Journal of General Virology (1995), **76**, 249–263. *Printed in Great Britain*

Classification and identification of geminiviruses using sequence comparisons

Malla Padidam, Roger N. Beachy and Claude M. Fauquet*

International Laboratory for Tropical Agricultural Biotechnology (ILTAB/ORSTOM-TSRI), Division of Plant Biology-MRC7, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA

The genomes and ORFs of 36 geminiviruses were compared to obtain phylogenetic trees and frequency distributions of all possible pairwise comparisons with an objective to classify geminiviruses. Such comparisons show that geminiviruses form two distinct clusters of leafhopper-transmitted viruses that infect monocots (subgroup I) and whitefly-transmitted viruses that infect dicots (subgroup III), irrespective of the part of the genome considered. Of the two leafhopper-transmitted viruses that infect dicots, tobacco yellow dwarf virus has a sequence most similar to subgroup I viruses, and that of beet curly top virus differed depending upon the ORF considered. The distributions of identities within subgroups are significantly different suggesting that the taxonomic status of a particular isolate within a subgroup can be quantified. All the recognized strains of any one virus have greater than 90% sequence identity. It was observed that the 200 nucleotide intercistronic regions of geminiviruses are more variable than the remainder of the genome. The amino acid sequences of the coat protein (CP) of subgroup III viruses are more conserved than the remainder of the genome. However, a short N-terminal region (60–70 amino acids) of the CP is more variable than the rest of the CP sequence and is a close representation of the genome. PCR primers based on conserved sequences can be used to clone and sequence the N-terminal sequences of the CP of the geminiviruses; this sequence is sufficient to classify a virus isolate. A possible taxonomic structure for geminiviruses is proposed after considering the sequence comparisons and biological properties.

Introduction

Geminiviruses are single-stranded DNA viruses with a distinctive geminate capsid structure. They are transmitted by whiteflies or leafhoppers, and can cause significant diseases in many crop plants (Davies & Stanley, 1989; Goodman, 1981; Lazarowitz, 1992; Stanley, 1985). Because of their economic importance and the relative ease with which the genomes can be cloned, many geminiviruses have been characterized in detail during the past decade.

Geminiviruses, according to the International Committee on Taxonomy of Viruses (ICTV) (Francki *et al.*, 1991) are subdivided into three subgroups based on the insect vector, host range and genome structure. Subgroup I includes viruses with monopartite genomes that are transmitted by leafhoppers to monocotyledonous plants; the type member of this group is maize streak virus (MSV). The viruses transmitted by leafhoppers to dicotyledonous plants are grouped into subgroup II and beet curly top virus (BCTV) is the type member. Viruses belonging to subgroup III have bipartite genomes (except some isolates of tomato yellow leaf curl virus) and are transmitted by whiteflies to dicotyledonous plants; bean golden mosaic virus (BGMV) is considered as the type member of this subgroup. It has been proposed, and accepted, by the ICTV that the geminivirus group would become the *Geminiviridae* family comprising three genera called 'Geminivirus subgroup I, II, and III' (Mayo & Martelli, 1993).

The complete nucleotide (nt) sequence has been determined for more than 36 virus species and strains of the *Geminiviridae* and additional viruses are being sequenced. In addition, nucleic acid hybridization and PCR techniques have been used to study the molecular variability of some of these viruses (Gilbertson *et al.*, 1991*a*; Hughes *et al.*, 1992; Polston *et al.*, 1989). With the increased number of isolates being studied, particularly at the nucleic acid sequence level, it appears that some of the viruses do not fit the current classification. Many viruses isolated from tomato in different countries bear the same name 'tomato yellow leaf curl virus

* Author for correspondence. Fax +1 619 554 6330. e-mail iltab@scripps.edu

Table 1. *Geminivirus sequences compared*

Geminivirus name; abbreviation	Reference	GenBank accession no.
Subgroup I		
Leafhopper-transmitted, infecting monocots:		
Chloris striate mosaic virus, Australian isolate; CSMV	Andersen *et al.* (1988)	M20021
Digitaria streak virus, Vanuatu isolate; DSV	Donson *et al.* (1987)	M23022
Maize streak virus, Kenyan isolate; MSV-(Ke)	Howell (1984)	X01089
Maize streak virus, Nigerian isolate; MSV-(Ni)	Mullineaux *et al.* (1984)	K02026, X01633
Maize streak virus, South African isolate; MSV-(Sa)	Lazarowitz (1988)	Y00514
Miscanthus streak virus, Japanese isolate; MiSV	Chatani *et al.* (1991)	D01030
Panicum streak virus, Kenyan isolate; PSV	Briddon *et al.* (1992)	X60168
Sugarcane streak virus, Natal isolate; SSV	Hughes (1991)	M82918
Wheat dwarf virus, Czechoslovakian isolate; WDV-(CJI)	Woolston *et al.* (1988)	X02869
Wheat dwarf virus, Swedish isolate; WDV-(Sw)	MacDowell *et al.* (1985)	X02869
Leafhopper-transmitted, infecting dicots:		
Tobacco yellow dwarf virus, Australian isolate; TYDV	Morris *et al.* (1992)	M81103
Subgroup II		
Leafhopper-transmitted, infecting dicots:		
Beet curly top virus, Californian isolate; BCTV	Stanley *et al.* (1986)	M24597, X04144
Subgroup III		
Whitefly-transmitted, infecting dicots, isolates from the New World:		
Abutilon mosaic virus, West Indian isolate; AbMV	Frischmuth *et al.* (1990)	X15983, X15984
Bead dwarf mosaic virus, Colombian isolate; BDMV	Hidayat *et al.* (1993)	M88179, M88180
Bean golden mosaic virus, Brazilian isolate; BGMV-(Bz)	Gilbertson *et al.* (1993)	M88686, M88687
Bean golden mosaic virus, Dominican isolate; BGMV-(Dr)	Faria *et al.* (1994)	L01635, L01636
Bean golden mosaic virus, Guatemalan isolate; BGMV-(Ga)	Faria *et al.* (1994)	M91604, M91605
Bean golden mosaic virus, Puerto Rican isolate; BGMV-(Pr)	Howarth *et al.* (1985)	M10070, M10080
Pepper huasteco virus, Mexican isolate; PHV	Torres-Pacheco *et al.* (1993)	X70418, X70419
Potato yellow mosaic virus, Venezuelan isolate; PYMV	Coutts *et al.* (1991)	D00940, D00941
Squash leaf curl virus, Californian isolate; SLCV	Lazarowitz & Lazdins (1991)	M38182, M38183
Tomato golden mosaic virus; TGMV	Hamilton *et al.* (1984)	K02029, K02030
Tomato mottle virus, Florida isolate; TMoV	Abouzid *et al.* (1992)	L14460, L14461
Whitefly-transmitted, infecting dicots, isolates from the Old World:		
African cassava mosaic virus, Kenyan isolate; ACMV-(Ke)	Stanley & Gay (1983)	J02057, J02058
African cassava mosaic virus, Nigerian isolate; ACMV(Ni)	Morris *et al.* (1990)	X17095, X17096
Indian cassava mosaic virus, Indian isolate; ICMV	Hong *et al.* (1993*b*)	Z24758, Z24759
Mungbean yellow mosaic virus, Thailand isolate; MYMV	Morinaga *et al.* (1993)	D14703, D14704
Tomato leaf curl virus, Australian isolate; ToLCV-(Au)	Dry *et al.* (1993)	S53251
Tomato leaf curl virus, Indian isolate 1; ToLCV-(In1)	M. Padidam, unpublished	U15015, U15017
Tomato leaf curl virus, Indian isolate 2; ToLCV-(In2)	M. Padidam, unpublished	U15016
Tomato yellow leaf curl virus, Egyptian isolate; TYLCV-(Eg)	N. Abdallah, pers. comm.	
Tomato yellow leaf curl virus, Israeli isolate; TYLCV-(Is)	Navot *et al.* (1991)	X15656
Tomato yellow leaf curl virus, Sardinian isolate; TYLCV-(Sr)	Kheyr-Pour *et al.* (1991)	X61153
Tomato yellow leaf curl virus, Sicilian isolate; TYLCV-(Si)	G. P. Accotto, pers. comm.	
Tomato yellow leaf curl virus, Thailand isolate 1; TYLCV-(Th1)	Rochester *et al.* (1994)	M59838, M59839
Tomato yellow leaf curl virus, Thailand isolate 2; TYLCV-(Th2)	S. Attathom, pers. comm.	

(TYLCV)' and one may interpret these as different strains of the same virus. However, some of the isolates of TYLCV have sequences more similar to those of geminiviruses infecting other hosts than viruses that infect tomato (Kheyr-Pour *et al.*, 1991; Hong *et al.*, 1993*a*; Rochester *et al.*, 1994). Furthermore, the TYLCV-Israel and -Sardinia isolates are atypical members of geminivirus subgroup III, as they apparently lack the B component of the genome (Kheyr-Pour *et al.*, 1991; Navot *et al.*, 1991). In addition, TYLCV-Thailand isolate has two components but does not require the B component for infection (Rochester *et al.*, 1990). A geminivirus recently isolated from a dicotyledonous plant, tobacco yellow dwarf virus (TYDV), is transmitted by leafhoppers and is thus a candidate for geminivirus subgroup II, yet its sequence is similar to members of subgroup I (Morris *et al.*, 1992). Therefore, there is a need for clarification of the current methods of geminivirus classification and for simple criteria for classification.

Comparisons among nucleic acid and protein sequences of viral origin, along with comparisons among structural and biological criteria have long been used to identify and classify plant viruses (Shukla & Ward, 1988, 1989). In the present study, we compared the genomes of 36 geminiviruses to determine if sequence comparisons can be used to classify geminiviruses. Complete nt sequences of the genome components, nt sequence of the

intercistronic regions (ICR), and nt and amino acid (aa) sequences of the individual ORFs were aligned to obtain all possible pairwise percentage identities and phylogenetic trees. Percentage identities between all possible pairs were then plotted as frequency distributions to observe if distributions for subgroups are distinct. The analysis shows that it is possible to classify geminiviruses based on the sequence comparisons, and that a short region of the genome is sufficient to classify an isolate. Highly conserved PCR primers can be used to clone and sequence this short region to classify new virus isolates.

Methods

Sequences compared. The 36 geminivirus sequences compared are shown in Table 1. Sequence correction was made for BGMV-(Pr) A component, as suggested by Dr A. Howarth (personal communication). A guanine residue was inserted at nucleotide number 395 which extends the ORF for coat protein 60 nt in the upstream direction. An adenine residue in position 360 of AbMV-A component was changed to guanine after comparing with other New World viruses, extending the coat protein ORF sequence 30 bases upstream. We feel that the sequence of MYMV (Morinaga *et al.*, 1993) may have errors as 'precoat' protein (V2) ORF terminates prematurely and AC2 ORF is absent.

Pairwise comparison and phylogenetic analysis. Sequences were aligned using the clustal method of aligning multiple sequences (MegAlign program) available with the DNASTAR package for the Apple Macintosh computer (version 1.02, DNASTAR Inc., Madison, Wis., USA). The Clustal algorithm of MegAlign makes no *a priori* assumption of relatedness. It preserves gaps that occur in earlier alignments through later stages. Each alignment stage employs two sequence alignment methods: Wilbur & Lipman (1983) for input sequences and Myers & Miller (1988) for ancestral consensus sequence. Percentage similarity between sequences (i) and (j) is given as $100 \times$ sum of matches/[length − gap residues (i) − gap residues (j)]. Different gap and gap length penalties were used initially to see the effect on alignment, but no significant differences were observed. Subsequently, a gap penalty of 10 and a gap length penalty of 10 were used throughout. A random sequence of equal length and composition was included in all alignments to show pairwise percentage identities that are not significantly different from random identity.

All possible pairwise percentage identities were plotted as frequency distribution to examine if distributions of percentage identities within subgroups are different. The differences between distributions of pairwise percentage identities was tested using the Mann–Whitney rank sum test.

Phylogenetic analyses were done by a cladistic parsimony method using the computer program PAUP version 3.1.1 developed by D. L. Swoford (distributed by the Illinois Natural History Survey, Champaign, Ill., USA). Optimum trees were obtained with the heuristic method with the tree-bisection–reconnection branch-swapping option. One hundred bootstrap replications were performed to place confidence estimates on groups contained in the most parsimonious tree.

Phylogenetic analyses were also done using the UPGMA distance matrix and neighbourhood-joining method available with the MegAlign program. In this program, a preliminary phylogeny is derived from the distance between pairs of input sequences and the application of the UPGMA algorithm guides the alignment of ancestral sequences (Sneath & Sokal, 1973). The final phylogeny is produced by applying the neighbourhood-joining method to the distance and alignment data (Saitou & Nei, 1987). The trees generated by both PAUP and MegAlign were nearly identical and the trees presented here are those generated using the PAUP program, except for the tree based on the complete nucleotide sequence of A components for which both PAUP and MegAlign trees are presented.

Results

Geminivirus subgroups I and II (leafhopper-transmitted geminiviruses, LTGs) have a single-component genome of about 2·7 kb, except for BCTV which has a genome of 2·99 kb. The organization of the genomes of subgroup I viruses comprises four cistrons while the genome organization of subgroup II viruses is similar to that of the A component of subgroup III viruses (whitefly-transmitted geminiviruses, WTGs) and comprises five or six cistrons.

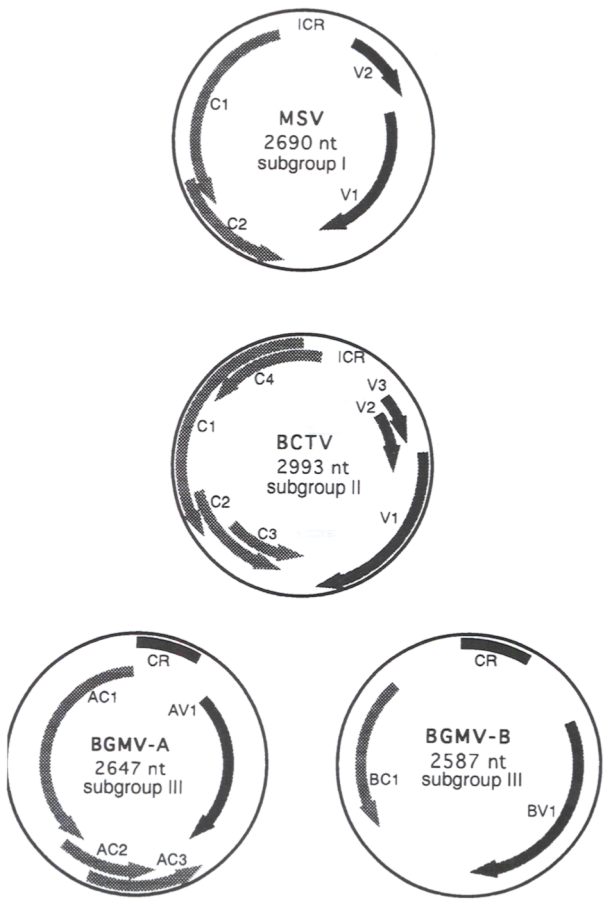

Fig. 1. Genome maps of type members of geminiviruses. ORF V1/AV1 codes for coat protein and ORF C1/AC1 product is essential for replication. ORFs BV1 and BC1 are required for virus movement. Other ORFs are involved in *trans*-activation, movement and regulating ssDNA vs dsDNA levels. ICR, intercistronic region; CR, common region which is identical in both A and B components of subgroup III viruses. B component has not been isolated for some of the subgroup III tomato yellow leaf curl isolates (see Lazarowitz, 1992, for details on genome structure).

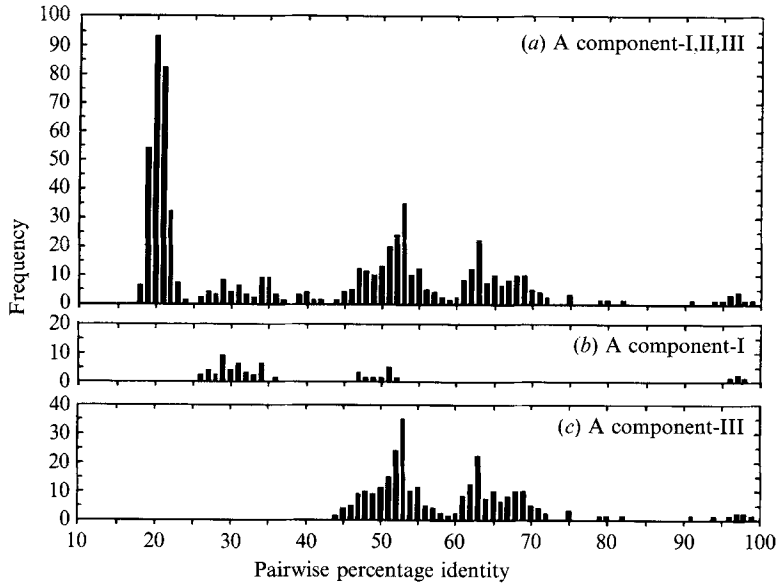

Fig. 2. (*a*) Distribution of the percentage identities of nucleotide sequences of the genomes of subgroup I and II geminiviruses and the A component of subgroup III geminiviruses. Frequency distributions for subgroup I and III geminiviruses are shown separately in (*b*) and (*c*), respectively. All possible pairwise percentage identities between all geminiviruses (630), subgroup I geminiviruses (50) and subgroup III geminiviruses (276) are plotted.

The B component of WTGs codes for two proteins. Genome maps for type members of the three subgroups are shown in Fig. 1. All the viruses belonging to subgroup III have a bipartite genome except TLCV-(Au), TYLCV-(Eg), TYLCV-(Is), TYLCV-(Si) and TYLCV-(Sr), for which a B component has not been isolated. The A and B components of bipartite viruses have no homology except for a 200 nt region that is almost identical (called common region, CR).

Four to eight ORFs have been identified in different viruses (Fig. 1). ORF AV1 codes for the coat protein (CP) while AC1 codes for a protein that is essential for replication (replicase). Other ORFs are involved in regulation of ssDNA vs dsDNA levels and in virus spread (see Lazarowitz, 1992, for a review on genome structure and gene function of geminiviruses).

Comparison of genomes

Initially, we compared the complete nt sequences of 36 geminiviruses to determine if they fall into distinct clusters. The published sequences were numbered in such a way that the first T of the sequence TAATATTAC, a nonanucleotide that is absolutely conserved in all the geminiviruses sequenced to date, is considered as nucleotide no. 1. Our analyses produced 630 possible alignments between the A components of WTGs plus LTGs and pairwise percentage identities (hereafter referred to as identities) are plotted as frequency distributions (Fig. 2*a*). A multimodal distribution with discontinuities is apparent. The identities between viruses in subgroup I and subgroup III ranged from 18–23%, as compared to 18–21% between a random sequence and the viral sequences, suggesting lack of significant homology between subgroup I and III viruses. The relationship among the 36 viruses is shown as phylogenetic trees in Fig. 3. The topologies of trees generated using the cladistic parsimony method of the PAUP program (Fig. 3*a*) and the UPMGA distance matrix method of the MegAlign program (Fig. 3*b*) were identical, suggesting the robustness of grouping. Also, in the bootstrap replication analysis of the most parsimonious trees generated with PAUP program, viruses within a branch occurred more than 50% of the time (Fig. 3*a*). The trees have distinct branches for subgroup I and subgroup III viruses. The subgroup III viruses formed clusters in the tree according to their geographical origin of isolation with distinct branches for New World and Old World viruses.

The distribution of identities for subgroups I and III are shown separately in Fig. 2(*b*) and 2(*c*). The identities among subgroup I viruses ranged from 26–98% (mean 39·3%, SD 18·4) compared to 44–99% among subgroup III viruses (mean 58·6%, SD 10·1), and the distributions are significantly different ($P < 0{\cdot}05$). Among the subgroup III viruses identities within New World viruses ranged from 61–97% (mean 69·1%, SD 8·2) and within Old World viruses ranged from 53–99% (mean 65·0%, SD 8·9). The identities between New World and Old World viruses ranged from 44–56% (mean 51·1%, SD 2·8). The distribution of identities between New World and Old World viruses is significantly different from the distribution of identities within New World or Old World viruses ($P < 0{\cdot}05$).

Sequence identities between BCTV (subgroup II) and viruses in subgroup III ranged from 35–41% as compared to 20–22% with subgroup I viruses. This is not unexpected because the genome organization of

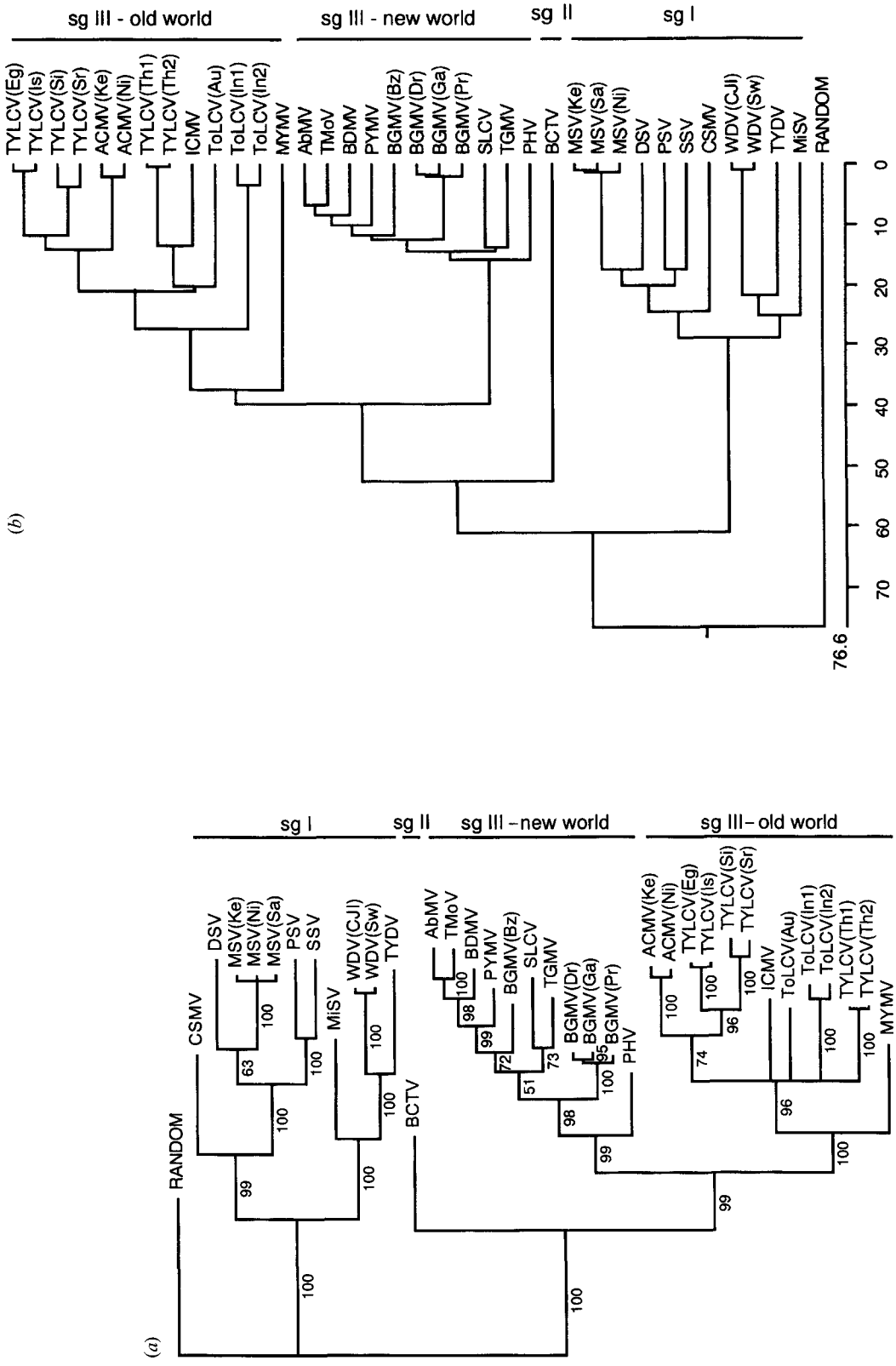

Fig. 3. (*a*) Phylogenetic tree obtained from the alignment of nucleotide sequences of 36 geminiviruses using the PAUP program. Analysis resulted in three equally parsimonious trees and the tree shown is the bootstrap 50% majority-rule consensus tree. The numbers below the branches refer to number of times (in percentages) in which the given branch is supported. Vertical distances are arbitrary, and horizontal distances reflect number of nucleotide differences between branch nodes. Total sequence of subgroup (sg) I and II geminiviruses and A component sequence of subgroup III geminiviruses were used for comparison. A random sequence of ACMV was included in the analysis. Note that the tree is unrooted and the random sequence was not used as an outgroup. (*b*) Phylogenetic tree obtained from the alignment of nucleotide sequences of 36 geminiviruses using the UPGMA distance matrix and neighbourhood-joining method of the MegAlign program. The scale below the tree measures the distance between sequences.

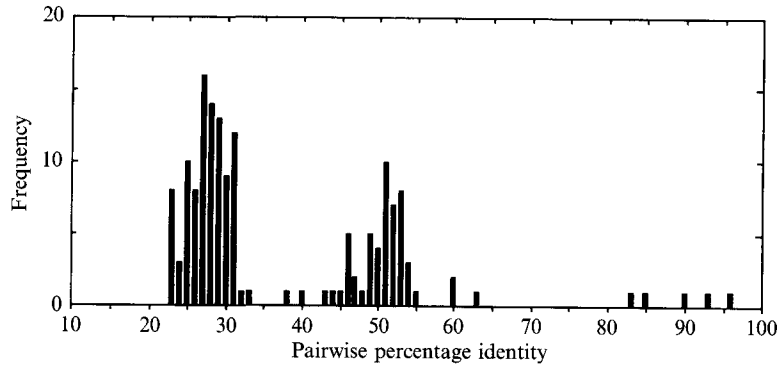

Fig. 4. Distribution of all the possible pairwise percentage identities (153) among B component sequences of subgroup III geminiviruses.

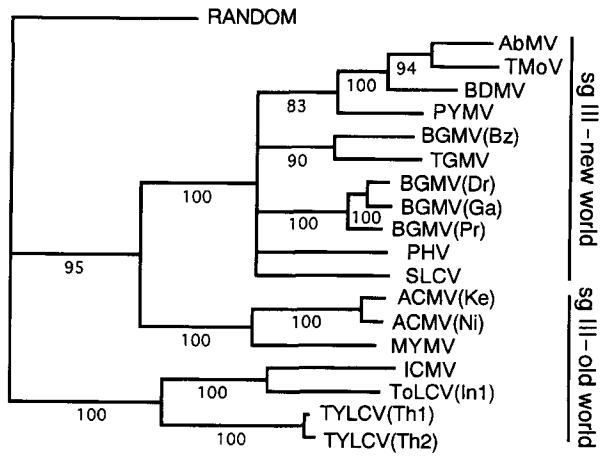

Fig. 5. Phylogenetic tree obtained from the alignment of B component nucleotide sequences of subgroup (sg) III geminiviruses using the PAUP program. Analysis resulted in only one most parsimonious tree and the tree shown is the bootstrap 50% majority-rule consensus tree. The numbers below the branches refer to number of times (in percentages) in which the given branch is supported. Vertical distances are arbitrary, and horizontal distances reflect number of nucleotide differences between branch nodes. Note that the tree is unrooted and the random sequence was not used as an outgroup.

BCTV is similar to subgroup III viruses, while only the CP sequence is similar to subgroup I viruses (Stanley *et al.*, 1986). The identity between TYDV and BCTV is 22%, and is not different from random identities. TYDV (subgroup I) is more similar to WDV (34% identity) than to other viruses (18–24% identity).

A high degree of identity (91–99%) was observed between MSV-(Ke), (Ni) and (Sa) isolates; WDV-(CJI) and (Sw) isolates; BGMV-(Dr), (Ga) and (Pr) isolates; ACMV-(Ke) and (Ni) isolates; ToLCV-(In1) and (In2) isolates; TYLCV-(Eg) and (Is) isolates; TYLCV-(Si) and (Sr) isolates; and TYLCV-(Th1) and (Th2) isolates (Figs 2 and 3). When these identities are considered as a separate distribution (mean 96·2%, SD 2·1) it is significantly different from the rest of the distribution (P 0·01). Because of such high identities and other biological properties they can be considered as strains of the same species. When these high identities are excluded from the analysis, the distributions of identities for subgroup I (mean 34·8%, SD 8·7) and subgroup III (mean 57·5%, SD 7·9), and the distributions of identities between New World and Old World viruses (mean 51·1%, SD 2·8) and within New World (mean 67·6%, SD 5·0) or Old World viruses (mean 62·9%, SD 4·0) are significantly different ($P < 0{\cdot}05$).

When the B components of subgroup III viruses are compared (153 possible pairings) a frequency distribution (Fig. 4) and the corresponding phylogenetic tree (Fig. 5) were developed. The percentage identities within New World or Old World viruses ranged from 27–96% as compared to 23–31% between New World and Old World viruses. Sequence identities between 83–96% were observed between BGMV-(Ga), (Dr) and (Pr) isolates; ACMV-(Ke) and (Ni) isolates; and TYLCV-(Th1) and (Th2) isolates. Phylogenetic clustering shows different branches for New World and Old World viruses (Fig. 5). However, Old World viruses ACMV and MYMV branched along with the New World viruses.

Comparison of intercistronic regions

The intercistronic region (ICR), also called CR for subgroup III viruses, contains promoter elements and a conserved nonanucleotide (TAATATTAC) which is part of a conserved hairpin loop assumed to be involved in DNA replication (Heyraud *et al.*, 1993). Approximately 200 nt of ICR sequences (from the start codon of AC1 to the end of the hairpin loop) were compared and are shown in Fig. 6. The distribution is somewhat different from the observed distribution of sequence comparisons for the entire A and B genomes (Figs 2 and 4) and appears somewhat bimodal. The identities within subgroup I viruses ranged from 21–98% (mean 34·6%, SD 17·7); within New World viruses identities ranged from 41–99% (mean 52·0%, SD 12·7), and within Old World viruses identities identities ranged from 32–100%

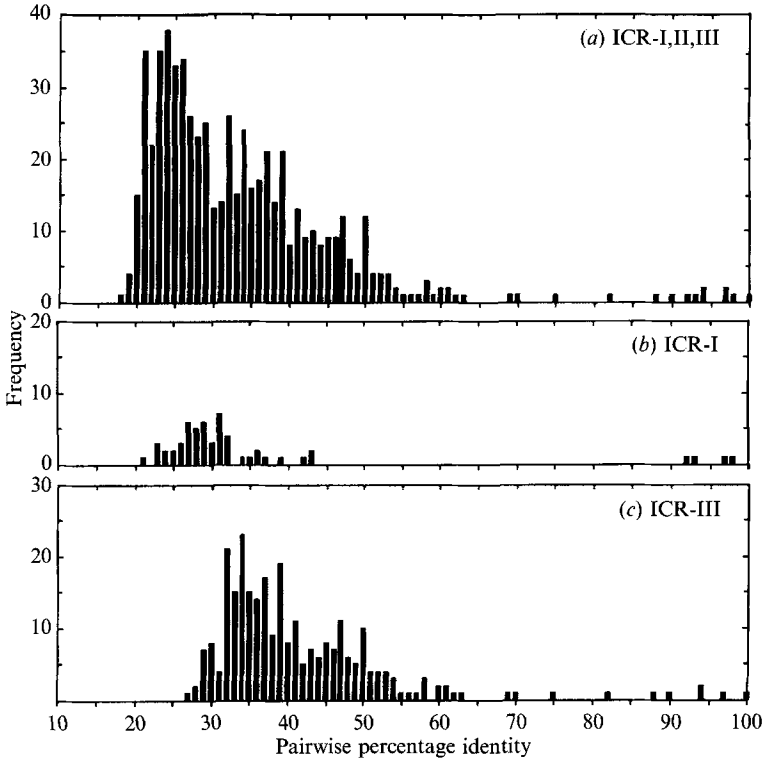

Fig. 6. (*a*) Distribution of the percentage identities among intercistronic regions (ICR) of 36 geminiviruses. Frequency distributions for subgroup I and III geminiviruses are shown separately in (*b*) and (*c*), respectively. All possible pairwise percentage identities between all geminiviruses (630), subgroup I geminiviruses (50) and subgroup III geminiviruses (276) are plotted.

(mean 47·0%, SD 12·4). Viruses whose A component sequences are more than 91% identical showed more than 82% identity in their ICRs.

Fig. 7 shows the phylogenetic tree based on the alignment of ICRs, and presents the same general pattern as the tree for the complete nt sequence of the A components (Fig. 3), i.e. three major branches for LTGs (except BCTV), WTGs from the New World and WTGs from the Old World. However, WTGs ACMV and ICMV formed a separate branch.

Comparison of coat protein sequences

The distributions of sequence identities for CP (ORF AV1) nt and aa sequences are shown in Fig. 8. The distribution of identities for nt sequences (Fig. 8*a*, *b*) is similar to the one obtained with complete A component sequences (Fig. 2*b*, *c*). However, the distribution for aa identities for subgroup III viruses (WTGs) is continuous from 62–98% (Fig. 8*d*). Interestingly, the distribution is discontinuous for subgroup I and II viruses (LTGs) for aa sequences and ranged from 17–99% (Fig. 8*c*). The identities between WTGs and LTGs ranged from 18–25% (random 18–21%) for nt sequences and 10–16% (random 10–13%) for aa sequences suggesting lack of relationship between CPs and LTGs and WTGs.

The phylogenetic trees obtained after aligning the CP nt or aa sequences (Fig. 9*a*, *b*) showed a branching pattern similar to the tree obtained from the complete A component sequences (Fig. 3), with the exception of BCTV which is placed with the subgroup I branch (Fig. 9*a*).

While aligning the CP sequence we observed that the N-terminal 60–70 aa are more variable than the remainder of the sequence for subgroup III viruses. This region precedes a stretch of 10 invariant aa (ACMV G68–S77). For subgroup I and II viruses, the variability is distributed over the entire CP sequence. However, there is a stretch of five conserved aa around the N-terminal 110 region (MSV W110–D114). The sequences 5′ of these conserved aa were compared and the distributions are shown in Fig. 10. The distribution of CP 5′ nt sequence identities for subgroups I, II and III (Fig. 10*a*, *b*) and the CP N-terminal aa identities for subgroups I and II (Fig. 10*c*) are similar to the complete A component sequence identity distribution (Fig. 2*b*, *c*), with a discontinuity around 90% of identity. The discontinuity around 90% identity is not as clear cut for the N-terminal aa sequence of subgroup III (Fig. 10*d*) compared with the 5′ nt sequence (Fig. 10*b*). However, it is a closer representation of the complete A component sequence distribution (Fig. 2*c*) than the complete aa sequence of the CP (Fig. 8*d*). The branching pattern of phylogenetic trees for 5′ nt and N-terminal aa CP sequences is similar to the tree based on the complete A component sequences (data not shown).

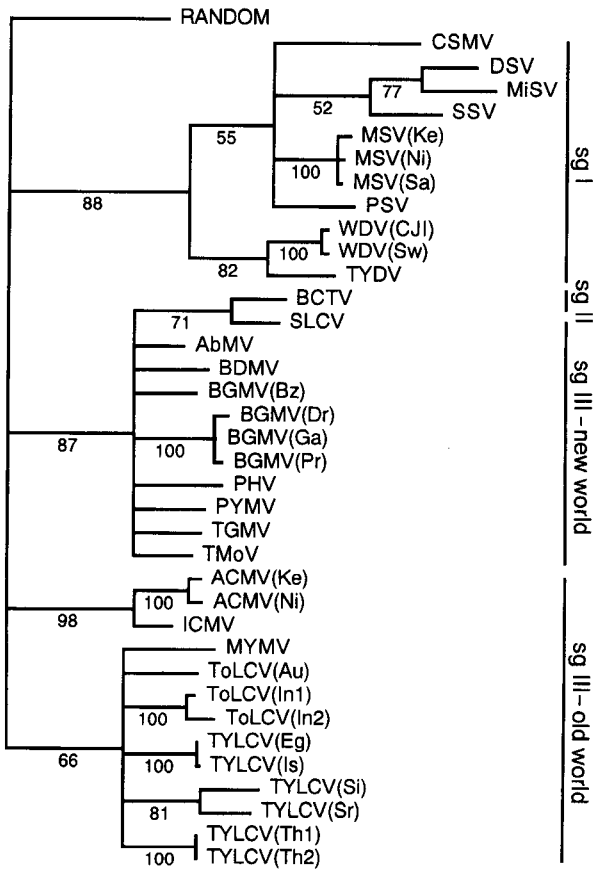

Fig. 7. Phylogenetic tree obtained from the alignment of intercistronic regions (ICR) of 36 subgroup (sg) I, II and III geminiviruses using the PAUP program. Analysis resulted in two equally parsimonious trees and the tree shown is the bootstrap 50% majority-rule consensus tree. The numbers below the branches refer to number of times (in percentages) in which the given branch is supported. Vertical distances are arbitrary, and horizontal distances reflect number of nucleotide differences between branch nodes. Note that the tree is unrooted and the random sequence was not used as an outgroup.

Comparison of replicase sequences

The distributions of identities of nt and aa sequences of the AC1 protein were similar to the ones based on complete A component sequences (Fig. 2); subgroup I virus sequences are more variable than subgroup III viruses and there is a discontinuity between 84–94% identities for both nt and aa sequences (data not shown). Unlike CP, both nt and aa identity distributions were similar.

Phylogenetic trees derived from the alignment of replicase nt and aa sequences (Fig. 11*a*, *b*) are similar to the general pattern with three main branches. Among the New World viruses, the PHV replicase is different from the other New World viruses, occupying a position between New World and Old World viruses. The position of BCTV, in both nt and aa sequence trees, is included with the subgroup III New World viruses (Fig. 11*a*, *b*), which is understandable as the BCTV has a genome that, with the exception of the CP sequence, is similar to subgroup III viruses. BCTV was originally isolated in the New World.

Unlike the situation with CP of subgroup III viruses, variable aa in the replicase are not restricted to a particular region. We wanted to see if a short N-terminal sequence could be identified as being representative of identities in the replicases. The sequence 5′ of a stretch of six conserved aa (MSV F84–Q89, ACMV F86–Q91) was compared for all subgroups. The identities for 5′ nt and N-terminal aa sequences of the replicase were not representative of the identities in the replicase (data not shown).

Comparison of other ORFs

We also compared nt and aa sequences of ORFs AV2, AC2, AC3, AC4, BV1 and BC1. The frequency distributions for identities and phylogenetic trees were similar to that obtained with the total sequence and the clustering of viruses was similar in the derived trees (data not shown).

Discussion

In this study we compared the genomes and ORFs of 36 geminiviruses. Alignment of nt sequences (Figs 3 and 5) and plotting of pairwise percentage identities (Figs 2 and 4) show that the geminiviruses form two distinct clusters of whitefly-transmitted, dicot-infecting viruses and leafhopper-transmitted, monocot infecting viruses. Within the whitefly-transmitted viruses, New World and Old World viruses always form separate groups (Figs 3 and 5). This confirms the earlier phylogenetic study of geminiviruses by Howarth & Vandemark (1989) and extends the results to 22 other isolates characterized since the early report. Of the two leafhopper-transmitted, dicot-infecting viruses, the TYDV sequence is more similar to leafhopper-transmitted, monocot-infecting viruses than to other viruses (Figs 3, 7, 9 and 11; Morris *et al.*, 1992). BCTV is different from TYDV as it has a hybrid genome; the CP is similar to LTGs (Fig. 9) and the other ORFs are similar to WTGs from the New World (Figs 3 and 12; Stanley *et al.*, 1986).

Comparison of intercistronic regions and nt and aa sequences of all ORFs (Figs 6–12 and data not shown) show the same general pattern of virus clustering as when sequences of the entire genome were compared (Figs 2–5). Although the alignment of ICRs resulted in a phylogenetic tree (Fig. 7) that is similar to the tree based on complete sequences, there is less sequence identity in

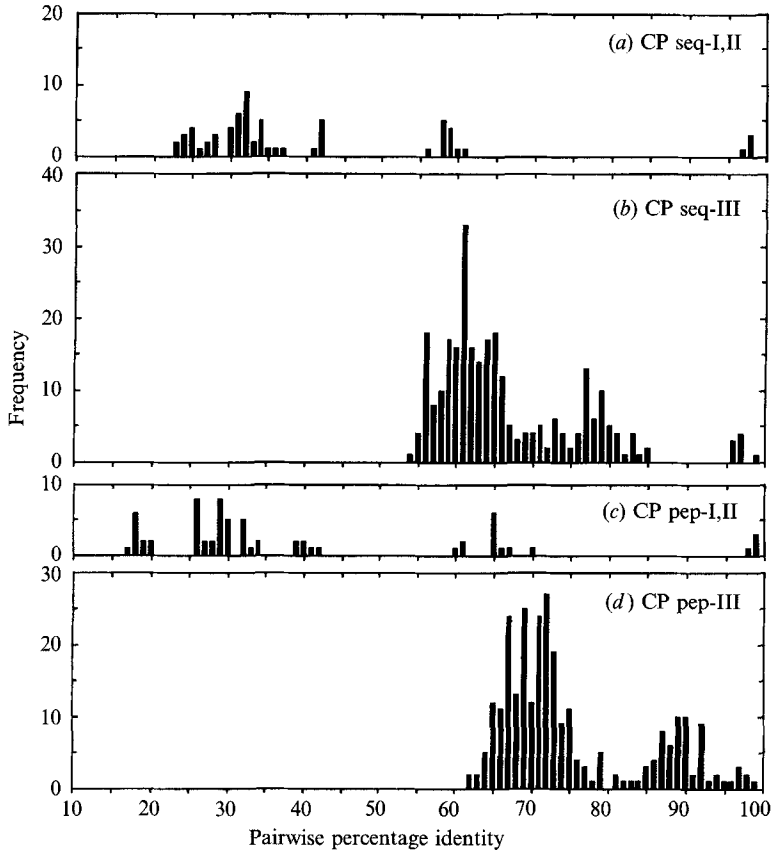

Fig. 8. Distribution of the percentage identities for the coat protein (AV1) nucleotide (*a*, *b*) and amino acid sequences (*c*, *d*) for subgroups I and II (*a*, *c*) and subgroup III (*b*, *d*) geminiviruses. All possible pairwise percentage identities between subgroups I and II geminiviruses (66) and subgroup III geminiviruses (276) are plotted.

this region than in other regions. The CP aa sequence is more conserved than the remainder of the genome for WTGs (Figs 2*c*, 8*b*, *d*). This may reflect the fact that allowed mutations in CP are under the constraints of viral structure, vector transmission, host specificity and other unknown functions. However, the CP sequence of LTGs are as variable as the rest of the genome (Figs 2*a*, 8*a*, *c*) and therefore we should conclude that the constraints are less important or less numerous for these viruses.

When all possible pairwise percentage identities among 36 geminivirus A component sequences are plotted as a distribution (Fig. 2), the data clearly showed a discontinuity of distribution around the level of 90% identity. The identities within subgroup I, New World or Old World viruses were significantly different than between the viruses of these groups. This suggests that the position of a particular virus within a subgroup can be quantified using all possible pairwise percentage identities. The discontinuity was also observed when different regions of the genome were compared (Figs 4, 6, 8 and 10; data not shown) showing that it reflects the entire genome rather than a particular region. The only exception was the CP aa sequence of WTGs. The isolates that have greater than 90% identities are recognized strains of the same virus species.

With the availability of the PCR technique, it is relatively easy to study sequence variations compared to biological properties. We wanted to see if a short region of the genome representative of variability within the entire virus can be cloned by PCR. The N-terminal sequence of CP fulfills this objective (Fig. 10). Degenerate primers based on the conserved aa immediately downstream of the N-terminal sequence and on the conserved nonanucleotide in the intercistronic region can be used to clone and sequence this region. Primers based on conserved sequences in CP, replicase and ICR have been previously used to clone and study variability in geminiviruses (Gilbertson *et al.*, 1991*b*; Rojas *et al.*, 1993; Rybicki & Hughes, 1990). We show that a shorter N-terminal sequence of CP is an informative as the entire sequence of the genome. Larger fragments of the entire genome can be sequenced for isolates that have less than 90% identity to previously sequenced isolates and/or have interesting biological properties.

Historically, the subdivision of geminiviruses has been proposed according to the genome composition and organization, the host range and vector transmission. The situation is made more complicated by the fact that TYDV, a member of the geminivirus subgroup I by its genome organization, infects a dicotyledonous plant, and that several viruses belonging to the geminivirus

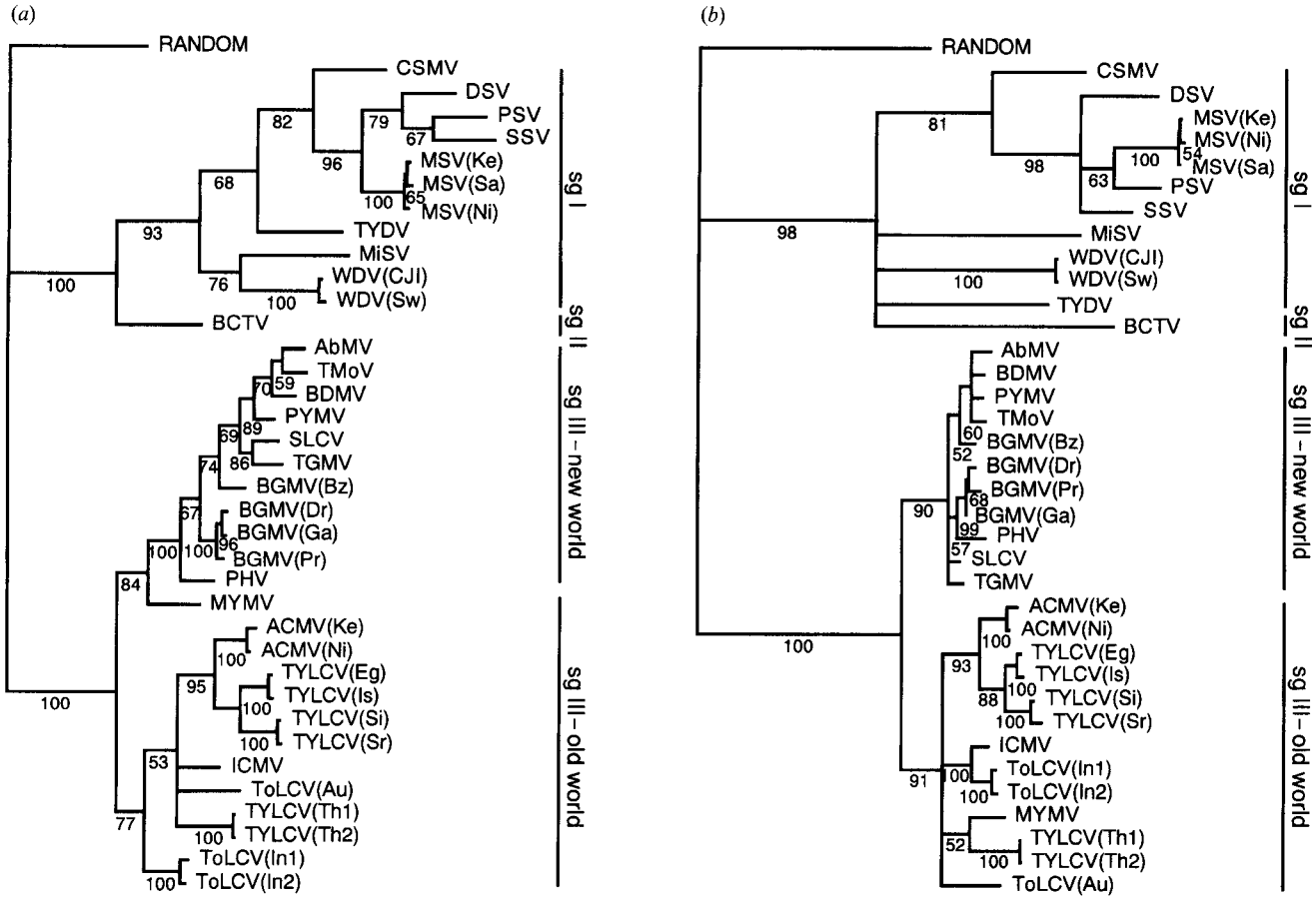

Fig. 9. Phylogenetic tree obtained from the alignment of the coat protein (AV1) nucleotide (*a*) and amino acid sequences (*b*) of 36 subgroup (sg) I, II and III geminiviruses using the PAUP program. Analysis resulted in only one most parsimonious tree for nucleotide sequence and six equally parsimonious trees for amino acid sequence. The trees shown are the bootstrap 50% majority-rule consensus tree. The numbers below the branches refer to number of times (in percentages) in which the given branch is supported. Vertical distances are arbitrary, and horizontal distances reflect number of nucleotide differences between branch nodes. Note that the tree is unrooted and the random sequence was not used as an outgroup.

subgroup III have a single component. Consequently, there are those who propose creating a geminivirus subgroup IV to include the monopartite WTGs and others who argue that TYDV should be included in the geminivirus subgroup II because it infects dicots.

The present study set out to look at geminivirus sequences and to consider the biological and molecular properties of these viruses when proposing a taxonomic structure of the *Geminiviridae*. Considering the results that were obtained in this study we propose that any new virus isolate having more than 90% sequence identity to a previously characterized virus genome should be called a strain of an already described virus species. By this criteria we suggest that the BGMV-(Bz) isolate should be considered as a separate species from BGMV-(Dr), BGMV-(Ga) and BGMV-(Pr) isolates. BGMV-(Bz) is more closely related to other New World viruses than to BGMV-(Dr), BGMV-(Ga) and BGMV-(Pr) (Figs 3, 5, 7, 9 and 11). BGMV-(Bz) also differs from other BGMV isolates in sap transmissibility and pathogenicity on different bean cultivars (Gilbertson *et al.*, 1993). Among the Old World viruses that infect tomatoes, ToLCV-(Au), ToLCV-(In1), TYLCV-(Is), TYLCV-(Sa) and TYLCV-(Th1) isolates should be considered as separate species as the identities between the viruses are less than 90%, and ToLCV-(In2), TYLCV-(Eg), TYLCV-(Si) and TYLCV-(Th2) as strains of ToLCV-(In1), TYLCV-(Is), TYLCV-(Sr) and TYLCV-(Th1), respectively. Among the viruses infecting cassava, ACMV-(Ke) and (Ni) isolates are to be considered as strains of the same virus species and ICMV as a different species. An isolate infecting cassava in Malawi was recently characterized and considered to be distinct from ACMV and ICMV (Hong *et al.*, 1993*b*); the conclusions of these authors fit with our proposal. Recently, the CPs of 12 MSV isolates from different countries of Africa have been sequenced (R. W. Briddon, personal communication). We compared these sequences and found that all MSV isolates

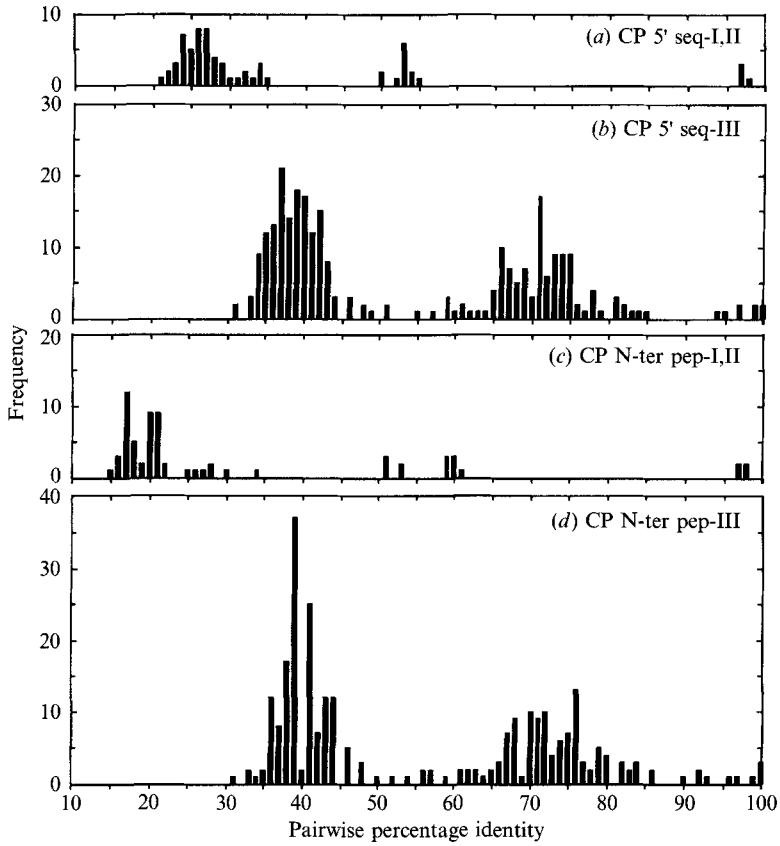

Fig. 10. Distribution of the identities for the coat protein (AV1) N-terminal nucleotide (*a*, *b*) and amino acid sequences (*c*, *d*) for subgroup I and II geminiviruses (*a*, *c*) and subgroup III (*b*, *d*) geminiviruses. For subgroup I and II geminiviruses the N-terminal ~ 110 amino acids (~ 330 nucleotides) before a stretch of five conserved amino acids and for subgroup III geminiviruses ~ 66 amino acids (~ 198 nucleotides) before a stretch of 10 invariable amino acids, were compared. All possible pairwise percentage identities between subgroup I and II geminiviruses (66) and subgroup III geminiviruses (276) are plotted.

have greater than 95% identity (data not shown), again demonstrating that strains of the same virus species have greater than 90% identity. A similar quantification of the taxonomic level of virus strain has been proposed to classify potyviruses (Shukla & Ward, 1988).

From this study it is also clear that there are three well defined virus clusters: cluster A-LTGs that infect monocots and dicots, with one component and 4 cistrons; cluster B-WTGs infecting dicots in the New World, having two components and 5(6)+2 cistrons; and cluster C-WTGs infecting dicots in the Old World, having one or two components and 5(6) or 5(6)+2 cistrons. It is apparent that clusters B and C share many properties and are closely related in many respects, but the geographical distributions reflected in each cistron indicate that these viruses have evolved over a long period of time, and constitute two well defined entities.

TYDV, if we consider all its cistrons, is a member of cluster A differing only from other members by its host range; thus, we must consider that this criteria may be of limited usefulness. We must also conclude that the host range criteria cannot be correlated with sequences.

The case of BCTV is more interesting, as its position in the phylogenetic tree differs according to the cistron considered. It is possible that this virus resulted from the recombination of a LTG and a WTG. Therefore, might it be useful to create a specific taxonomic level for a virus that has all but one of the properties of another cluster? The current definition of a genus is 'a group of species sharing common characters'; one may consider that the situation of BCTV is sufficiently unique for it to be considered as a separate genus.

Some of the WTGs of cluster C have a single component, but if we consider the other cistrons, they are closely related to members having two components. Furthermore, TYLCV-(Th) has two components but infection of tomato plants with only the A component can cause a severe disease; the presence of the B component serves as a symptom enhancer (Rochester *et al.*, 1990). On the contrary ToLCV-(In) requires the B component for infectivity (M. Padidam, unpublished); therefore, genome composition may not be an absolute taxonomic criteria.

Taking into account the sequence comparisons and the biological properties of geminiviruses a possible taxonomic structure of the *Geminiviridae* family could be suggested. We propose to establish the three clusters described above as genera. TYDV is included in cluster A. BCTV, clearly a recombinant of cluster B, could be a sole member of another genus. Subfamilies can be established to include the marked differences between clusters B and C and BCTV. The proposed classification

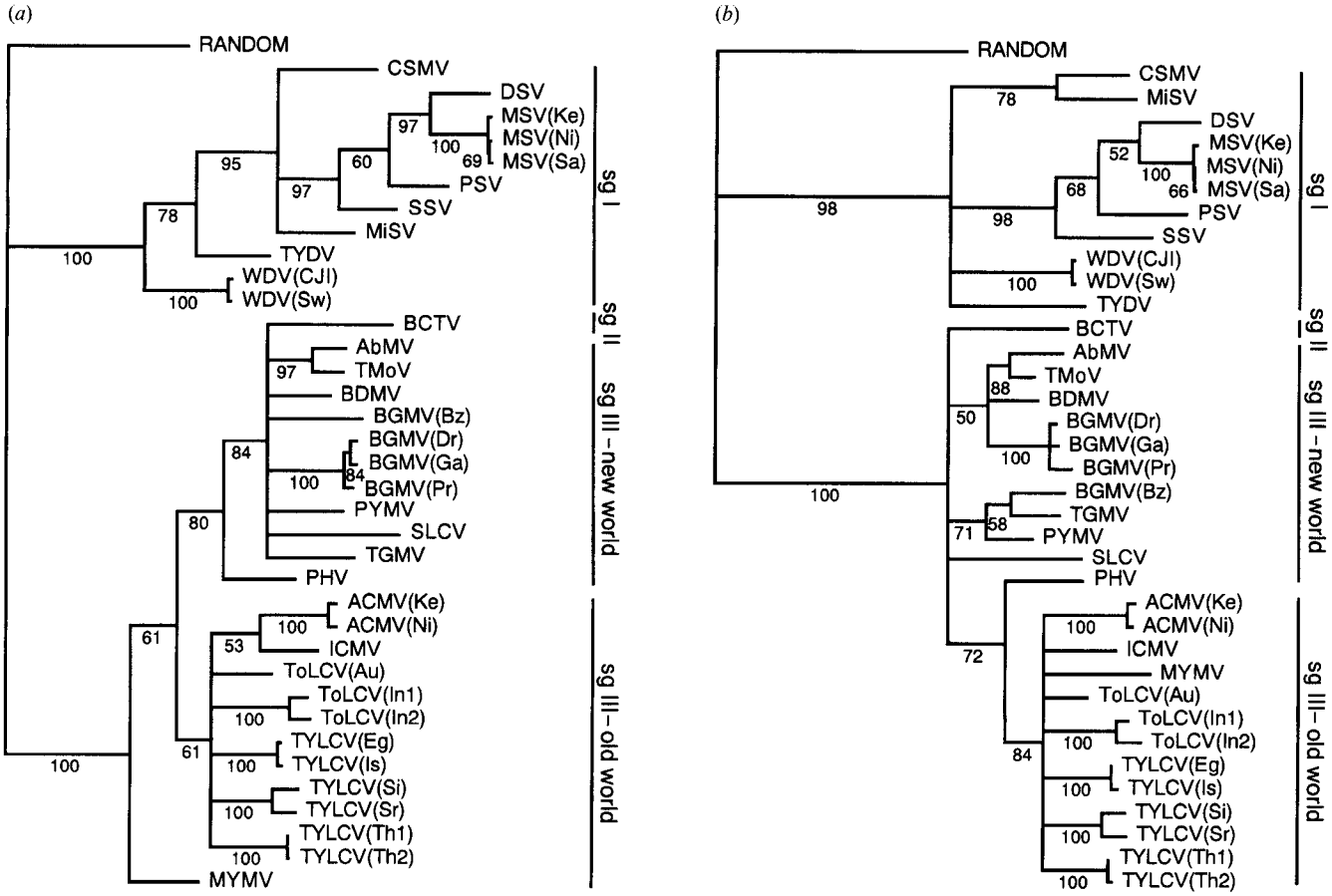

Fig. 11. Phylogenetic tree obtained from the alignment of the replicase (AC1) nucleotide (*a*) and amino acid sequences (*b*) of 36 subgroup (sg) I, II and III geminiviruses using the PAUP program. Analysis resulted in only one most parsimonious tree for nucleotide sequence and 20 equally parsimonious trees for amino acid sequence. The trees shown are the bootstrap 50% majority-rule consensus trees. The numbers below the branches refer to number of times (in percentages) in which the given branch is supported. Vertical distances are arbitrary, and horizontal distances reflect number of nucleotide differences between branch nodes. Note that the tree is unrooted and the random sequence was not used as an outgroup.

is shown schematically in Fig. 12. The proposal will be discussed with the members of the ICTV geminivirus study group prior to submission to the ICTV executive committee.

It is interesting to note that each genus can be identified by the frequency distribution of the pairwise percentage identities (Fig. 2). The distribution of identities between genera are significantly different from the distribution of identities within genera. Although the limits are not precise, there is definitely a concept of quantification of the taxonomic levels within the *Geminiviridae* family.

The results of sequence comparisons may have a bearing on the evolution of geminiviruses. Previous work (Howarth & Goodman, 1986) and our phylogenetic analyses show that subgroup I and subgroup III viruses are distant relatives descended from a common ancestral sequence. It is possible that this ancestral geminivirus was infecting monocots, had a single-component genome and was leafhopper-transmitted. This may explain the higher sequence variability among subgroup I viruses. TYDV, a member of subgroup I, later acquired the ability to infect dicots. Similar changes might have facilitated ancestral virus to acquire whitefly transmission and marked changes in ORFs to give rise to single component WTGs in the Old World. This then would suggest that viruses with two-component genomes evolved later on the evolutionary scale. The second component that is present in some of the WTGs from the Old World and in all the WTGs from the New World may have contributed to the enhanced movement of the virus within the plant and/or efficient transmission by whiteflies. BCTV appears to have evolved from a monopartite subgroup III virus by acquiring the CP sequence from a member of subgroup I.

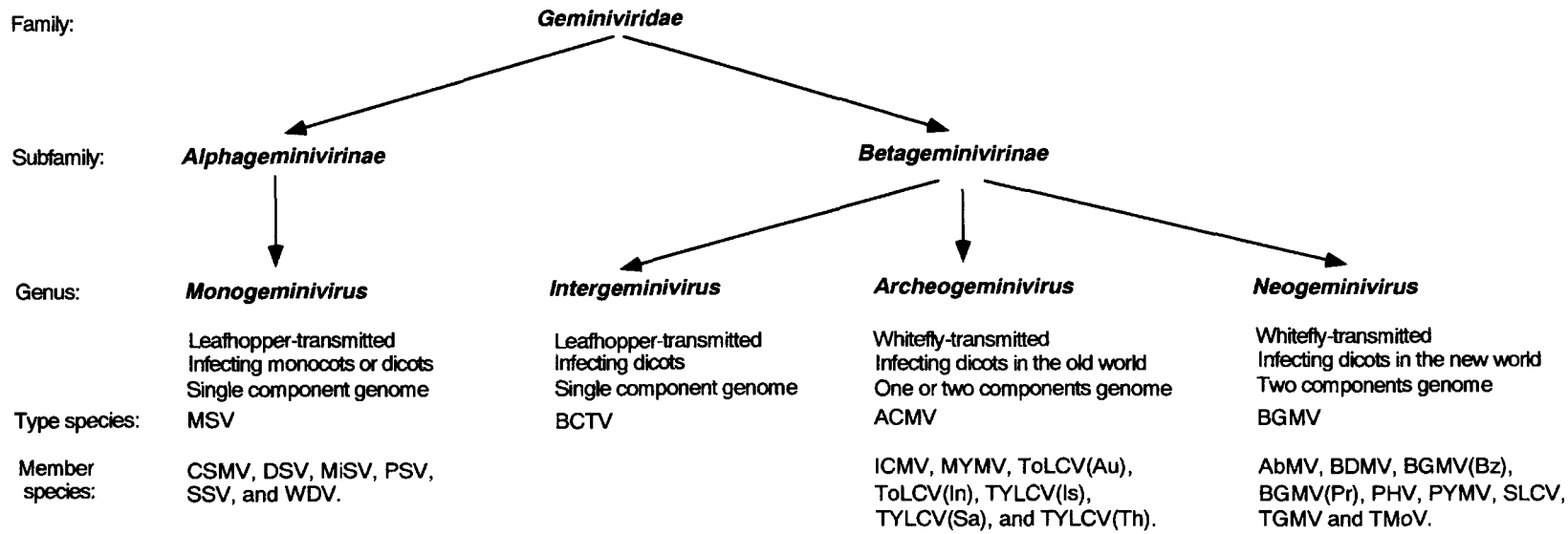

Fig. 12. Proposed taxonomic structure for the family *Geminiviridae*.

The authors would like to thank N. Abdallah, G. P. Accotto, S. Attathom, R. W. Briddon and D. P. Maxwell for providing sequence data before publication. We are also grateful to B. Sobral for help in phylogenetic analysis. This work was supported by USAID grant no. 3-5-97653 and ORSTOM.

References

ABOUZID, A. M., POLSTON, J. E. & HIEBERT, E. (1992). The nucleotide sequence of tomato mottle virus, a new geminivirus isolated from tomatoes in Florida. *Journal of General Virology* **73**, 3225–3229.

ANDERSEN, M. T., RICHARDSON, K. A., HARRISON, S. & MORRIS, B. A. M. (1988). Nucleotide sequence of the geminivirus Chloris striate mosaic virus. *Virology* **164**, 443–449.

BRIDDON, R. W., LUNESS, P., CHAMBERLIN, L. C. L., PINNER, M. S., BRUNDISH, H. & MARKHAM, P. G. (1992). The nucleotide sequence of an infectious insect-transmissible clone of the geminivirus *Panicum* streak virus. *Journal of General Virology* **73**, 1041–1047.

CHATANI, M., MATSUMOTO, Y., MIZUTA, H., IKEGAMI, M., BOULTON, M. I. & DAVIES, J. W. (1991). The nucleotide sequence and genome structure of the geminivirus miscanthus streak virus. *Journal of General Virology* **72**, 2325–2331.

COUTTS, R. H. A., COFFIN, R. S., ROBERTS, E. J. F. & HAMILTON, W. D. O. (1991). The nucleotide sequence of the infectious cloned DNA components of potato yellow mosaic virus. *Journal of General Virology* **72**, 1515–1520.

DAVIES, J. W. & STANLEY, J. (1989). Geminivirus genes and vectors. *Trends in Genetics* **5**, 77–81.

DONSON, J. K., ACCOTTO, G. P., BOULTON, M. I., MULLINEAUX, P. M. & DAVIES, J. W. (1987). The nucleotide sequence of a geminivirus from *Digitaria sanguinalis*. *Virology* **161**, 160–169.

DRY, I. B., RIGDEN, J. E., KRAKE, L. R., MULLINEAUX, P. M. & REZAIAN, M. A. (1993). Nucleotide sequence and genome organization of tomato leaf curl geminivirus. *Journal of General Virology* **74**, 147–151.

FARIA, J. C., GILBERTSON, R. L., HANSON, S. F., MORALES, F. J., AHLQUIST, P., LONIELLO, A. O. & MAXWELL, D. P. (1994). Bean golden mosaic geminivirus type II isolates from the Dominican Republic and Guatemala. *Phytopathology* **84**, 321–329.

FRANCKI, R. B. I., FAUQUET, C. M., KNUDSON, D. L. & BROWN, F. (1991). Classification and Nomenclature of Viruses. Fifth Report of the International Committee on Taxonomy of Viruses. *Archives of Virology*, Supplementum 2.

FRISCHMUTH, T., ZIMMAT, G. & JESKE, H. (1990). The nucleotide sequence of abutilon mosaic virus reveals prokaryotic as well as eukaryotic features. *Virology* **178**, 461–467.

GILBERTSON, R. L., HIDAYAT, S. H., MARTINEZ, R. T., LEONG, S. A., FARIA, J. C., MORALES, F. & MAXWELL, D. P. (1991*a*). Differentiation of bean-infecting geminiviruses by nucleic acid hybridization probes *and* aspects of bean golden mosaic in Brazil. *Plant Disease* **75**, 336–342.

GILBERTSON, R. L., ROJAS, M. R., RUSSEL, D. R. & MAXWELL, D. P. (1991*b*). Use of the asymmetric polymerase chain reaction and DNA sequencing to determine genetic variability of bean golden mosaic geminivirus in the Dominican Republic. *Journal of General Virology* **72**, 2843–2848.

GILBERTSON, R. L., FARIA, J. C., AHLQUIST, P. & MAXWELL, D. P. (1993). Genetic diversity in geminiviruses causing bean golden mosaic disease: the nucleotide sequence of the infectious cloned DNA components of a Brazilian isolate of bean golden mosaic geminivirus. *Phytopathology* **83**, 709–715.

GOODMAN, R. M. (1981). Geminiviruses. *Journal of General Virology* **54**, 9–21.

HAMILTON, W. D. O., STEIN, V., COUTTS, R. H. A. & BUCK, K. Q. W. (1984). Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: potential coding regions and regulatory sequences. *EMBO Journal* **3**, 2197–2205.

HEYRAUD, F., MATZEIT, V., KAMMAN, M., SCHAEFER, S., SCHELL, J. & GRONENBORN, B. (1993). Identification of the initiation sequence for viral strand DNA synthesis of wheat dwarf virus. *EMBO Journal* **12**, 4445–4452.

HIDAYAT, H. S., GILBERTSON, R. L., HANSON, S. F., MORALES, F. J., AHLQUIST, P., RUSSEL, D. R. & MAXWELL, D. P. (1993). Complete nucleotide sequences of the infectious cloned DNAs of bean dwarf mosaic geminivirus. *Phytopathology* **83**, 181–187.

HONG, Y. G., FARGETTE, D., SWANSON, M. M., MCGRATH, P. F. & HARRISON, B. D. (1993*a*). Sequence and epitope variation among coat proteins of tomato leaf curl geminiviruses from different regions. *IXth International Congress of Virology*, Glasgow, UK, abstract no. W68-2.

HONG, Y. G., ROBINSON, D. J. & HARRISON, B. D. (1993*b*). Nucleotide sequence evidence for the occurence of three distinct whitefly-transmitted viruses. *Journal of General Virology* **74**, 2437–2443.

HOWARTH, A. J. & GOODMAN, R. M. (1986). Divergence and evolution of geminivirus genomes. *Journal of Molecular Evolution* **23**, 313–319.

HOWARTH, A. J. & VANDEMARK, G. J. (1989). Phylogeny of geminiviruses. *Journal of General Virology* **70**, 2717–2727.

HOWARTH, A. J., CATON, J., BOSSERT, M. & GOODMAN, R. M. (1985). Nucleotide sequence of bean golden mosaic virus and a model for gene regulation in geminiviruses. *Proceedings of the National Academy of Sciences, USA* **82**, 3572–3576.

HOWELL, S. H. (1984). Physical structure and genetic organisation of the genome of maize streak virus (Kenyan isolate). *Nucleic Acids Research* **12**, 7359–7375.

HUGHES, F. L. (1991). *Molecular investigations of subgroup I geminiviruses*. PhD thesis, University of Cape Town, South Africa.

HUGHES, F. L., RYBICKI, E. P. & VON WECHMAN, M. B. (1992). Genome typing of southern African subgroup 1 geminiviruses. *Journal of General Virology* **73**, 1031–1040.

KHEYR-POUR, A., BENDAHMANE, M., MATZEIT, V., ACCOTTO, G. P., CRESPI, S. & GRONENBORN, B. (1991). Tomato yellow leaf curl virus from Sardinia is a whitefly-transmitted monopartite geminivirus. *Nucleic Acids Research* **19**, 6763–6769.

LAZAROWITZ, S. G. (1988). Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus. *Nucleic Acids Research* **16**, 229–250.

LAZAROWITZ, S. G. (1992). Geminiviruses: genomes structure and gene function. *Critical Reivews in Plant Science* **11**, 327–349.

LAZAROWITZ, S. G. & LAZDINS, I. N. (1991). Infectivity and complete nucleotide sequence of the cloned genomic components of the bipartite squash leaf curl geminivirus with a broad host range phenotype. *Virology* **180**, 58–69.

MACDOWELL, S. W., MACDONALD, H., HAMILTON, W. D. O., COUTTS, R. H. A. & BUCK, K. W. (1985). The nucleotide sequence of cloned wheat dwarf virus DNA. *EMBO Journal* **49**, 2173–2180.

MAYO, M. A. & MARTELLI, G. P. (1993). New families and genera of plant viruses. *Archives of Virology* **133**, 496–498.

MORINAGA, T., IKEGAMI, M. & MIURA, K. (1993). The nucleotide sequence and genome structure of mungbean yellow mosaic geminivirus. *Microbiology and Immunology* **37**, 471–476.

MORRIS, B., COATES, L., LOWE, S., RICHARDSON, K. & EDDY, P. (1990). Nucleotide sequence of the infectious cloned DNA components of African cassava mosaic virus (Nigerian strain). *Nucleic Acids Research* **18**, 197–198.

MORRIS, B. A. M., RICHARDSON, K. A., HALEY, A., ZHAN, X. & THOMAS, J. E. (1992). The nucleotide sequence of the infectious cloned DNA component of tobacco yellow dwarf virus reveals features of geminiviruses infecting monocotyledonous plants. *Virology* **187**, 633–644.

MULLINEAUX, P. M., DONSON, J., MORRIS-KRSINICH, B. A. M., BOULTON, M. I. & DAVIES, J. W. (1984). The nucleotide sequence of maize streak virus DNA. *EMBO Journal* **3**, 3063–3068.

MYERS, E. W. & MILLER, W. (1988). Optimal alignments in linear space. *Computer Applications in Biosciences* **4**, 11–17.

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.

POLSTON, J. E., DODDS, J. A. & PERRING, T. M. (1989). Nucleic acid probes for detection and strain discrimination of cucurbit geminiviruses. *Phytopathology* **79**, 1123–1127.

ROCHESTER, D. E., KOSITRATANA, W. & BEACHY, R. N. (1990). Systemic movement and symptom production following agroinoculation with a single DNA of tomato yellow leaf curl geminivirus (Thailand). *Virology* **178**, 520–526.

ROCHESTER, D. E., FAUQUET, C. M., DEPAULO, J. J. & BEACHY, R. N. (1994). Complete nucleotide sequence of the geminivirus, tomato yellow leaf curl virus (Thailand isolate). *Journal of General Virology* **75**, 477–485.

ROJAS, M. R., GILBERTSON, R. L., RUSSEL, D. R. & MAXWELL, D. P. (1993). Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. *Plant Disease* **77**, 340–347.

RYBICKI, E. P. & HUGHES, F. L. (1990). Detection and typing of maize streak virus and other distantly related geminiviruses of grasses by polymerase chain reaction amplification of a conserved viral sequence. *Journal of General Virology* **71**, 2519–2526.

SAITOU, N. & NEI, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

SHUKLA, D. D. & WARD, C. W. (1988). Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. *Journal of General Virology* **69**, 2703–2710.

SHUKLA, D. D. & WARD, C. W. (1989). Identification and classification of potyviruses on the basis of coat protein sequence data and serology. *Archives of Virology* **106**, 171–200.

SNEATH, P. H. A. & SOKAL, R. R. (1973). *Numerical Taxonomy*. San Francisco: W. H. Freeman.

STANLEY, J. (1985). The molecular biology of geminiviruses. *Advances in Virus Research* **30**, 139–177.

STANLEY, J. & GAY, M. R. (1983). Nucleotide sequence of cassava latent virus DNA. *Nature* **301**, 260–262.

STANLEY, J., MARKHAM, P. G., CALLIS, R. J. & PINNER, M. S. (1986). The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. *EMBO Journal* **5**, 1761–1767.

TORRES-PACHECO, I., GARZON-TIZNADO, J. A., HERRERA-ESTRELLA, L. & RIVERA-BUSTAMANTE, R. F. (1993). Complete nucleotide sequence of pepper huasteco virus: analysis and comparison with bipartite geminiviruses. *Journal of General Virology* **74**, 2225–2231.

WILBUR, W. J. & LIPMAN, D. J. (1983). Rapid similarity searches of nucleic acid and protein data banks. *Proceedings of the National Academy of Sciences, USA* **80**, 726–730.

WOOLSTON, C. J., BARKER, R., GUNN, H., BOULTON, M. I. & MULLINEAUX, P. M. (1988). Agroinfection and nucleotide sequence of cloned wheat dwarf virus DNA. *Plant Molecular Biology* **11**, 35–43.

(*Received 11 April 1993; Accepted 13 September 1994*)