tr>
<tr>
<td>D 1</td>
<td>GGGACACT</td>
<td>clc802a, clc804a</td>
<td>Cotton</td>
<td></td>
</tr>
<tr>
<td>E 1</td>
<td>GGGACACT</td>
<td>clc72b, clc806b</td>
<td>Cotton</td>
<td></td>
</tr>
<tr>
<td>E 2</td>
<td>GGGACACT</td>
<td>clc802a, clc804a</td>
<td>Cotton</td>
<td></td>
</tr>
<tr>
<td>F 3</td>
<td>TGATCA</td>
<td>p3</td>
<td>Hibiscus rosa-sinensis</td>
<td></td>
</tr>
<tr>
<td>G 1</td>
<td>ATCGGTGT</td>
<td>O1, p12</td>
<td>Okra, A. rosea</td>
<td></td>
</tr>
<tr>
<td>G 2</td>
<td>ATCGGTGT</td>
<td>U38239†</td>
<td>Tomato</td>
<td></td>
</tr>
<tr>
<td>G 5</td>
<td>ATCGGTGT</td>
<td>clc75</td>
<td>Cotton</td>
<td></td>
</tr>
<tr>
<td>H 4</td>
<td>GGT/CGTC</td>
<td>p17, p19, p20, p28, p29, U15015†</td>
<td>H. rosa-sinensis, Solanum nigrum, Cyamopsis tetragonoloba, tomato, Luffa acutangula, Tomato</td>
<td></td>
</tr>
<tr>
<td>I 6</td>
<td>GGT/CGTC</td>
<td>p1</td>
<td>Tomato</td>
<td></td>
</tr>
</tbody>
</table>

* Designations in bold type refer to sequences determined by Zhou et al. (1998) or Sanz et al. (1999).
† Tomato leaf curl virus from Bangalore, India (EMBL acc. no. U38239).
‡ Tomato leaf curl virus from Delhi, India (EMBL acc. no. U15015).

Discussion

Our experiments were not intended to establish the aetiology of the different begomovirus-associated diseases found in Pakistan. This must be the subject of further research. Our aim was to study the kinds of begomovirus DNA-A occurring in a range of plant species and to assess the variety of interactions among the virus genomes. Our results provide much new information on these topics. They indicate that begomovirus co-infection is common in Pakistan. It was detected in 14 out of 43 plants tested by virus type-specific PCR, and in 12 out of 18 tested by the overlapping sequence method. At least 5 plants contained sequences typical of three types of DNA-A. Although the same samples were not used
for the two kinds of test, the overlapping sequence method seemed to be better at detecting the occurrence of multiple viral sequences in malvaceous species (8/12 samples) than the PCR method (14/43 samples), which used primers of greater specificity, based on more variable regions of DNA-A. Indeed, even the remarkably high figure obtained with the overlapping sequence method must be considered a minimum, because it is unlikely that all the kinds of begomovirus sequence present in the samples could be amplified with the primers used, and that all those amplified were cloned and sequenced. However, although the overlapping sequence approach was effective for detecting multiple sequences in single plants, it provides no information on the relative concentrations of the different kinds of DNA-A in them.

Evidence was obtained of co-infection in cotton, in other malvaceous species and in non-malvaceous species. Sequences typical of viruses from non-malvaceous plants were found in malvaceous ones, and sequences of CLCuV-PK were found in non-malvaceous plants. One factor favouring the spread of begomoviruses among these plants is that many dicotyledonous species in Pakistan are hosts of whiteflies of the B. tabaci complex (Ali et al., 1995), which are the known or likely vectors of all the viruses. Another possible factor, which may favour virus establishment in the vector-inoculated plants, is that begomovirus replication and/or movement proteins, which are host-adapted, may be able to mediate in trans the replication and systemic invasion of closely related co-infecting viruses that would otherwise be unable to infect the relevant plant species systemically. This phenomenon has been observed in experiments (Lazarowitz, 1991; Frischmuth et al., 1993; Ingham et al., 1995; Hou et al., 1998), and might partly explain the frequency and variety of the begomovirus sequence mixtures found in naturally infected plants in Pakistan.

The increased number of nucleotide sequences now available for Pakistani begomoviruses has provided a clearer picture of the extents of variation and affinity among them. Considerable similarities were found among the viral CP genes, with most of those from non-malvaceous plants being classified in the clusters containing CP genes obtained from malvaceous plants. This conclusion parallels that drawn from comparisons of epitope profiles of begomovirus particles (Harrison et al., 1997). In contrast, comparisons of IR sequences, the most variable part of begomovirus DNA-A (Patidam et al., 1995), provided much new evidence of variation and recombination. Additional kinds of IR sequence were obtained from infected malvaceous and non-malvaceous species. Moreover, some of the sequences from malvaceous plants incorporated elements indistinguishable from sequences of viruses infecting non-malvaceous species; and some sequence elements obtained from infected non-malvaceous plants were typical of viruses from malvaceous species. Many putative recombinant sequences had a recombination site close to ori, in agreement with previous findings (Stanley, 1995; Zhou et al., 1998; Sanz et al., 1999). Interestingly, we found no evidence for recombination in the 5’ half of the IR, or in the region (250 nt) of the AC1 gene immediately adjoining it. The association of the 5’ half of the IR and its cognate AC1 gene was therefore maintained. This is probably required for virus viability because, for the genomic DNA to be replicated, a specific amino-acid motif in the N-terminal portion of the AC1-encoded Rep protein must recognize the iteron(s) in the 5’ half of the IR (Lazarowitz et al., 1992; Fontes et al., 1994; Jupin et al., 1995). Also, iteron sequences differ among begomoviruses, and the recognition is nucleotide sequence-specific. In contrast, a putative recombination site was detected in the AV2 gene, near its junction with the IR. The frequency of supposed recombinant sequences among Pakistani begomoviruses presumably reflects the many opportunities for recombination presented by multiple infection of individual plants, together with the existence of shared sequence motifs at various points in the viral DNA-A.

The pattern of relationships among Pakistani begomoviral DNA-A sequences resembles a network, with frequent evolutionary interactions among viruses from malvaceous species, and less frequent interactions between these viruses and those from non-malvaceous species. This picture confirms and extends that provided by the complementary data of Sanz et al. (1999), who found that begomoviral sequences from cotton had essentially the same kinds of nucleotide diversity as those from other malvaceous species, and argued that all these sequences represent a single undifferentiated population. We conclude that begomoviruses in Pakistan, and probably in the whole Indian subcontinent, have a complex lineage, largely irrespective of their preferred host species, and distinct from the lineages of begomoviruses associated with similar diseases in other geographical areas, such as the Americas or the African/Mediterranean region.

We have not found the complete DNA-A of CLCuV-PK in Pakistan outside the area affected by the cotton leaf curl epidemic (Punjab and the extreme north of Sindh), although we now have evidence from the sequences found in mixed infections, and from the recombinant sequences, that pieces of sequence that are typical of each of the main types of CLCuV-PK DNA-A (a, 26, 72b and 804a) occur in begomovirus-infected malvaceous or non-malvaceous species in central or south Sindh. A form of CLCuV-PK DNA-A could presumably have emerged in the Punjab by rounds of recombination among such viruses. We have also obtained evidence that the begomovirus epidemic in cotton, in turn, has resulted in numerous multiple infections, and probably yet more rounds of recombination, which have ensured that further types of begomovirus continued to emerge in Pakistan. Whether cotton leaf curl disease in Pakistan is caused by any one, or some combination, of these begomovirus variants, perhaps in association with another virus-like agent (Liu et al., 1998; Mansoor et al., 1999), remains to be determined.

Our work was supported in part by contract CI1*-CT94-0052 from...
the European Commission. SCRI is grant-aided by the Scottish Executive Rural Affairs Department.

References


Received 14 January 2000; Accepted 3 March 2000