

Diversity and Global Distribution of Whitefly-Transmitted Geminiviruses of Cotton

J. K. Brown
Department of Plant Sciences
University of Arizona, Tucson, AZ

Abstract

Geminivirus diseases of cotton are on the rise, worldwide, yet few have been studied in adequate detail to permit the implementation of rational approaches to disease control. The rising costs of managing the whitefly vector, coupled with substantial losses caused by geminivirus-incited diseases now hinder cotton production by requiring inputs that are beyond economic feasibility. The need for geminivirus disease resistant cultivars in diverse cotton producing areas and against different viral genotypes presents a new challenge. To meet this need, information about the identity, distribution, and relevant biotic characteristics of cotton-infecting geminiviruses is needed. This project addresses this problem through the molecular analysis of the genomes of cotton-infecting geminivirus from cotton throughout the world. Here, sequence similarities of the coat protein gene and of the non-coding IR/CR involved in regulating virus replication and transcription were examined by comparative sequence analysis to achieve virus identification. This is the first effort to determine virus identity and to map the distribution of geminiviruses on a global basis. The outcome of this effort will be a data base containing biotic and molecular information that will permit rapid and accurate geminivirus identification, and the selection of relevant viral species for development of cotton cultivars with disease resistance to the geminiviruses specific to individual production areas.

Introduction

Until quite recently, whitefly-transmitted (WFT) geminiviruses were known primarily to infect non-cultivated plants in dry subtropical and tropical climates. The increasingly wider distribution and higher population levels of the whitefly vector in agroecosystems are directly implicated in the development of new geminivirus epidemics in cotton-vegetable agroecosystems throughout the world. This is due to the increased use of pesticides to control primary insect and vector pests, and to the cultivation of high yielding cotton cultivars that are not tolerant or resistant to geminivirus pathogens.

Whitefly-transmitted geminivirus-incited diseases result in reduced growth and significant reductions in yield and quality of cotton. Virus infection at seedling or subsequent, early developmental stages typically results in little to no production in most widely grown vegetable and cotton cultivars. To date, there are few geminivirus-resistant crop varieties available, though much effort is in progress to control geminivirus diseases through development of resistant varieties that are suitable to specific locations and market needs.

Only a few geminiviruses of cotton have been studied to date. Cotton leaf crumple disease was first reported in the irrigated desert southwest in the 1950's (Brown, 1992). This disease of cotton was shown to be caused by a whitefly-transmitted geminivirus, CLCV. CLCV is not usually economically limiting to cotton production in the

US because in most years, although virus is widely prevalent, infection occurs late in the season, doing minimal damage (Brown and Nelson, 1984; 1987; Brown et al., 1987; Butler et al., 1986). During the 1950 to the 1980's, declining cotton production in Central America and the Caribbean Basin was attributed to whitefly-transmitted geminiviruses (J. Bird, per comm). In the Eastern Hemisphere, cotton production has experienced increasing threats from diseases caused by locally occurring geminiviruses that are transmitted by the indigenous *B. tabaci* populations (Brown, 1992; 1998). Among the additional countries reporting geminivirus diseases of cotton are Cameroon, India, Malawi, Mali, South Africa, Sudan, and Pakistan (J. K. Brown, 1998, APS Cotton Compendium, in press) though for the most part, these viruses are generally poorly studied. In Pakistan, at least one whitefly-transmitted geminivirus has been described from cotton and is thought to be a primary causal agent in the 1993-1995 epidemic (Mansoor et al., 1993).

Members of the *Bemisia tabaci* (Genn.) complex, among which are the tobacco, cotton, or sweet potato whitefly, and *B. argentifolli*, or the B biotype of *B. tabaci*, are the only known whitefly vectors of subgroup III geminiviruses (Brown, 1996). *B. tabaci* is readily capable of establishing to extreme population levels, particularly in crops grown under irrigated, arid conditions in both field and greenhouse systems. In addition, this whitefly has the potential to colonize a wide range of dicotyledonous species, among which are primarily vegetable and fiber species of great importance to worldwide agricultural production efforts. Recent studies indicate that there are numerous populations of *B. tabaci* that vary somewhat in their capacity to develop high population densities and cause direct feeding damage, in the extent of their host ranges, and in the efficacy with which they can transmit geminiviruses (Bedford et al., 1994; Brown and Bird, 1995; Brown et al., 1995a,b,c). The establishment of the B biotype in cotton-vegetable agroecosystems in the sunbelt states of the US and throughout Latin America during 1986-1996, is considered the primary factor driving force behind the emergence of geminiviruses in cotton-vegetable agroecosystems.

Many distinct virus species are thought to infect cotton on a worldwide basis, though very little specific information is available. Consequently, there is much to be learned about the identity, the distribution, and the specific threats that these emerging geminiviruses pose to cotton production. Most problematic in devising measures to control these new diseases is the paucity of knowledge about the biotic and molecular characteristics of the most problematic viruses and their molecular epidemiology. For example, most viruses are as yet unidentified and remain unstudied. As a result, information about their distribution, host range, and virus-vector relationships are not available. This report describes the results of a two year study of the molecular sequences of taxonomically key regions of the genomes of cotton-infecting geminivirus in an initial effort toward identifying and mapping whitefly-transmitted geminiviruses of cotton on a global basis.

Materials and Methods

Total nucleic acid extraction, PCR, and DNA sequencing. Nucleic acids were extracted from leaves using the method of Doyle and Doyle, 1990 with minor modifications. Leaves were ground to powder while frozen in liquid nitrogen and immediately suspended in warm (60°C) cetyltrimethylammonium bromide (CTAB) buffer. Homogenates were extracted with an equal amount of chloroform:isoamyl alcohol (24:1), phases were separated by centrifugation at 9,000 x g for 10 min at 4°C, and supernatants were precipitated overnight with 2/3 vol isopropyl alcohol. Nucleic acids were collected by centrifugation at 9,000 x g for 10 min at 4°C, washed with 1 vol wash buffer containing 76% ethanol and 0.2 M sodium acetate, and collected by microcentrifugation at 4°C. Pellets were air- or vacuum-dried and resuspended (1:1 wt:vol) in Tris-EDTA (TE) buffer, pH 8.0.

The CR of the DNA A component and the coat protein gene (*AVI*) were amplified using polymerase chain reaction (PCR) [Saiki, 1988] and two different pairs of degenerate primers (Idris and Brown, 1998). PCR primers were synthesized at the Biotechnology Facility, University of AZ, Tucson. PCR was carried out as described (Idris and Brown, 1998; Wyatt and Brown, 1996). PCR products of the expected sizes were cloned into the plasmid vector (pCR® 2.1) using the TA Cloning Kit (Invitrogen, San Diego, CA) per manufacturer's instructions. Fragments containing putative viral sequences were identified by size approximation in miniprep screening, and clones were confirmed positive for viral inserts using the respective PCR primers used for initial amplification. Viral DNA sequences were obtained for a minimum of three clones each by automated sequencing at the Molecular Genetics

Facility, University of Georgia (Athens, GA) using an automated Applied Biosystems 373A (version 2.1.1) Sequencer. Clones were sequenced in both directions using the same primers employed to obtain PCR products or universal primers on the vector.

Reference viral CR and AVI sequences used are of representative geminiviruses and were obtained by truncating full length sequences obtained from GenBank. Cotton geminivirus sequences were assembled with the aid of SeqEdit in the DNASTAR software package (DNASTAR, Madison, WI). AVI and CR sequences were aligned using the Clustal multiple sequence alignment program 'MegAlign' (DNASTAR) which aligns using a distance matrix method with stepwise, cumulative comparative clustering, generating an arithmetic mean for all subsequent pairwise comparisons. Clustal calculates the distances between all pair-wise character comparisons, and makes no *a priori* assumptions about evolutionary histories. Reconstructed trees were analyzed by maximum parsimony (MP) with Phylogenetic Analysis Using Parsimony (PAUP) program, version 3.1.1 [Swofford, 1993 #240]. A single most parsimonious tree(s) was sought using a heuristic search method with stepwise addition of sequences and the tree-bisection-reconnection (TBR) random branch-swapping options [Swofford, 1993]. Bootstrap values were calculated for major nodes using the 50% majority rule to place confidence limits on the tree exhibiting the most parsimonious reconstruction. Geminivirus sequences used in comparative sequence analyses were obtained from GenBank.

Results

The results of a preliminary analysis of viral common region sequences by maximum parsimony (PAUP) (Fig 1) and mean calculated pairwise distances expressed as percent similarities are shown, along with iterations identified in the common region of select cotton isolates (Table 1; Table 2). Also presented are results of maximum parsimony analysis of the coat protein gene (Fig 2.) and corresponding mean % similarities calculated from the pairwise mean distance matrix (Table 3) for select geminivirus isolates from cotton. Distance and phylogenetic analyses of the coat protein gene of these isolates indicate that they have a distinct origin in either an Old or a New World site. For example, from these analyses it can clearly be determined that cotton leaf curl virus-Pak 1 (CLCuV-Pak1) is distinct from all other geminiviruses found thus far in either locale, and is an Old World virus species, quite distinct from the well-studied cotton leaf crumple virus of the southwestern US and Mexico desert.

Among the New World isolates examined here, cotton leaf crumple virus (CLCV), and possibly, strains thereof, was documented as the sole geminivirus species in Arizona (CLCVAZCgr94 &95), California (CLCVCalif 94&95), and in Caborca and the Mexicali Valley of Sonora, Mexico (MexCaborcaCYM, MexicaliVLCr 17&18), all three sites in which the disease has been previously documented. In addition, a close relative or strain of CLCV was documented for the first time in Guatemala (Guat57cot94). Also, at least one and possibly two apparently distinct and as yet undescribed geminiviruses of cotton were found in Texas (CotV9Tx61, Txcotgrmos, TxcotLCr, TxMontAl and TxcotLCr & CotV7Tx) and another in Guatemala (Guat2cot).

Further studies will now be required to substantiate these preliminary findings, including the complete DNA sequence from infectious clones and examination of the potential for *trans* activation of replication and movement functions between the components of different isolates. Additional geminivirus isolates from the Dominican Republic (1992), Brazil (1997-98), Sudan (1997), India (1998), and numerous other cotton growing regions of the world are presently under investigation using these approaches.

Discussion

Recent advances in the application of molecular biological methods to the characterization of geminiviruses have facilitated the investigation of cotton infecting geminiviruses. New approaches involve the application of polymerase chain reaction (PCR) and universal subgroup III that amplify the middle or core region of the coat protein gene (Wyatt and Brown, 1996), or geminivirus-specific primers to direct the amplification of key regions of the virus genome (Idris and Brown, 1998). Partial genome sequences of relevant genes or genomic regions are useful for establishing virus identity and relationships between viruses in the absence of a complete genomic sequence, the latter, a time-consuming and arduous task. Used in conjunction with biotic information about the isolate, and when compared for multiple viruses, coat gene and LIR/CR sequences provide important clues about relationships between whitefly-transmitted geminiviruses, as based upon the specific targeted or marker sites within the viral genome.

The geminivirus coat protein gene (AV1) and large viral intergenic/common region (LIR/CR) sequences have been shown to be useful in establishing the relative identity of geminiviruses (Brown et al., 1998, in preparation). The coat protein gene is informative because it contains sequences that are both highly conserved and regions that are variable to such an extent that phylogenetic inferences can be correlated to biotic and geographic characteristics (Padidam et al., 1995). The coat contains sequences that are conserved to function in the formation of the characteristic 'geminat' coat protein that encapsidates the ssDNA genome, plays a role in virus movement in the plant (Pooma et al., 1996), and is required for vector-mediated transmission (Brown, 1996).

The large intergenic region is considered to be an informative sequence of the geminivirus genome because it contains viral regulatory sequences essential for the disease cycle (Eagle and Hanley-Bowdoin, 1997) and for potentially interacting with other geminiviruses when they occur in a mixture in the same host, a possible means by which additional genotypes can emerge, possibly with distinct biological properties (Arguello-Astorga et al., 1994). Specifically, these sequences are important for predicting the likelihood of cross-replication, or pseudo-recombination, between compatible, and therefore, closely related isolates (in question), and is conserved at key sites to perform essential functions in the disease cycle.

Phylogenetic analysis of geminiviruses based upon the coat protein or LIR/CR sequences position the most closely related viruses within the same 'cluster' or group, whereas, those that are not as closely related are placed on different branches with their closest sequence relatives. Inclusion of the leafhopper/ planthopper relatives within the Geminiviridae in the analysis reveals clear separation of whitefly-transmitted viruses from subgroup I and II viruses. Within the whitefly subgroup, the viruses are clearly further separated by geography of origin (Eastern or Western Hemisphere), and at times by a further sub-geographic separation (Brown, 1996; Brown and Wyatt, 1995; 1996; Padidam et al., 1995).

Using this approach, the introduction of a geminivirus from one geographic world region into another can be readily detected. In line with predictive approach for the rational selection of relevant isolates for resistance screening or as sources of viral genes toward pathogen-derived resistance, we are examining the prospects of applying phylogenetic relationships to development of disease resistant germplasm with well-defined breadths of virus resistance. In this approach, it is possible to make predictions about the genotype and numbers of closely or distantly related geminiviruses in germplasm from breeding programs and in plants protected by virus-derived resistance. Though the approaches differ in possible resistance mechanisms, both hope to achieve resistance to as many viruses and strains as possible to accommodate for extant virus pathogens and those that may emerge from extant relatives in the future.

We hypothesize that germplasm with resistance to a geminivirus from the same or an adjacent geographic region may also afford protection against other closely related viruses, and provide less or little protection against more divergent viruses from distant geographic regions, the viruses having evolved either under more or less the same conditions and/or possibly from a biogeographically related common ancestor. Consequently, germplasm or transgenic plants expressing the gene of a particular geminivirus genotype will be most likely be effective against most closely related viruses as opposed to those evolving in biogeographic isolation. Clearly, this hypothesis cannot claim *a priori* knowledge about any particular mechanism operating in a resistant genotype, or whether there is a traceable evolutionary basis for a mechanism that is necessarily congruent with geminivirus pathogen evolution.

None the less, the broad theoretical and practical utilities of this diagnostic and predictive tool can not be underestimated.

The initial goal of this effort is to determine the identity and map the geographic distribution of the most prevalent geminiviruses of cotton on a global basis. Second, viral sequence data will be used to establish the phylogenetic relationships between cotton-infecting geminiviruses in relation to well-studied geminiviruses, using the analogous sequences as reference sequences having a biotic and geographic basis. Third, particular regulatory sequences will be evaluated to predict if interactions may be possible between viruses if present in mixed infections, a result that also lends insights to relationships at the virus strain versus species level.

The long range goal is to catalog and map whitefly-transmitted geminiviruses of cotton on a global basis, and establish a rational means for the selection of relevant viruses and/or strains toward developing customized geminivirus disease resistant cotton varieties. The most important virus species will then be selected according to the criteria of association with substantial disease losses and a widespread distribution in cotton. These viruses will be subjected to molecular cloning to obtain a full length infectious virus clones and their complete DNA sequences, thus achieving for the first time, the isolation of these viruses in 'pure culture', reproduction of disease symptoms after inoculating cotton with the infectious clones, and the end result being, characterized geminiviruses of cotton for disease resistance efforts.

References

- Bedford, I.D., P.G. Markham, J.K. Brown and R.C. Rosell. 1994. Geminivirus transmission and biological characterization of whitefly (*Bemisia tabaci*) biotypes from different world regions. *Ann. appl. Biol.* 125: 311-325.
- Brown, J. K. 1994. Current Status of *Bemisia tabaci* Genn. as a pest and vector in world agroecosystems. *FAO Plant Prot. Bull.* 42: 3-32.
- Brown, J. K. 1996. Chapter 5 in: *Molecular Biology and Epidemiology of Subgroup III, Geminiviridae*. Plant-Microbe Interactions Review Series, G. Stacey and N. Keen, eds (Chapman and Hall) pp 125-195.
- Brown, J. K., D. Frohlich, and R. Rosell, 1995a. The Sweetpotato/Silverleaf Whiteflies: Biotypes of *Bemisia tabaci* Genn., or a Species Complex? *Ann. Rev. Entomology* 40: 511-534.
- Brown, J. K., S. Coats, I. D. Bedford, P. G. Markham, J. Bird, and D. R. Frohlich. 1995c. Characterization and distribution of esterase electromorphs in the whitefly, *Bemisia tabaci* (Genn.) (Homoptera:Aleyrodidae). *Biochemical Genetics*: 33: 205-214.
- Brown, J.K. 1992. Virus Diseases of Cotton, pages 275-330 in: *Cotton Diseases*. R. J. Hillocks, ed. Commonwealth Agricultural Bureaux International, Oxon, United Kingdom. 415 pp.
- Brown, J.K. and J. Bird. 1992. Whitefly-transmitted geminiviruses in the Americas and the Caribbean Basin: past and present. *Plant Dis.* 76:220-225.
- Brown, J.K. and Nelson, M.R. 1984. Geminiate particles associated with cotton leaf crumple disease in Arizona. *Phytopathology* 74:987-990.
- Brown, J.K. and Nelson, M.R. 1987. Host range and vector relationships of cotton leaf crumple virus. *Plant Dis.* 71:522-524.
- Brown, J.K., Mihail, J.D., and Nelson, M.R. 1987. The effects of cotton leaf crumple virus on cotton inoculated at different growth stages. *Plant Dis.* 71:699-703.

Eagle, P. A. and L. Hanley-Bowdoin. 1997. *cis* Elements that contribute to geminivirus transcriptional regulation and the efficiency of DNA replication. *Journal of Virology* 71: 6947-6955.

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

Idris, A. M. and J. K. Brown. 1998. Sinaloa tomato leaf curl virus: biological and molecular evidence for a new subgroup III geminivirus. *Phytopathology* (in press).

Lazarowitz, S. G., L. C. Wu, S. G. Rogers, and J. S. Elmer. 1992. Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. 1992. *Plant Cell* 4:799-809.

Padidam, M., R. N. Beachy, and C. F. Fauquet. 1995. Classification and identification of geminiviruses using sequence comparisons. *J. Gen. Virol.* 76: 249-253.

Swofford, D. 1993. *PAUP: Phylogenetic Analysis Using Parsimony*, version 3.1.1. Washington D.C.: Smithsonian Institution.

Wyatt, S. D., and J. K. Brown. 1996. Detection of subgroup III geminivirus isolates in leaf extracts by degenerate primers and polymerase chain reaction. *Phytopathology* 86: 1288-1293.

Table 1. Whitefly-transmitted geminivirus isolates that were PCR-positive for presence of subgroup III geminiviruses and the * denotes those used in this study.

<u>Designation</u>	<u>Symptom Cotton</u>	<u>Geographic</u>	<u>Source Plant</u>	<u>Year</u>
abmvroth		WestIndies	<i>Abutilon</i>	1850?
azhibis93		AZ, USA	hibiscus	1993
clcvaz	leaf crumple	AZ, USA	cotton	1982
clcvaz92	leaf crumple	AZ, USA	cotton	1992
clcvazcgr94*	leaf crumple	AZ, USA	cotton	1994
clcvazcgr95*	leaf crumple	AZ, USA	cotton	1995
mparvbm97		AZ,USA	<i>M. parviflora</i>	1997
drcot	mosaic	Dom Rep	cotton	1992
egyptcot95	leaf curl	Egypt	cotton	1995
guat2cot94*	leaf crumple	Guatemala	cotton	1994
guat57cot94*		Guatemala	cotton	1994
guatabut94		Guatemala	<i>Abutilon</i>	1994
guatcot294*	mosaic	Guatemala	cotton	1994
guatcot92ym	yellow mosaic	Guatemala	cotton	1992
guatmalvweed22 *		Guatemala	Malv weed	1994
guatsidaspp		Guatemala	Sida spp.	1994

hibis94		AZ, USA	hibiscus	1994
indiacot96	leaf curl	Punjab, India	cotton	1996
mexcabcym96*	yellow mottle	Sonora, Mex	cotton	1996
mexcablcr96*	leaf crumple	Sonora, Mex	cotton	1996
mexicalivlcr96*	leaf crumple	Sonora, Mex	cotton	1996
mx94okra		Mexico	okra	1995
pakclcuv*	leaf curl	Faisal, Pak	cotton	1993
pakcot	leaf curl/mild	Pakistan	cotton	1996
prrmv	veinal yellows	Puerto Rico	<i>R. minima</i>	1997
sudan okra	leaf curl/suspect	Sudan	okra	1995
sudan okra	leaf curl/suspect	Sudan	okra	1997

Table 2. Range of mean pairwise distance comparisons of large intergenic/common region sequences for selected geminivirus isolates from cotton.

Isolate and #taxa	Greatest % Similarity with Reference Virus	Least % Similarity with Reference Virus
CLCVAz n=3	54.1-54.5 TTMoV	42.3-42.6 TLCrV 43.6 ACMV
CLCVCab n=2	54.1-54.5 ToMoB 54.9 BGMV (PR&Jam)	43.2 TLCrV 43.6 ACMV
CLCVMexV n=5	53.4 TTMoV 54.3 ToLCVCh	43.0-43.2 TLCrV 43.7 ACMV
Texcot-1 (Leaf Curl) n=2	54.1 TTMoV 54.8% BGM(PR&Jam)	41.6 TLCrV, ACMV
Texcot-2 mosaic/stunting n=4	56.6% AbMV	40.8% ACMV

Table 3. Range of mean pairwise distance comparisons of coat protein gene (AVI) sequences for selected geminivirus isolates from cotton.

Isolate #Taxa	Range within Each Taxon- (% Similarity)	Greatest and Least % Similarity with Reference Virus
CLCVAz n=4	.0012-.012	98.9-98.8
CLCVCalif n=4	.011-.030	97-98.9
CLCVCab n=4	.001-.003	99.7-99.9
CLCVMexV n=2	.003	99.7-99.9
Guat2 n=2	.008	99.2
Guat57 n=1	-	-
Texcot-1 (Leaf Curl) n=2	.009	99.1
Texcot-2 mosaic n=4	.000-.012	0-98.8

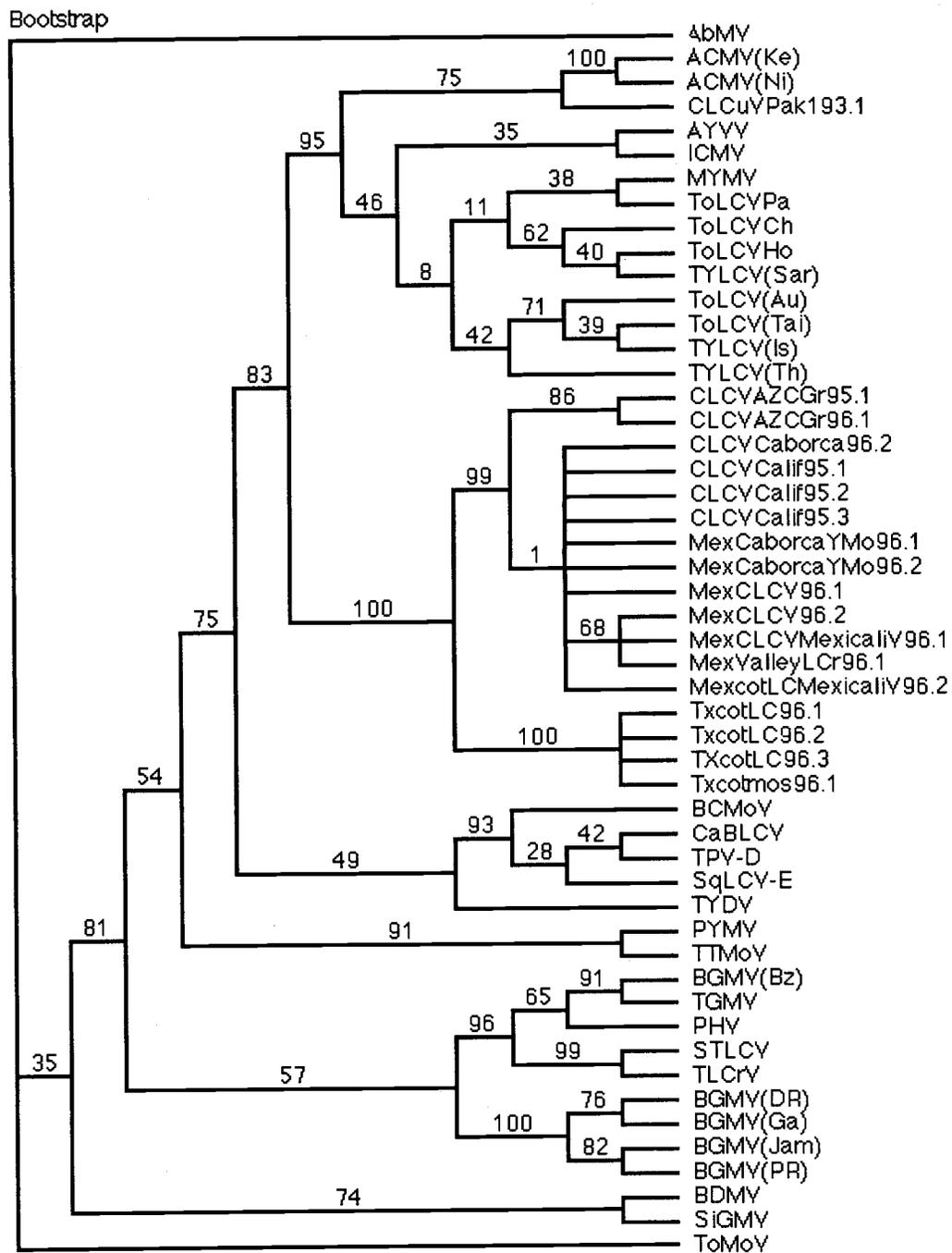


Figure 1. Single most parsimonious tree (unrooted) for the intergenic/common region relationships of 24 geminivirus isolates and reference Geminiviridae taxa. Branch lengths are approximately proportional to the number of nucleotide changes. Maize streak virus (subgroup I) and beet curly top virus (subgroup II) are the outgroups. Bootstrap sampling (10 replicates/100 iterations) employed the heuristic search option and TBR branch swapping. Bootstrap percentages are indicated on nodes for the 50% majority rule.

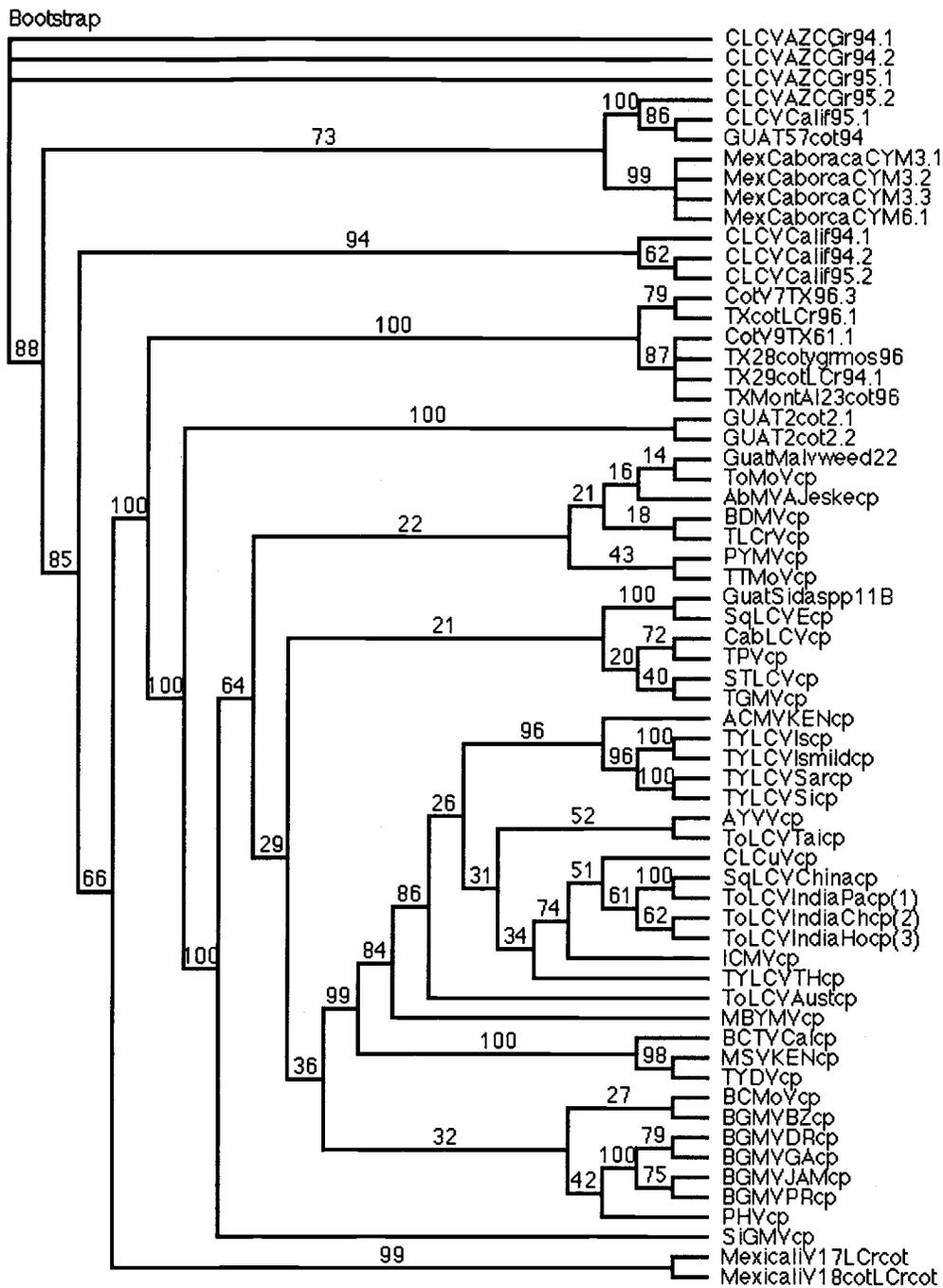


Figure 2. Single most parsimonious tree (unrooted) for coat protein gene (AV1) relationships of 24 geminivirus isolates and reference Geminiviridae taxa. Branch lengths are approximately proportional to the number of nucleotide changes. Maize streak virus (subgroup I) and beet curly top virus (subgroup II) are used as outgroups. Bootstrap sampling (10 replicates; 100 iterations) employed the heuristic search option with TBR branch swapping. Bootstrap percentages are indicated on nodes using the 50% majority rule.