Available online at www.sciencedirect.com

SCIENCE DIRECT®

Virology 324 (2004) 462–474

VIROLOGY

www.elsevier.com/locate/yviro

Diversity of DNA 1: a satellite-like molecule associated with monopartite begomovirus–DNA β complexes

Rob W. Briddon,[a,*] Simon E. Bull,[a] Imran Amin,[b] Shahid Mansoor,[b] Ian D. Bedford,[a] Narayan Rishi,[c] Surender S. Siwatch,[d] Yusuf Zafar,[b] Aly M. Abdel-Salam,[e] and Peter G. Markham[a]

[a] *Department of Disease and Stress Biology, John Innes Centre, Norwich NR4 7UH, UK*
[b] *National Institute of Biotechnology and Genetic Engineering, Faisalabad, Pakistan*
[c] *Department of Plant Pathology, CCS Haryana Agricultural University, Hisar 1125004, India*
[d] *Department of Plant Breeding, CCS Haryana Agricultural University, Hisar 1125004, India*
[e] *Plant Pathology Department, Faculty of Agriculture, Cairo University, Giza, Egypt*

Received 6 November 2003; returned to author for revision 22 January 2004; accepted 30 March 2004

Available online 14 May 2004

Abstract

DNA 1 components are satellite-like, single-stranded DNA molecules associated with begomoviruses (family *Geminiviridae*) that require the satellite molecule DNA β to induce authentic disease symptoms in some hosts. They have been shown to be present in the begomovirus–DNA β complexes causing cotton leaf curl disease (CLCuD) and okra leaf curl disease (OLCD) in Pakistan as well as *Ageratum* yellow vein disease (AYVD) in Singapore. We have cloned and sequenced a further 17 DNA 1 molecules from a diverse range of plant species and geographical origins. The analysis shows that DNA 1 components are associated with the majority of begomovirus–DNA β complexes, being absent from only two of the complexes examined, both of which have their origins in Far East Asia. The sequences showed a high level of conservation as well as a common organization consisting of a single open reading frame (ORF) in the virion sense, a region of sequence rich in adenine and a predicted hairpin structure. In phylogenetic analyses, there was some evidence of grouping of DNA 1 molecules according to geographic origin, but less evidence for grouping according to host plant origin. The possible origin and function of DNA 1 components are discussed in light of these findings.

Keywords: Geminivirus; Begomovirus; Satellite; DNA 1; Evolution; Diversity

Introduction

Geminiviruses are plant-infecting, single-stranded DNA viruses that are encapsidated in twinned icosahedral particles. The family *Geminiviridae* is divided into four genera with viruses assigned according to their vector species, genome arrangement, and host range (Fauquet et al., 2003). The economically most important, geographically most widespread, and numerous geminiviruses are in the genus *Begomovirus*. Begomoviruses infect only dicotyledonous plant species and are transmitted exclusively by the whitefly *Bemisia tabaci*. The first geminivirus to be characterized at the sequence level, *African cassava mosaic virus* (ACMV; Stanley and Gay, 1983), has the typical genome arrangement of the majority of begomoviruses, consisting of two genomic components. Genomic component DNA A encodes all viral factors required for DNA replication, gene expression, and insect transmission, while the second (DNA B) encodes factors required for cell-to-cell movement in plants.

In the late 1980s, a second group of begomoviruses was identified. These viruses are monopartite, lacking the component equivalent to DNA B. The most prominent and economically significant monopartite begomovirus is *Tomato yellow leaf curl virus* (TYLCV; Navot et al., 1991). TYLCV is believed to have originated in the Middle East,

* Corresponding author. Department of Disease and Stress Biology, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK. Fax: +44-1603-450045.
E-mail address: rob.briddon@fsmail.net (R.W. Briddon).

0042-6822/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2004.03.041

but has since spread throughout the Mediterranean basin as well as to Africa, the Americas, and Asia (Czosnek and Laterrot, 1997).

Recently, a third group of begomoviruses has been identified. These viruses are monopartite but require a satellite molecule (DNA β) to induce typical disease symptoms in the plant species from which they were isolated (Briddon et al., 2001; Jose and Usha, 2003; Saunders et al., 2000, 2003; Zhou et al., 2003). DNA β components are symptom-modulating, single-stranded DNA satellites that require the helper begomovirus for replication, spread in plant tissues, and plant-to-plant transmission by the whitefly vector of begomoviruses (*B. tabaci*). However, the function of DNA β in the disease process remains unclear at this time.

In addition to the begomovirus and DNA β components, three diseases [*Ageratum* yellow vein disease (AYVD) originating from Singapore, cotton leaf curl disease (CLCuD) from Pakistan, and okra leaf curl disease (OLCD) from Pakistan] are associated with a third single-stranded DNA component termed DNA 1 (Mansoor et al., 1999, 2001; Saunders and Stanley, 1999). These satellite-like molecules encode a single product with similarity to the replication associated protein (Rep; a rolling-circle replication initiator protein) of nanoviruses; another family of plant-infecting single-stranded DNA viruses. Consequently, DNA 1 molecules are capable of autonomous replication in the cells of host plants, but require the helper begomovirus for spread in plants and insect transmission. DNA 1 appears to have no role in the disease process, being dispensable both for infectivity to and symptom induction in host plants.

Viruses of the family *Nanoviridae* have multicomponent genomes. The precise number of components associated with each virus has yet to be established. For the four confirmed members of the family {three in the genus *Nanovirus* [*Faba bean necrotic yellows virus* (FBNYV), *Subterranean clover stunt virus* (SCSV), and *Milk vetch dwarf virus* (MVDV)] and one in the genus *Babuvirus* [*Banana bunchy top virus* (BBTV)]}, between 6 and 11 components have been identified (Boevink et al., 1995; Horser et al., 2001; Katul et al., 1998; Sano et al., 1998), of which several encode Reps. However, for each species, only a single molecule encodes the Rep responsible for *trans*-replication of the non-Rep encoding components (Timchenko et al., 2000). This "master" Rep is essential for infectivity of the virus, functioning to *trans*-replicate all the remaining *bona fide* virus components. The remaining Rep-encoding components can be seen as satellite-like molecules that depend upon the helper nanovirus for systemic movement within plants and insect transmission. Whether these satellite-like components have any function in the aetiology of the disease has yet to be established.

In this study, we have assessed the geographical and sequence diversity of DNA 1 components. We show that the geographical range of DNA 1 approaches that of DNA β, but does not, apparently, extend to Far East Asia. DNA 1 has a very highly conserved sequence and arrangement. The implications of these findings are discussed.

Results

DNA 1 components are associated with the majority of begomovirus–DNA β complexes

DNA 1 components were amplified using primers designed to highly conserved regions of the Rep-encoding genes of these molecules and have been found to produce full-length products for the majority of isolates tested (Bull et al., 2003). For isolates in which no convincing PCR product was obtained, the presence of DNA 1 was established by Southern blot hybridization using an heterologous probe (CLCuD DNA 1 clone CLCuD101-Pak) and washing at low stringency (results not shown). With the exception of honeysuckle yellow vein mosaic disease (HYVMD) and *Eupatorium* yellow vein disease (EYVD; Saunders et al., 2003), all disease isolates shown by Briddon et al. (2003) to be associated with a DNA β component were also associated with a DNA 1. The presence of DNA 1 was also shown in three further isolates; AYVD originating from Uganda, *Ageratum* enation disease originating from Nepal, and *Sida* yellow vein disease originating from Nigeria, although we were unable to obtain clones of the DNA 1 components from these.

Sequence and structure of DNA 1 components

Clones of 17 DNA 1 components were obtained and sequenced in their entirety, in both orientations, with no ambiguities remaining. These sequences are available in the EMBL, DDJB, and GenBank nucleotide sequence databases under the accession numbers listed in Table 1. For the purpose of this paper and in-keeping with the nomenclature adopted for DNA β (Briddon et al., 2003), we shall name the DNA 1 molecules after the diseases with which they are associated; AYVD, CLCuD, OLCD, okra yellow vein disease (OYVD), tomato leaf curl disease (TomLCD), tobacco leaf curl disease (TobLCD), hollyhock leaf crumple disease (HLCrD), *Hibiscus* leaf curl disease (HLCD). The plant from which clone Cot101-Egy was isolated exhibited no discernable virus-like symptoms, with the possible exception of a very mild downward leaf curl. Neither a begomovirus nor a DNA β component could be detected either by cross-hybridization or by PCR using universal primers. Consequently, this DNA 1 clone is named after the plant species from which it was isolated. In addition, the sequences of four further DNA 1 molecules were included in the analysis presented here; two associated with CLCuD originating from Pakistan (CLCuD101-Pak and CLCuD102-Pak; Mansoor et al., 1999) and two associ-

Table 1
Origin and features of DNA 1 molecules

Clone	Origin (country/town [state]/year)	Plant species	Symptoms[a]	Associated begomovirus[b]	Primers	Size (nucleotides)	Predicted size of Rep [no. of amino acids, kDa]	Infectivity[c]	EMBL accession no.
CLCuD101-Pak[d,e,▲]	Pakistan/Faisalabad/1998	Cotton (*Gossypium hirsutum*)	LC,E,LE,VS,VD	CLCuMV	–	1375	315[36.5]	CLCuMV(YES)	AJ132344
CLCuD102-Pak[d,e,▲]	Pakistan/Faisalabad/1998	Cotton (*Gossypium hirsutum*)	LC,E,LE,VS,VD	CLCuMV	–	1376	315[36.5]	CLCuMV(YES)	AJ132345
AYVD101-Pak	Pakistan/Faisalabad/1994	*Ageratum conyzoides*	VY	nd	DNA101/102	1371	315[36.4][f]	ACMV (10/10)	AJ512951
AYVD102-Pak	Pakistan/Faisalabad/2000	*Ageratum conyzoides*	VY	nd	DNA101/102	1373	315[36.5][f]	ACMV (9/10)	AJ512952
AYVD103-Pak	Pakistan/Faisalabad/2000	*Ageratum conyzoides*	VY	nd	UNI101/102	1389	315[36.5]	ACMV (10/10)	AJ512948
AYVD104-Pak	Pakistan/Faisalabad/1994	*Ageratum conyzoides*	VY	nd	DNA101/102	1364	315[36.7]	ACMV (10/10)	AJ512949
HLCD101-Pak	Pakistan/Faisalabad/1994	*Hibiscus rosa-sinensis*	E,VS,VD	nd	DNA101/102	1375	315[36.6]	ACMV (9/10)	AJ512950
HLCD102-Pak	Pakistan/Faisalabad/1994	*Hibiscus rosa-sinensis*	E,VS,VD	nd	DNA101/102	1102	–	ACMV (0/10)	AJ512953
OLCD101-Pak	Pakistan/Bahawalpur/1997	Okra	LC,YM	nd	DNA101/102	1377	315[36.3]	ACMV (10/10)	AJ512954
TomLCD101-Pak	Pakistan/Rahim Yar Khan/1997	Tomato	LC,E	nd	DNA101/102	1370	315[36.5][f]	ACMV (9/10)	AJ512955
TobLCD101-Pak	Pakistan/Bahawalpur/1999	Tobacco	LC	nd	DNA101/102	1377	315[36.5]	ACMV (10/10)	AJ512956
CLCuD101-Ind	India/Dabwali[Rajasthan]/1995	Cotton (*Gossypium hirsutum*)	LC,E,LE,VS,VD	nd	DNA101/102	1379	315[36.7]	ACMV (9/10)	AJ512957
AYVD101-Ind	India/[Punjab]/1997	*Ageratum conyzoides*	VY	nd	DNA101/102	1380	315[36.5]	ACMV (9/10)	AJ512958
AYVD102-Ind	India/Lucknow/1996	*Ageratum conyzoides*	YV	nd	UNI101/102	1383	315[36.7]	ACMV (10/10)	AJ512959
Cot101-Egy	Egypt/Cairo/1995	Cotton (*Gossypium barbadense*)	mildLC	none[g]	DNA101/102	1379	315[36.5]	ACMV (10/10)	AJ512960
OYVD101-Egy	Egypt/[Fayoum]/1995	Okra	VY	nd	DNA101/102	1104	–	nd	AJ512961
HLCrD101-Egy	Egypt/Cairo/1995	Hollyhock	VY	HLCrV	DNA101/102	1103	–	nd	AJ512962
AYVD101-Sin[h]	Singapore	*Ageratum conyzoides*	VY	AYVV	–	1367	315[36.4]	AYVV (9/10)	AJ238493
AYVD201-Sin[i]	Singapore	*Ageratum conyzoides*	VY	AYVV	–	1360	295[34.1]	NO[j]	AJ416153
AYVD101-Ken[e,■]	Kenya	*Ageratum conyzoides*	VY	nd	DNA101/102	1386	315[36.5]	nd	AJ512963
AYVD102-Ken[e,■]	Kenya	*Ageratum conyzoides*	VY	nd	DNA101/102	1064	–	ACMV (0/10)	AJ512964

nd: not determined.

[a] Symptoms are denoted as leaf curling (LC), enations (E), leaf-like enations (LE), vein swelling (VS), vein darkening (VD), vein yellowing (VY), and yellow mosaic (YM).

[b] The begomovirus associated with plant from which the DNA 1 was cloned (*Cotton leaf curl Multan virus* [CLCuMV], *Hollyhock leaf crumple virus* [HLCrV], and *Ageratum yellow vein virus* [AYVV]).

[c] Infectivity following co-inoculation with either its cognate helper begomovirus or ACMV to *N. benthamiana*. For inoculations with ACMV that were done as part of this analysis, infectivity is given as the number of plants in which the DNA 1 component could be detected/number of plants inoculated.

[d] Mansoor et al. (1999).

[e] Clones with the same symbol (▲ or ■) were obtained from the same plant.

[f] The amino acid sequence of Rep was reconstructed by removal of frame-shifts or stop codons.

[g] No begomovirus was detected in this plant.

[h] Saunders and Stanley (1999).

[i] Saunders et al. (2002).

[j] Saunders and Stanley, personal communication.

ated with AYVD originating from Singapore (AYVD101-Sin, Saunders and Stanley, 1999; and AYVD201-Sin, Saunders et al., 2002).

Of the 21 sequences analyzed, 17 represent potentially full-length DNA 1 molecules. Four (HLCD102-Pak, OYVD101-Egy, HLCrD101-Egy, and AYVD102-Ken) are believed to represent deletion mutants, because each has a size below 1300 nucleotides (Table 1).

All the presumed full-length, intact DNA 1 components cloned have a highly conserved structure. They have a single predicted open reading frame (ORF), an adenine-rich (A-rich) region, and a predicted hairpin structure. A consensus genome map for DNA 1, showing the relative positions of these features, is presented in Fig. 1A.

The typical size of DNA 1 molecules presumed to be full-length was between 1364 and 1389 nucleotides with an average of 1375 nucleotides. As such the average size of DNA 1 is marginally larger than that of DNA β at 1354 nucleotides (average of the full-length DNA β molecules characterized by Briddon et al., 2003).

Analysis of the hairpin structure

With the exception of HLCD102-Pak, OYVD101-Egy, HLCrD101-Egy, and AYVD102-Ken, the sequences of all the DNA 1 components contain a predicted hairpin structure with a loop containing the sequence TAGTATTAC (the so-called "nonanucleotide sequence" [NS]) that is common to nanoviruses. Only a single molecule (AYVD102-Ken) has the NS TAATATTAC, common to geminiviruses. For both geminiviruses and nanoviruses, the NS forms part of the origin of replication and is nicked by Rep to initiate virion-strand DNA replication (Hafner et al., 1997; Heyraud-Nitschke et al., 1995; Laufs et al., 1995).

The sequences of the DNA 1 hairpin structures fall into five groups (Fig. 1B). The largest (indicated as Group 1 in Fig. 1B) consists of molecules isolated from a diverse group of host plants including *Ageratum conyzoides*, three malvaceous species, and tomato originating from the Indian subcontinent and Singapore (AYVD101-Pak, AYVD102-Pak, AYVD104-Pak, AYVD101-Sin, HLCD101-Pak, CLCuD101-Ind, OLCD101-Pak, and TomLCD101-Pak). Group 2 composes two DNA 1 molecules previously isolated from CLCuD-affected cotton originating from Pakistan (CLCuD101-Pak, CLCuD102-Pak; Mansoor et al., 1999) that share a stem loop sequence with a molecule isolated from tobacco (TobLCD101-Pak) and two molecules isolated from *A. conyzoides* (AYVD101-Ind and AYVD103-Pak) also originating from the India subcontinent. A further group (Group 3) consists of two molecules originating from Africa and isolated from cotton (Cot101-Egy) and *A. conyzoides* (AYVD101-Ken). The hairpin sequences of these three groups are very similar with a core stem sequence GAGGCT/GG.

The hairpin sequences of AYVD201-Sin and AYVD102-Ken were each unique in this analysis (Groups 4 and 5, respectively). That of AYVD102-Ken, however, shows high sequence identity to the hairpin sequences of several geminiviruses including *East African cassava mosaic Malawi virus* (EACMMV), as illustrated in Fig. 1B. This is in contrast to the remaining DNA 1 components where the hairpin and flanking sequences show high levels of sequence similarity either to other DNA 1 components or to the Rep-encoding components of nanoviruses. For AYVD201-Sin, the sequences of the hairpin and surrounding region show no particular similarity to geminiviruses, nanoviruses, DNA β, or DNA 1 components.

Three of the molecules cloned (HLCD102-Pak, OYVD101-Egy, and HLCrD101-Egy) lacked the sequences

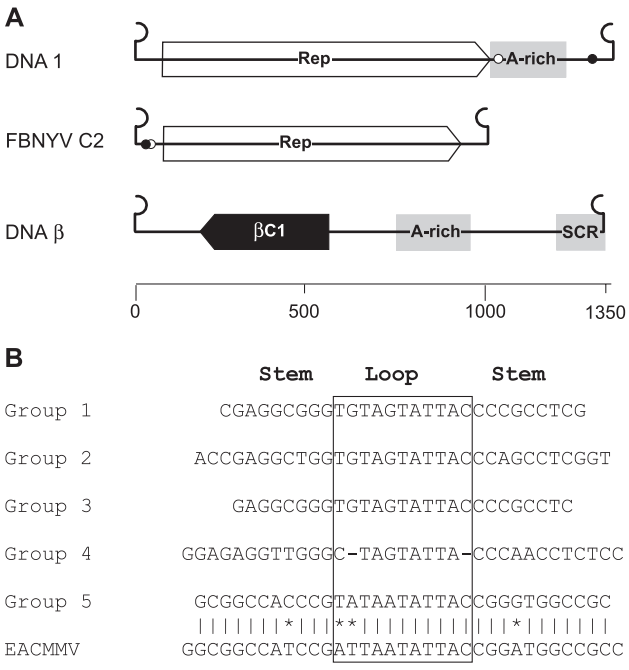


Fig. 1. Genome arrangement of DNA 1 components (A). This is compared to the consensus genome arrangement of DNA β molecules (Briddon et al., 2003) and a master Rep-encoding component of the nanovirus *Faba bean necrotic yellows virus* (component C2; accession no. X80879; Katul et al., 1995). The circular components are represented as linear DNA, starting at the nick site for the initiation of DNA replication located within the ubiquitous stem–loop structure. The position and orientation of predicted open reading frames for the replication-associated protein (Rep; of DNA 1 and the nanovirus components) and the DNA β C1 gene (βC1) are shown, as well as the A-rich regions (of DNA 1 and DNA β) and the satellite conserved region (SCR) of DNA β. The positions of consensus TATA box sequences (TATAAA; filled circle) and polyadenylation signals (AATAAA; open circle) are shown. Panel B shows an alignment of the hairpin sequences of DNA 1 components (Group 1 = TomLCD101-Pak, AYVD101-Pak, AYVD102-Pak, AYVD104-Pak, HLCD101-Pak, OLCD101-Pak, CLCuD101-Ind, and AYVD101-Sin; Group 2 = CLCuD101-Pak, CLCuD102-Pak, TobLC101-Pak, AYVD101-Ind, and AYVD103-Pak; Group 3 = Cot101-Egy and AYVD101-Ken; Group 4 = AYVD201-Sin; Group 5 = AYVD102-Ken). The positions of the stem and loop sequences are indicated. Spaces (–) are introduced to optimize the alignment. Additionally, the sequence of the hairpin of AYVD102-Ken is compared to that of the begomovirus *East African cassava mosaic Malawi virus* (EACMMV; accession number AJ006460). In this comparison, nucleotide identity is indicated by a bar (|) whereas differences are indicated with an asterisk (*).

containing the predicted hairpin structure of the presumed full-length DNA 1 molecules.

Analysis of the A-rich region

A region of sequence high in adenine (A-rich) is maintained by all the DNA 1 components immediately downstream of the Rep gene (Fig. 1), a feature in common with DNA β components (Briddon et al., 2003). For the majority of DNA 1 molecules, this A-rich region is approximately 180–200 nucleotides long with an A content of between 46% and 52% (the typical overall A content of the DNA 1 molecules examined is between 29% and 33%). For three closely related DNA 1 components originating from Africa (AYVD101-Ken, AYVD102-Ken, and Cot101-Egy), the A-rich region is shorter (approximately 100 nucleotides) with a lower A content (43%). These molecules also have an overall lower A content (approximately 28%). The A-rich sequences of the DNA 1 components are thus somewhat shorter than the equivalent sequences of DNA β molecules (typically 100–200 nucleotides for DNA 1 molecules compared to160–280 nucleotides for DNA β molecules).

Analysis of the potential coding region

With the exception of AYVD101-Pak, AYVD102-Pak, and TomLCD101-Pak, all the full-length DNA 1 molecules contain one large open reading frame (ORF) in the virion sense. This is predicted to encode a protein of 315 amino acids with a molecular weight of approximately 36.5 kDa (Table 1), except for AYVD201-Sin (as discussed later). For clones AYVD101-Pak and AYVD102-Pak, this ORF is disrupted by a single frame-shift mutation in each case, whereas that of TomLCD101-Pak is disrupted by a stop codon (Fig. 2). For the amino acid sequence analyses, the Reps of these three molecules were reconstructed. For these three isolates, we were able to show the presence of molecules containing intact Rep-encoding sequences, using specific primers to amplify only the Rep gene (results not shown), but were unable to obtain clones of these.

An alignment of the predicted amino acid sequences of the Reps of DNA 1 molecules and several other rolling-circle replication initiator proteins is shown in Fig. 3. With the exception of the Rep of AYVD201-Sin, the sequences of the Reps of DNA 1 components are very highly conserved (above 86% amino acid sequence similarity; Table 2). Among these, the Rep sequences of two molecules isolated from Africa (Cot101-Egy and AYVD101-Ken) are the most distinct. Their Rep amino acid sequences show less than 90% similarity to those of the other molecules, whereas all others show above 90% sequence similarity.

The Rep of AYVD201-Sin is very distinct from all others being both shorter (295 amino acids) and having a distinct sequence (less than 53% amino acid sequence similarity to the Reps of other molecules; Table 2). This molecule may be defective because it apparently replicates to only low levels in plant cells (Saunders et al., 2002).

For all the full-length DNA 1 components, a consensus TATA box is located 5′ of the stem–loop structure (Fig. 1). This is a feature in common with the majority of satellite-like, Rep-encoding components of nanoviruses. However, for all the master Rep-encoding components (MRCs) of nanoviruses, the TATA box is situated 3′ of the hairpin structure. A consensus polyadenylation motif (AATAAA) occurs at the 3′ terminus of the Rep gene, within the A-rich region, of all full-length DNA 1 molecules (Fig. 1).

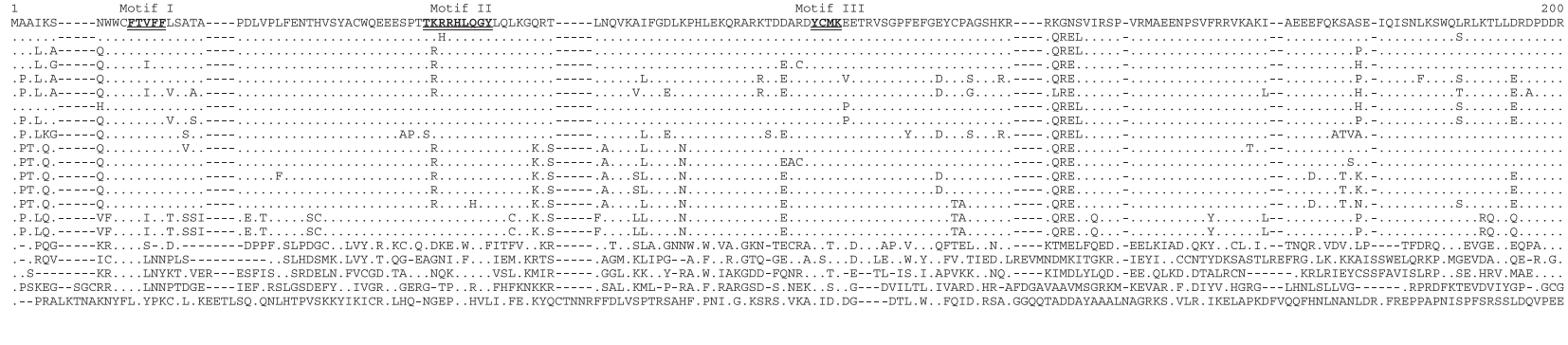
Phylogenetics of DNA 1 components

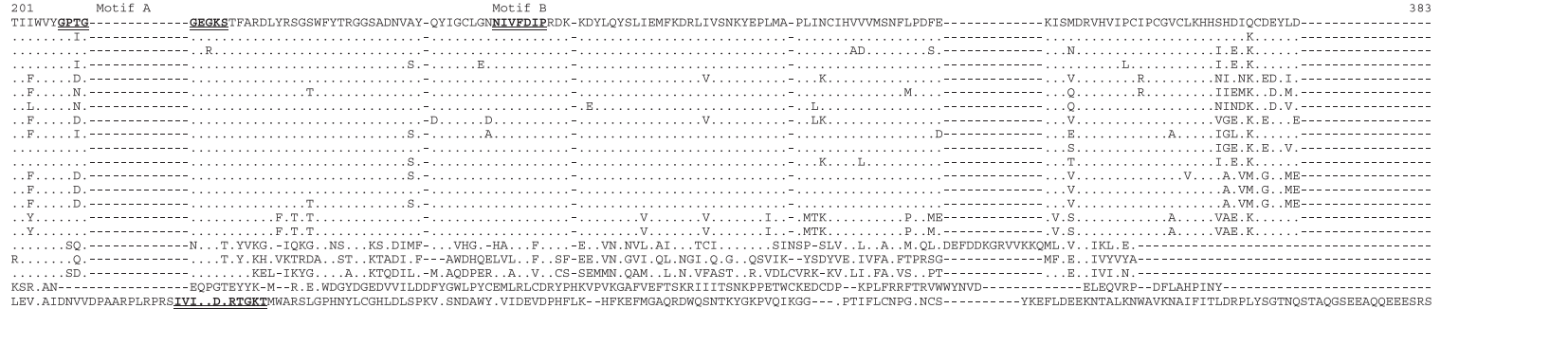
The putative phylogenetic relationships of DNA 1 molecules deduced from alignments of their complete nucleotide sequences, the predicted amino acid sequences of their Reps, the nucleotide sequences of their intergenic regions, and the nucleotide sequences of their A-rich regions are shown in Fig. 3. For all the trees, there is little evidence for clustering of DNA 1 molecules according to the host from which they were isolated. For example, the DNA 1 molecules isolated from cotton do not group. There is, however, some evidence of grouping according to geographic origin. Three molecules originating from Africa (Cot101-Egy, AYVD101-Ken, and AYVD102-Ken) form a cluster that is distinct from the DNA 1s originating from Asia. However, two further molecules originating from Africa (HLCrD101-Egy and OYVD101-Egy), both of which are defective and lack the hairpin region, do not segregate with the other African DNA 1s.

In these analyses, AYVD201-Sin is distinct from and basal to the remaining DNA 1 components for all but the tree based on the A-rich sequence. The position of this molecule in all dendrograms except that based on the intergenic sequence is not well supported by bootstrapping, indicating that its relationship to the remaining DNA 1 molecules is unclear.

For the trees based on the complete nucleotide sequence (A) and the amino acid sequences of Rep (B), it is evident that the DNA 1 molecules, with the exception of AYVD201-Sin,

Fig. 2. Alignment of the predicted amino acid sequences of the replication-associated proteins (Rep) of DNA 1 molecules. These are compared to the Rep sequences of other rolling-circle replicons; the begomovirus *Cotton leaf curl Multan virus* (CLCuMV; accession number AJ132439), the circovirus *Beak and feather disease virus* (BFDV; AF311297), a satellite-like (component C1; X80879), and the master Rep (component C2; Y11405) of the nanovirus *Faba bean necrotic yellows virus* (FBNYV). Gaps (–) were introduced to optimize the alignment. Only amino acids that differ from AYVD101-Pak are shown. Amino acid identity is indicated with a dot (.). The position of the putative motifs involved in rolling-circle DNA replication (Motifs I to III; Ilyina and Koonin, 1992) including the presumed nicking tyrosine of Motif III and the NTP-binding site (Motifs A and B; $EGX_4GKT_{22}DD$; Walker et al., 1982) of DNA 1 is indicated. The positions of the frame-shift mutations of clones AYVD101-Pak and AYVD102-Pak are indicated by single underlining. The position of the stop codon in TomLCD101-Pak is indicated by an asterisk (*).


1
Motif I
Motif II
Motif III
200
AYVD101-Pak
AYVD102-Pak
HLCD101-Pak
AYVD102-Ind
AYVD104-Pak
CLCuD101-Ind
OLCD101-Pak
TomLCD101-Pak
AYVD101-Sin
AYVD101-Ind
AYVD103-Pak
CLCuD101-Pak
TobLCD101-Pak
CLCuD102-Pak
Cot101-Egy
AYVD101-Ken
AYVD201-Sin
FBNYV C2
FBNYV C1
BFDV
CLCuMV


201
Motif A
Motif B
383
AYVD101-Pak
AYVD102-Pak
HLCD101-Pak
AYVD102-Ind
AYVD104-Pak
CLCuD101-Ind
OLCD101-Pak
TomLCD101-Pak
AYVD101-Sin
AYVD101-Ind
AYVD103-Pak
CLCuD101-Pak
TobLCD101-Pak
CLCuD102-Pak
Cot101-Egy
AYVD101-Ken
AYVD201-Sin
FBNYV C2
FBNYV C1
BFDV
CLCuMV

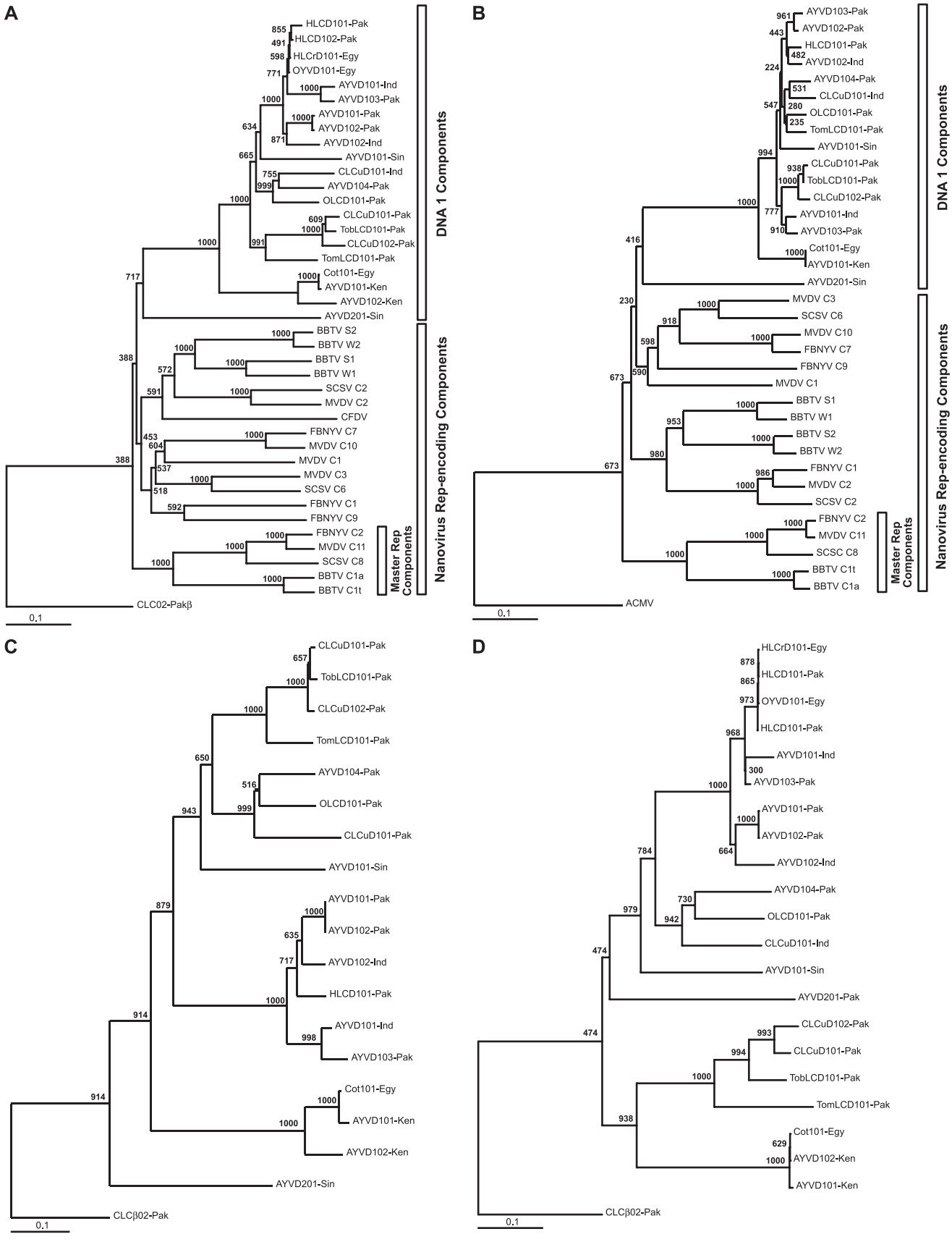

A
HLCD101-Pak
HLCD102-Pak
HLCrD101-Egy
OYVD101-Egy
AYVD101-Ind
AYVD103-Pak
AYVD101-Pak
AYVD102-Pak
AYVD102-Ind
AYVD101-Sin
CLCuD101-Ind
AYVD104-Pak
OLCD101-Pak
CLCuD101-Pak
TobLCD101-Pak
CLCuD102-Pak
TomLCD101-Pak
Cot101-Egy
AYVD101-Ken
AYVD102-Ken
AYVD201-Sin
BBTV S2
BBTV W2
BBTV S1
BBTV W1
SCSV C2
MVDV C2
CFDV
FBNYV C7
MVDV C10
MVDV C1
MVDV C3
SCSV C6
FBNYV C1
FBNYV C9
FBNYV C2
MVDV C11
SCSV C8
BBTV C1a
BBTV C1t
CLC02-Pakβ
0.1
DNA 1 Components
Nanovirus Rep-encoding Components
Master Rep Components
B
AYVD103-Pak
AYVD102-Pak
HLCD101-Pak
AYVD102-Ind
AYVD104-Pak
CLCuD101-Ind
OLCD101-Pak
TomLCD101-Pak
AYVD101-Sin
CLCuD101-Pak
TobLCD101-Pak
CLCuD102-Pak
AYVD101-Ind
AYVD103-Pak
Cot101-Egy
AYVD101-Ken
AYVD201-Sin
MVDV C3
SCSV C6
MVDV C10
FBNYV C7
FBNYV C9
MVDV C1
BBTV S1
BBTV W1
BBTV S2
BBTV W2
FBNYV C1
MVDV C2
SCSV C2
FBNYV C2
MVDV C11
SCSC C8
BBTV C1t
BBTV C1a
ACMV
0.1
DNA 1 Components
Nanovirus Rep-encoding Components
Master Rep Components
C
CLCuD101-Pak
TobLCD101-Pak
CLCuD102-Pak
TomLCD101-Pak
AYVD104-Pak
OLCD101-Pak
CLCuD101-Pak
AYVD101-Sin
AYVD101-Pak
AYVD102-Pak
AYVD102-Ind
HLCD101-Pak
AYVD101-Ind
AYVD103-Pak
Cot101-Egy
AYVD101-Ken
AYVD102-Ken
AYVD201-Sin
CLCβ02-Pak
0.1
D
HLCrD101-Egy
HLCD101-Pak
OYVD101-Egy
HLCD101-Pak
AYVD101-Ind
AYVD103-Pak
AYVD101-Pak
AYVD102-Pak
AYVD102-Ind
AYVD104-Pak
OLCD101-Pak
CLCuD101-Ind
AYVD101-Sin
AYVD201-Pak
CLCuD102-Pak
CLCuD101-Pak
TobLCD101-Pak
TomLCD101-Pak
Cot101-Egy
AYVD102-Ken
AYVD101-Ken
CLCβ02-Pak
0.1

Table 2
Pairwise percentage nucleotide sequence identity (above the diagonal) and predicted amino acid sequence similarities Rep (below the diagonal) of DNA 1 components

Clone	CLCuD101-Pak	CLCuD102-Pak	AYVD101-Pak	AYVD102-Pak	AYVD103-Pak	AYVD104-Pak	HLCD101-Pak	HLCD102-Pak	OLCD101-Pak	TomLCD101-Pak	TobLCD101-Pak	CLCuD101-Ind	AYVD101-Ind	AYVD102-Ind	Cot101-Egy	OYVD101-Egy	HLCrD101-Egy	AYVD101-Sin	AYVD201-Sin	AYVD101-Ken	AYVD102-Ken
CLCuD101-Pak	100	94.7	78.5	78.4	81.8	75.0	77.2	83.9	75.4	81.4	96.2	73.3	78.8	78.1	69.5	84.0	84.0	72.7	48.1	70.3	66.6
CLCuD102-Pak	96.8	100	76.9	77.0	79.4	75.2	76.0	82.2	73.1	81.0	93.1	71.7	77.1	75.4	72.1	82.2	82.2	72.0	48.7	70.7	66.0
AYVD101-Pak	92.7	91.7	100	99.4	86.8	80.9	90.5	93.6	82.9	80.7	79.7	76.8	87.6	90.2	72.3	93.5	93.5	79.0	48.7	70.4	66.4
AYVD102-Pak	92.7	92.4	97.5	100	86.4	81.0	90.4	93.7	83.1	80.8	79.6	76.9	87.2	90.0	72.5	93.6	93.6	79.1	51.3	70.6	66.4
AYVD103-Pak	95.0	93.7	93.7	94.0	100	76.4	88.3	91.9	76.1	77.2	82.4	73.4	95.4	85.0	69.4	91.8	91.8	76.0	49.2	70.5	65.7
AYVD104-Pak	92.1	91.7	93.7	94.6	92.7	100	80.0	83.8	85.4	80.2	75.8	69.1	75.6	79.7	69.1	83.9	83.8	77.3	48.1	69.2	65.5
HLCD101-Pak	91.7	90.8	95.6	96.2	94.0	94.3	100	97.7	79.8	77.7	78.3	78.0	87.7	91.8	70.3	97.7	97.7	78.5	51.1	69.8	66.0
HLCD102-Pak	-	-	-	-	-	-	-	100	85.5	82.3	85.3	81.6	92.2	92.4	75.2	99.7	99.9	81.7	51.0	76.1	70.4
OLCD101-Pak	91.4	91.1	94.9	96.5	92.4	95.2	94.9	-	100	79.6	75.7	80.7	77.3	79.2	71.1	85.5	85.5	75.3	48.3	68.5	63.7
TomLCD101-Pak	92.4	92.1	93.7	95.2	93.0	95.9	94.6	-	94.9	100	82.2	77.4	75.6	76.4	70.3	83.4	83.4	74.6	48.5	70.7	69.6
TobLCD101-Pak	99.0	97.1	93.7	93.7	95.2	93.0	92.7	-	92.4	93.3	100	74.1	80.1	77.3	72.0	85.3	85.3	73.4	48.1	70.6	73.4
CLCuD102-Ind	91.1	90.8	93.0	93.3	91.4	94.6	93.3	-	94.0	94.0	92.1	100	74.5	74.5	67.4	81.7	81.6	76.5	47.4	67.0	62.3
AYVD101-Ind	94.9	94.0	94.3	94.3	96.8	93.3	94.0	-	93.7	93.3	95.9	92.1	100	85.1	71.4	92.1	92.1	74.8	49.0	70.7	66.8
AYVD102-Ind	92.7	91.7	95.9	96.5	94.9	94.0	95.9	-	93.7	94.0	93.0	93.7	94.0	100	67.3	92.4	92.5	78.8	49.9	68.9	64.6
Cot101-Egy	88.9	89.5	87.9	87.9	89.5	88.3	88.3	-	87.3	89.5	89.8	88.3	90.2	87.6	100	75.1	75.1	67.1	48.1	99.6	91.1
OYVD101-Egy	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100	99.9	81.7	50.7	75.9	70.5
HLCrD101-Egy	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100	81.7	51.5	75.9	70.5
AYVD101-Sin	91.1	90.2	91.7	93.0	91.4	92.4	92.7	-	92.4	-	91.4	91.7	91.7	93.7	86.7	-	-	100	48.4	66.3	61.3
AYVD201-Sin	51.4	51.7	52.7	52.7	50.7	52.3	50.3	-	52.7	52.3	51.7	50.3	50.3	52.7	51.0	-	-	50.7	100	47.5	46.7
AYVD101-Ken	88.9	89.5	87.9	87.9	89.5	88.3	88.3	-	87.3	89.5	89.8	88.3	90.2	87.6	100	-	-	86.7	51.0	100	90.5
AYVD102-Ken	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100

Nucleotide sequence identities above 90% and amino acid sequence similarities above 95% are highlighted in grey.

are more closely related to each other than the Rep-encoding components of nanoviruses are to each other. This is indicated by the considerably shorter mutation distances of DNA 1s, particularly for the Rep sequence tree. This may indicate that either the constraints on variation are tighter for DNA 1 than nanovirus Rep-encoding components or, more likely, that DNA 1 evolved relatively recently.

Infectivity of DNA 1 components

Assessment of the biological integrity of the cloned DNA 1 components was based upon the finding of Saunders and Stanley (1999) that these molecules can be maintained (encapsidated and moved *in trans*) by the bipartite begomovirus *African cassava mosaic virus* (ACMV); a virus not normally associated with DNA 1. Three of the molecules (CLCuD101-Pak, CLCuD102-Pak, and AYVD101-Sin) have been shown previously to be infectious to plants in the presence of their cognate helper begomovirus (Mansoor et al., 1999; Saunders and Stanley, 1999). Although it was shown capable of autonomous replication in leaf disk assays (Saunders et al., 2002), AYVD201-Sin was not capable of maintenance in plants by its cognate helper begomovirus (Saunders and Stanley, personal communication).

The majority of the cloned DNA 1 components were shown to be competent for maintenance by ACMV in *Nicotiana benthamiana* (Table 1). Fig. 4 shows the maintenance of three DNA 1 molecules by ACMV. The levels of each molecule maintained in plants differed, possibly indicating some adaptation between DNA 1 and its helper begomovirus. Surprisingly, three DNA 1 clones (AYVD102-Pak, AYVD103-Pak, and TomLCD101-Pak) that have mutated Rep coding sequences were capable of maintenance by ACMV (Table 1, Fig. 4). Only a single molecule tested (HLCD102-Pak) was not maintained by either ACMV or its cognate helper begomovirus. This molecule lacks both the hairpin sequence, which is the

Fig. 3. Neighbor-joining phylogenetic dendrograms based on alignments of the complete nucleotide sequences of DNA 1 components (A), the predicted amino acid sequences of their Rep genes (B), the nucleotide sequences of their intergenic regions (excluding the A-rich sequences) (C), and the nucleotide sequences of the A-rich regions (D). Vertical distances are arbitrary. Horizontal distances are proportional to calculated mutation distances. Numbers at nodes indicate bootstrap scores (1000 replicates). The trees were rooted on the complete nucleotide sequence of cotton leaf curl disease DNA β (an unrelated molecule of a similar size; AJ292769) and the predicted amino acid sequence of Rep of *African cassava mosaic virus* (J02057) for the nucleotide and amino acid sequence based analyses, respectively. Additional sequences used in the comparisons were *Banana bunchy top virus* (BBTV) components 1a, 1t, S1, S2, W1, and W2 (AR010225, AR010233, AF216221, AF216222, L32166, and L32167, respectively), *Subterranean clover stunt virus* (SCSV) components C2, C6, and C8 (U16731, U16736, and AJ290434, respectively), *Faba bean necrotic yellows virus* (FBNYV) components C1, C2, C7, and C9 (X80879, Y11405, AJ005964, and AJ005966, respectively), *Milk vetch dwarf virus* (MVDV) components C1, C2, C3, C10, and C11 (AB000920, AB000921, AB000922, AB009047, and AB027511, respectively), and *Coconut foliar decay virus* (CFDV, M29963).

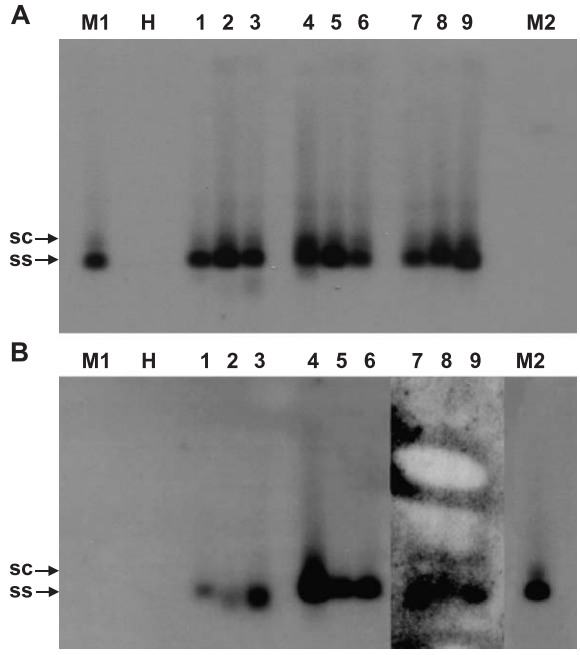


Fig. 4. Detection of ACMV DNA A and DNA 1s in biolistically inoculated *N. benthamiana*. Samples were extracted from individual plants inoculated with ACMV (lane M1) or co-inoculated with ACMV and clone OLCD101-Pak (lanes 1–3), ACMV and AYVD101-Pak (lanes 4–6), or ACMV and clone CLCuD101-Ind (lanes 7–9). The sample in lane M2 was extracted from a *N. benthamiana* plant graft inoculated with a field isolate of CLCuD originating from Pakistan. Equivalent amounts of nucleic acids (approximately 10 μg) were loaded in each lane. Lanes 7–9 of panel B were exposed to film approximately three times longer (48 h) than the remaining lanes. Blots were hybridized to ACMV DNA A (panel A) and CLCuD DNA 1 (clone CLCuD101-Pak; panel B) probes and were washed at high (0.1× SSC, 65 °C) and intermediate (2× SSC, 50 °C) stringency, respectively. The positions of single-stranded (ss) and supercoiled (sc) DNAs are shown.

presumed origin of virion-strand DNA replication, as well as an intact Rep coding sequence.

The symptoms induced following co-infection of *N. benthamiana* with ACMV and DNA 1 were qualitatively similar to plants infected with only ACMV. However, the onset of symptoms was delayed by approximately 1 day and plants were marginally less stunted than control plants infected with only ACMV (results not shown). This is consistent with previous findings (Mansoor et al., 1999; Saunders and Stanley, 1999).

Discussion

Recently, a new class of begomoviruses was identified that is associated with two additional single-stranded DNA components (Mansoor et al., 2003). These begomoviruses are monopartite and require a symptom-modulating satellite molecule (DNA β) to induce symptomatic infections in some hosts. The second additional component (DNA 1) is satellite-like, being capable of autonomous replication, but apparently dispensable for symptomatic infection (Mansoor et al., 1999; Saunders and Stanley, 1999).

The analysis conducted here has shown that, with the exception of two isolates from the Far East (HYVMD and EYVD), all begomovirus–DNA β complexes investigated are associated with a DNA 1 component. This indicates that the association of DNA 1 with begomovirus–DNA β complexes is the norm and that the geographic distribution of DNA 1 approaches that of DNA β.

The presumed full-length DNA 1 components have a conserved size (between 1364 and 1389 nucleotides) and three highly conserved features; an A-rich region, a predicted hairpin structure, and a single open reading frame in the virion sense that encodes a rolling-circle initiator protein.

The A-rich sequence is the only feature that distinguishes the begomovirus-associated DNA 1 components from Rep-encoding components of nanoviruses, from which they are believed to have originated. It has been suggested that the A-rich sequences function merely to increase the size of the molecule to half that of a begomovirus component (approximately 1350 nucleotides; Mansoor et al., 1999; Saunders and Stanley, 1999; Saunders et al., 2002). A strict size selection, for molecules that are a quarter, half, or unit (approximately 2800 nucleotides) length, occurs during begomovirus movement, necessitating such an increase (Etessami et al., 1989; Rojas et al., 1998). Additionally, Frischmuth et al. (2001) have shown that defective interfering DNA (naturally occurring half unit length DNA molecules derived from the DNA B components of many bipartite begomovirus; Stanley et al., 1990) is preferentially encapsidated in monomeric capsids. This suggests that DNA 1 and DNA β may be similarly encapsidated in monomeric, rather than geminate, capsids.

Were there not a selective advantage, the A-rich sequence of DNA 1 would be expected to be rapidly lost by genetic drift. However, the maintenance of this sequence by molecules occurring over such a wide geographic area and diversity of host shows that this is not the case, indicating that it may have some function. The A-rich region contains consensus polyadenylation motifs, possibly indicating that it is a diffuse transcription terminator. For DNA β components, the similar A-rich region has been suggested to possibly play a role in complementary-sense DNA replication (Briddon et al., 2003).

For several of the DNA 1 molecules, the A-rich region contains small (approximately 30 nucleotides) duplicated sequences (results not shown). This may indicate that the A-rich sequences have arisen by intramolecular duplication of sequences rather than by intermolecular recombination. Also, the A-rich region of the DNA 1 molecules is shifted downstream (closer to the stem–loop structure) in comparison to the DNA β molecules (Fig. 1). This is presumably necessary to accommodate the large coding sequence for Rep.

The only polypeptide predicted to be encoded by DNA 1 components is a rolling-circle replication initiator protein

that has sequence similarity to the products of nanovirus Rep-encoding components (Figs. 2 and 3; Mansoor et al., 1999; Saunders and Stanley, 1999; Saunders et al., 2002). With the exception of AYVD201-Sin, the Reps encoded by DNA 1 molecules are very highly conserved, all being 315 amino acids long and having more than 86% amino acid sequence similarity. Thus, there has been a size increase of the coding capacity of *rep* from approximately 280 amino acids in the nanoviruses to 315 amino acids in DNA 1. This may indicate that the size of nanovirus Reps is suboptimal for their function (possibly due to packaging constraints) or is adapted to the smaller genomic size of nanoviruses (necessitating a size increase when interacting with the larger DNA 1).

In contrast to DNA β components, which clearly segregate by host and geographic origin (Briddon et al., 2003), the DNA 1 molecules show only a loose grouping according to geographic origin, although the molecules originating from Africa were distinct. Segregation according to host plant origin was less evident, which may indicate that DNA 1 molecules have no particular affinity to the helper begomovirus with which they are associated. These results are consistent with DNA 1 components being highly mobile elements with a wider host range, relative to DNA β, which are readily exchanged between different begomovirus–DNA β complexes. This essentially parallels the situation in nanoviruses. The satellite-like, Rep-encoding components of viruses of the genus *Nanovirus* show no particular affinity to their associated virus or host, although the monocot-infecting babuvirus molecules are distinct from the dicot-infecting nanovirus components. The DNA 1 components are also far more conserved than, as well as distinct from, the master Rep-encoding nanovirus components. Clearly the constraints to genetic diversification of the master Rep-encoding components (MRCs) are greater than those acting on the satellite Rep-encoding components (SRCs). For the MRCs, these constraints must include the ability to be encapsidated and *trans*-replicate other genomic components. These are likely to be the determinative factors reducing the diversity of the MRCs relative to the SRCs. Therefore, the low diversity of DNA 1 components relative to the SRCs indicates either a strong constraint to sequence divergence or, more likely, that the association of DNA 1 with begomovirus–DNA β complexes is only a relatively recent phenomenon.

AYVD201-Sin is the most diverse DNA 1 molecule associated with begomovirus–DNA β complexes in this analysis. It has a distinct hairpin sequence, a shorter Rep coding sequence, and a distinct Rep amino acid sequence that is basal to all the other DNA 1 sequences in most of the phylogenetic analyses. Consequently, it may represent a distinct class of DNA 1 molecules, possibly with a different origin, and thus justify a separate taxonomic classification to the remaining DNA 1 molecules. However, this should await more detailed studies to determine whether the clone obtained by Saunders et al. (2002) is intact, functional, and representative. Although AYVD201-Sin was capable of autonomous replication in leaf disk assays, it was not maintained by AYVV in plants (Saunders and Stanley, personal communication).

Clone Cot101-Egy was obtained from an apparently symptomless plant in an area that has no recent reports of the occurrence of cotton-infecting geminiviruses. Despite extensive efforts, neither a geminivirus nor a DNA β satellite could be detected in total DNA samples extracted from this plant. Similarly, there was no evidence for the presence of a nanovirus infection (results not shown), raising the question about the origin of this DNA 1 component. Because levels of whitefly infestation in the cotton field were high at the time of sampling, the most likely explanation is that it originates from another host plant species harboring a begomovirus–DNA β–DNA 1 complex whose helper virus' host range does not include cotton. The presence of a polyphagous biotype of *B. tabaci* could thus lead to inoculation of the complex to cotton. Because DNA 1 does not seem to suffer the host range constraints of many begomoviruses, it is possible that sufficient cells would become infected to be able detect DNA 1 by PCR in the absence of the helper begomovirus and DNA β.

In the analysis presented here, three DNA 1 components originating from north Africa are distinct from the other molecules. They have a distinct Rep sequence, a shorter A-rich region, and are basal in the phylogenetic analyses (with the exception of AYVD201-Sin, as discussed earlier). Interestingly, previous studies have also shown that the African DNA β components are distinct, suggesting that these begomovirus–DNA β–DNA 1 complexes are evolving independently due to geographic isolation (Briddon et al., 2003).

Of interest are the three defective molecules (HLCD102-Pak, OYVD101-Egy, and HLCrD101-Egy) with very similar sequences (>99%). While we cannot categorically rule out cross-contamination between these isolates, the fact that they were collected, isolated, and cloned at different times, and their sequences are not identical, as well as all being isolated from malvaceous hosts, makes this unlikely. Consequently, this may indicate that there has been an exchange of infected material between Asia and Africa, although these molecules undoubtedly have their origins on the Indian subcontinent.

Co-inoculation of the DNA 1 components with ACMV to *N. benthamiana* indicated that the majority of molecules cloned were biologically active and capable of systemic spread by a begomovirus. Surprisingly, three DNA 1 clones with mutations in their Rep coding sequences (AYVD101-Pak, AYVD102-Pak, and TomLCD101-Pak) were also maintained by ACMV. Whether these molecules are capable of autonomous replication or were *trans*-replicated by ACMV was not determined. It is possible that maintenance of these molecules was due to recombination resulting in the insertion of the ACMV origin of replication into the DNA 1 molecule. Such recombination events are common for

geminiviruses and these satellite molecules (Briddon et al., 2000; Saunders et al., 2000, 2001). However, we cannot rule out the possibility that autonomous replication, at least in some cases, is not a prerequisite for maintenance of DNA 1 by a begomovirus.

The function(s) of the DNA 1 components of begomovirus–DNA β complexes remains unclear. The near ubiquitous association of DNA 1 components with complexes suggests they have a role to play. Although not essential, they may moderate the infection by "mopping-up" cellular resources. The reduction in the level of (helper) viral DNA in infected plants in the presence of DNA 1 certainly supports this hypothesis (Saunders et al., 2002). A similar role has been suggested for the defective interfering (DI) molecules of bipartite begomoviruses. DI molecules are deletion products derived from DNA B that are approximately half component length (Frischmuth and Stanley, 1991).

Given the common association of DNA 1 components with begomovirus–DNA β complexes, its absence for HYVMD and EYVD is interesting. Either these complexes do not require the function(s) provided by DNA 1 or they are provided by some other mechanism or factor. One possibility is that complexes lacking a DNA 1 instead use defective molecules derived from the other components. All begomovirus–DNA β complexes thus far analyzed produce a large number of defective or recombinant molecules (Briddon et al., 2000, 2001, 2003; Saunders et al., 2000, 2001; Stanley et al., 1997). This is particularly evident for HYVMD (R.W. Briddon, unpublished results). Whether the begomovirus–DNA β complexes causing HYVMD and EYVD evolved before the capture of DNA 1 by the progenitor begomovirus–DNA β complex or whether they subsequently lost their DNA 1 components is debatable. Characterization of further disease complexes occurring in the Far East may provide an answer to this question.

The origin of DNA 1 undoubtedly lies with the nanoviruses, and the available evidence suggests that the association of DNA 1 with begomovirus–DNA β complexes is a relatively recent occurrence. However, does the diversity of the molecules we have detected result from a single or multiple capture events followed by divergence? The lack of overall sequence diversity and the similarity of their stem loop structures would suggest that, with the exception of AYVD201-Sin, the present diversity of DNA 1 results from a single capture event. AYVD201-Sin is so distinct from other DNA 1 that, should it prove to be more than just a unique recombinant, it must indicate a second capture event.

Materials and methods

Origins and maintenance of disease isolates

The origins of plant materials and the symptoms exhibited are indicated in Table 1. Of these, cotton (*Gossypium hirsutum*) plants from Pakistan and India exhibiting typical symptoms of CLCuD (Briddon and Markham, 2000), *A. conyzoides* plants from Pakistan, India, and Kenya exhibiting typical symptoms of AYVD (Tan et al., 1995), and *Hibiscus rosa-sinensis* plants exhibiting typical symptoms of HLCD were maintained at 25 °C with supplementary lighting to give a 16-h photoperiod. At regular intervals, CLCuD-affected plants were moved to a growth cabinet at 35 °C with a photoperiod of 16 h to maintain symptoms. For all other isolates, leaves of infected plants were collected in the field and maintained at −20 °C before extraction of nucleic acids.

Production of DNA 1 clones

Nucleic acids were extracted from the leaves of plants by the CTAB method or using a Phytopure DNA extraction kit (Amersham). Clones of DNA 1 molecules were produced by PCR-mediated amplification with one of two pairs of "universal" primers described in Bull et al. (2003) (Table 1). PCR products were cloned into the pGem T-Easy vector (Promega) as per the manufacturers' instructions.

Sequence determination and analysis

The complete nucleotide sequences of DNA 1 clones were determined by dideoxynucleotide chain termination sequencing using the PCR-based BIG DYE kit (Perkin-Elmer Cetus) and specific internal primers (Genosys). Reaction products were resolved on an ABI 3700 automated sequencer. Sequence information was stored, assembled, and analyzed using version 10.1 of the program library of the Genetics Computer Group (Devereaux et al., 1984).

Sequence alignments and phylogenetic analyses were generated using CLUSTAL X (Thompson et al., 1997) running on a Unix mainframe computer. Phylogenetic dendrograms were viewed and manipulated using Treeview (Page, 1996).

Infectivity analysis

The biological integrity of cloned DNA 1 components was assessed by co-inoculation with clones of *African cassava mosaic virus*-Kenya (ACMV-[Ke]; (Stanley and Gay, 1983)) to *N. benthamiana* using a hand-held biolistics device. The production of a partial repeat of the DNA A component and a dimer of the DNA B component of ACMV-[Ke] was described previously by Klinkenberg et al. (1989). The excised DNA 1 components were coated onto gold particles (0.93 μm) with the undigested ACMV-[Ke] components in equal proportions. Plants were inoculated as described by Briddon et al. (1998) and maintained in insect-proof glasshouses at 25 °C with supplementary lighting to yield a 16-h photoperiod. Plants were checked daily for the appearance of symptoms. The presence of DNA 1 components in ACMV-infected plants was assessed

by Southern blot analysis using a heterologous DNA 1 probe labelled with α^{32}P-dCTP.

Acknowledgments

The authors thank Dr. J. Stanley for providing the *Eupatorium* isolate. This work was supported by the Common Fund for Commodities (overseen by the International Cotton Advisory Committee under project no. ICAC/07) and the Department for Environment, Food and Rural Affairs (DEFRA) under project number PH01 47T. The CFC/ICAC project was a collaboration between the University of Arizona (Tucson, USA), the National Institute of Biotechnology and Genetic Engineering (Faisalabad, Pakistan), the Cotton Research Institute (Faisalabad, Pakistan), and the John Innes Centre. Their inputs to this work are gratefully acknowledged. The authors also acknowledge the support of the Biotechnology and Biological Sciences Research Council. Viruses were held and manipulated with the authorization of DEFRA under license number PHL 11D/3825(7/2001).

References

Boevink, P., Chu, P.W.G., Keese, P., 1995. Sequence of subterranean clover stunt virus DNA: affinities with the geminiviruses. Virology 207, 354–361.

Briddon, R.W., Markham, P.G., 2000. Cotton leaf curl virus disease. Virus Res. 71, 151–159.

Briddon, R.W., Liu, S., Pinner, M.S., Markham, P.G., 1998. Infectivity of African cassava mosaic virus clones to cassava by biolistic inoculation. Arch. Virol. 143, 2487–2492.

Briddon, R.W., Mansoor, S., Bedford, I.D., Pinner, M.S., Markham, P.G., 2000. Clones of cotton leaf curl geminivirus induce symptoms atypical of cotton leaf curl disease. Virus Genes 20, 17–24.

Briddon, R.W., Mansoor, S., Bedford, I.D., Pinner, M.S., Saunders, K., Stanley, J., Zafar, Y., Malik, K.A., Markham, P.G., 2001. Identification of DNA components required for induction of cotton leaf curl disease. Virology 285, 234–243.

Briddon, R.W., Bull, S.E., Amin, I., Idris, A.M., Mansoor, S., Bedford I.D., Dhawan, P., Rishi, N., Siwatch, S.S., Abdel-Salam, A.M., Brown, J.K., Zafar, Y., Markham, P.G., 2003. Diversity of DNA β: a satellite molecule associated with some monopartite begomoviruses. Virology 312, 106–121.

Bull, S.E., Briddon, R.W., Markham, P.G., 2003. Universal primers for the PCR-mediated amplification of DNA 1: a satellite-like molecule associated with begomovirus-DNA β complexes. Mol. Biotechnol. 23, 83–86.

Czosnek, H., Laterrot, H., 1997. A worldwide survey of tomato yellow leaf curl viruses. Arch. Virol. 142, 1391–1406.

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

Etessami, P., Watts, J., Stanley, J., 1989. Size reversion of African cassava mosaic virus coat protein gene deletion mutants during infection of *Nicotiana benthamiana*. J. Gen. Virol. 70, 277–289.

Fauquet, C.M., Bisaro, D.M., Briddon, R.W., Brown, J.K., Harrison, B.D., Rybicki, E.P., Stenger, D.C., Stanley, J., 2003. Revision of taxonomic criteria for species demarcation in the family *Geminiviridae*, and an updated list of begomovirus species. Arch. Virol. 148, 405–421.

Frischmuth, T., Stanley, J., 1991. African cassava mosaic virus DI DNA interferes with the replication of both genomic components. Virology 183, 539–544.

Frischmuth, T., Ringel, M., Kocher, C., 2001. The size of encapsidated single-stranded DNA determines multiplicity of *African cassava mosaic virus* particles. J. Gen. Virol. 82, 673–676.

Hafner, G.J., Stafford, M.R., Wolter, L.C., Harding, R.M., Dale, J.L., 1997. Nicking and joining activity of banana bunchy top virus replication protein in vitro. J. Gen. Virol. 78, 1795–1799.

Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J., Gronenborn, B., 1995. Determination of the origin cleavage and joining domain of geminivirus Rep proteins. Nucleic Acids Res. 23, 910–916.

Horser, C.L., Harding, R.M., Dale, J.L., 2001. Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. J. Gen. Virol. 82, 459–464.

Ilyina, T.V., Koonin, E.V., 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. 20, 3279–3285.

Jose, J., Usha, R., 2003. Bhendi yellow vein mosaic disease in India is caused by association of a DNA β satellite with a begomovirus. Virology 305, 310–317.

Katul, L., Maiss, E., Vetten, H.F., 1995. Sequence analysis of a faba bean necrotic yellows virus DNA component containing a putative replicase gene. J. Gen. Virol. 76, 475–479.

Katul, l., Timchenko, T., Gronenborn, B., Vetten, H.J., 1998. Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome. J. Gen. Virol. 79, 3101–3109.

Klinkenberg, A., Ellwood, S., Stanley, J., 1989. Fate of African cassava mosaic virus coat protein deletion mutants after agroinoculation. J. Gen. Virol. 70, 1837–1844.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S.G., Schell, J., Gronenborn, B., 1995. In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. Proc. Natl. Acad. Sci. U.S.A. 92, 3879–3883.

Mansoor, S., Khan, S.H., Bashir, A., Saeed, M., Zafar, Y., Malik, K.A., Briddon, R.W., Stanley, J., Markham, P.G., 1999. Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. Virology 259, 190–199.

Mansoor, S., Amin, I., Hussain, M., Zafar, Y., Bull, S., Briddon, R.W., Markham, P.G., 2001. Association of a disease complex involving a begomovirus, DNA 1 and a distinct DNA beta with leaf curl disease of okra in Pakistan. Plant Dis. 85, 922.

Mansoor, S., Briddon, R.W., Zafar, Y., Stanley, J., 2003. Geminivirus disease complexes: an emerging threat. Trends Plant Sci. 8, 128–134.

Navot, N., Pichersky, E., Zeidan, M., Zamir, D., Czosnek, H., 1991. Tomato yellow leaf curl virus: a whitefly-transmitted geminivirus with a single genomic component. Virology 185, 151–161.

Page, R.D.M., 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.

Rojas, M.R., Noueiry, A.M., Lucas, W.J., Gilbertson, R.L., 1998. Bean dwarf mosaic geminivirus movement proteins recognize DNA in a form- and size-specific manner. Cell 95, 105–113.

Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T., Kojima, M., 1998. Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome. J. Gen. Virol. 79, 3111–3118.

Saunders, K., Stanley, J., 1999. A nanovirus-like component associated with yellow vein disease of *Ageratum conyzoides*: evidence for interfamilial recombination between plant DNA viruses. Virology 264, 142–152.

Saunders, K., Bedford, I.D., Briddon, R.W., Markham, P.G., Wong, S.M., Stanley, J., 2000. A novel virus complex causes Ageratum yellow vein disease. Proc. Natl. Acad. Sci. U.S.A. 97, 6890–6895.

Saunders, K., Bedford, I.D., Stanley, J., 2001. Pathogenicity of a natural recombinant associated with *Ageratum* yellow vein disease: implica-

tions for geminivirus evolution and disease aetiology. Virology 282, 38–47.

Saunders, K., Bedford, I.D., Stanley, J., 2002. Adaptation from whitefly to leafhopper transmission of an autonomously-replicating nanovirus-like DNA component associated with *Ageratum* yellow vein disease. J. Gen. Virol. 83, 909–915.

Saunders, K., Bedford, I.D., Yahara, T., Stanley, J., 2003. The earliest recorded plant virus disease. Nature 422, 831.

Stanley, J., Gay, M.G., 1983. Nucleotide sequence of cassava latent virus DNA. Nature 301, 260–262.

Stanley, J., Frischmuth, T., Ellwood, S., 1990. Defective viral DNA ameliorates symptoms of geminivirus infection in transgenic plants. Proc. Natl. Acad. Sci. U.S.A. 87, 6291–6295.

Stanley, J., Saunders, K., Pinner, M.S., Wong, S.M., 1997. Novel defective interfering DNAs associated with Ageratum yellow vein geminivirus infection of *Ageratum conyzoides*. Virology 239, 87–96.

Tan, P.H.N., Wong, S.M., Wu, M., Bedford, I.D., Saunders, K., Stanley, J., 1995. Genome organization of ageratum yellow vein virus, a monopartite whitefly-transmitted geminivirus isolated from a common weed. J. Gen. Virol. 76, 2915–2922.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal_X windows interface; flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Timchenko, T., Katul, L., Sano, Y., de Kouchkovsky, F., Vetten, H.J., Gronenborn, B., 2000. The master Rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. Virology 274, 189–195.

Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J., 1982. Distantly related sequences in the α- and β-subunits of the ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945–951.

Zhou, X., Xie, Y., Tao, X., Zhang, Z., Li, Z., Fauquet, C.M., 2003. Characterization of DNA β associated with begomoviruses in China and evidence for co-evolution with their cognate viral DNA-A. J. Gen. Virol. 84, 237–247.