

Sequence Analysis and Classification of Apparent Recombinant Begomoviruses Infecting Tomato in the Nile and Mediterranean Basins

C. M. Fauquet, S. Sawyer, A. M. Idris, and J. K. Brown

First author: International Laboratory for Tropical Agricultural Biotechnology, Danforth Plant Science Center, St. Louis, MO 63132; second author: Department of Mathematics, Washington University, St. Louis, MO 63105; and third and fourth authors: Department of Plant Sciences, University of Arizona, Tucson 85721.

Accepted for publication 25 January 2005.

ABSTRACT

Fauquet, C. M., Sawyer, S., Idris, A. M., and Brown, J. K. 2005. Sequence analysis and classification of apparent recombinant begomoviruses infecting tomato in the Nile and Mediterranean Basins. *Phytopathology* 95:549-555.

Numerous whitefly-transmitted viral diseases of tomato have emerged in countries around the Nile and Mediterranean Basins the last 20 years. These diseases are caused by monopartite geminiviruses (family *Geminiviridae*) belonging to the genus *Begomovirus* that probably resulted from numerous recombination events. The molecular biodiversity of these viruses was investigated to better appreciate the role and importance of

recombination and to better clarify the phylogenetic relationships and classification of these viruses. The analysis partitioned the tomato-infecting begomoviruses from this region into two major clades, *Tomato yellow leaf curl virus* and *Tomato yellow leaf curl Sardinia virus*. Phylogenetic and pairwise analyses together with an evaluation for gene conversion were performed from which taxonomic classification and virus biodiversity conclusions were drawn. Six recombination hotspots and three homogeneous zones within the genome were identified among the tomato-infecting isolates and species examined here, suggesting that the recombination events identified were not random occurrences.

Many diseases of tomato caused by whitefly-transmitted geminiviruses of the genus *Begomovirus* (family *Geminiviridae*) are known worldwide (5,6). Of these, 23 have been approved as distinct species, while at least another 24 are considered provisional species (available online from the Danforth Center) (9) and it is likely that additional undescribed begomoviruses occur and thereby, are yet unclassified. Typical symptom phenotype descriptors for tomato-infecting begomovirus diseases are leaf curl, yellow leaf curl, leaf crumple, and foliar mosaic and mottle, among others. Such descriptors are often employed in various combinations and are also used to categorize viruses provisionally, which can lead to confusion in nomenclature when viral identity is ultimately established. This often occurs because different viral species can cause similar or indistinguishable symptom phenotypes in the same plant host. Likewise, symptoms may vary for the same virus-host combination, depending on the particular cultivars, ecological and environmental conditions, or differences characteristic of a particular strain or variant. Furthermore, infection of the same host by two or more viruses also is common, further complicating disease diagnosis, and collectively, making symptom phenotype an unreliable criteria by which to establish virus nomenclature and identity. Subsequently, the value of expedient virus identification through DNA sequence comparisons and phylogenetic analysis with sequences for well-studied begomoviruses is evident.

Even so, when the complete sequence of a viral genome becomes available, it may still be difficult to appreciate the extent to which a new species differs from or is similar to previously described begomoviral species, primarily because begomoviruses

are capable of undergoing inter- and intraspecific recombinations (4-6,13,15). Recently, recombinant begomoviruses have been reported for viruses infecting bean (7) and pepper in Mexico, *Macroptilium lathyroides* (L.) in Puerto Rico (19), cotton in North America (17), *Sida* spp. from Central America (12), and cassava (8,11,25,31), cotton (16), and tomato (18,24) from Africa and the Mediterranean region.

Presently there is no clear understanding of the specific biological consequences of begomoviral recombination. However, it is clear that very recently noted recombinants such as *Tomato yellow leaf curl Malaga virus* from Spain (TYLCMaIV) (23) and *East African mosaic virus-Uganda2* (EACMV-UG2) (8,25,31) have emerged as new and important plant pathogens that occupy new niches and are expected to have the potential to form additional populations or genotypes that are capable of possibly participating in rapid evolution schemes. In this study, the molecular diversity and taxonomic relationships have been assessed for a group of recently described begomoviruses causing, or having an association with, recent outbreaks and/or epidemics in the Mediterranean Basin and on the African continent. Particular attention has been given to interspecific recombination, which has apparently occurred during the evolution of most if not all of these viruses.

MATERIALS AND METHODS

Virus sequences. At the time of the analysis, 19 complete monopartite sequences of begomovirus sequences were available for tomato-infecting viruses from Egypt, Italy, Iran, Israel, Portugal, Spain, and Sudan. Also included were the complete genome sequences for isolates of the Old World *Tomato yellow leaf curl virus* (TYLCV) identified during the last decade in Cuba, the Dominican Republic, and Japan following their introduction from the Mediterranean and Middle Eastern regions.

Corresponding author: J. K. Brown; E-mail address: jbrown@ag.arizona.edu

DOI: 10.1094/PHYTO-95-0549

© 2005 The American Phytopathological Society

Reference sequences with their species names and the respective acronym (in parenthesis) and GenBank accession number {in brackets} were TYLCV from Israel (TYLCV) {X15656}, TYLCV from Aichi (TYLCV-[Aic]) {AB014347}, TYLCV from Cuba (TYLCV-[CU]) {AJ223505}, TYLCV from the Dominican Republic (TYLCV-[DO]) {AF024715}, TYLCV from Egypt (TYLCV-[EG]) {L12219}; TYLCV from Iran (TYLCV-[IR]) {AJ132711}, TYLCV-mild from Israel (TYLCV-Mld) {X76319}, TYLCV from Portugal (TYLCV-[PO]) {AF105975}, TYLCV from Shizouka (TYLCV-[Shi]) {AB014346}, TYLCV from Spain (TYLCV-[SP72]) {AF071228}, TYLCV from Sudan (TYLCV-[SD]) {AY044138}, *Tomato yellow leaf curl Sardinia virus* (TYLCSV) {X61153}, TYLCSV from Spain (TYLCSV-[ES1]) {Z25751}, TYLCSV-[ES2] {L27708}, TYLCSV from Sicily (TYLCSV-[Sic]) {Z28390}, TYLCM from Spain (TYLCMaV) {AF271234}, *Tomato leaf curl Sudan virus* (ToLCSDV), Gezira isolate (ToLCSDV-[Gez]) {AY044137}, and Shambat isolate (ToLCSDV-[Sha]) {AY044139}.

Phylogenetic analysis. All sequences for the full genome (monopartite viruses) or the DNA-A component (bipartite viruses) for the selected viruses were aligned using the Clustal V method of aligning multiple sequences (14) available with the DNASTAR package (DNASTAR Inc., Madison, WI). A gap penalty of 10 and a gap length penalty of 10 were used throughout. The phylogenetic tree was estimated using the neighbor-joining method (26) with the unweighted pair-group method with arithmetic average (UPGMA) distance matrix (MegAlign program). Phylogenetic analysis also was conducted using a cladistic parsimony method and phylogenetic analysis using parsimony (PAUP) (version 4.10b, Sinauer Associates, Sunderland, MA) (30). One-thousand bootstrap replications were performed to place confidence estimates on major groups resolved in the tree. The same multiple alignments were subjected to two approaches to detect intermolecular recombination.

Recombination analyses. Two methods were used to search for intermolecular recombination for this set of tomato-infecting begomoviruses. The first employed the program GeneConv, developed by Sawyer (28), and the second involved a pairwise analysis using the pairwise comparison option (Align) available in DNASTAR (version 1.02). The multiple sequence alignment program (MegAlign 3.11) within the DNASTAR package was used to align viral sequences. The GeneConv method was employed to examine begomoviral sequences for evidence of intermolecular recombination.

Each system has advantages and disadvantages and the two types of analysis together provide a better picture of the distribution of the putative recombinant fragments in a subset of viral sequences, to produce a putative recombinant map for the sequences. GeneConv compares DNA sequences (usually 50 nucleotides [nts]) in a sliding window and polymorphisms that are

identified are compared with a pair of chosen sequences. A *P* value is then calculated, which provides an estimate of the probability of an outcome due to a random event. In the pairwise analysis approach, the profile of a pair of sequences is compared with that for an average profile of sequences that are selected a priori, based on knowledge that the selected sequences are related to the species or the strain clusters because they are all members of the same clade. The analyzed sequences determine the relevance of the chosen pair by analysis of sequence segments of 50 nts. The switch from one category to the other of the considered fragments identifies putative recombinant fragments, and no statistical test is calculated.

GeneConv method. In GeneConv, the term “fragment” is used to denote an aligned or homologous pair of segments in an alignment. The highest-scoring fragments in the alignment are listed and are assigned *P* values based on the assumption of a random distribution of polymorphic sites. The scores are defined as follows. First, all sites that are monomorphic in the alignment are discarded, so that only polymorphic sites are considered. For a given pair of sequences, matching bases are scored as +1 and mismatches as -m, where m depends on the pair of sequences. Fragments are assigned *P* values in a manner similar to the BLAST procedure (2,3,20). This *P* value is an approximation of the proportion of permutations of the polymorphic sites for which that pair of sequences has some fragment with the observed score or larger (28). The mismatch penalty m is given by a proportionality constant (“g scale”) divided by the ratio of the total number of site differences between the pair of viruses and the total number of alignment polymorphisms. The g scale is rounded up to the nearest integer. Lower g scale values should be more sensitive to older recombination events or for higher mutation rates (1,21).

Given a g scale value, the program finds all non-overlapping maximal fragments that have Bonferroni-corrected (22) *P* values of 0.05 or less. The *P* values are Bonferroni or multiple comparison-corrected by multiplying the BLAST-like Karlin-Altschul *P* value by the number of pairs of sequences in the alignment (22). Permutation tests based on a score equivalent to this *P* value show that both the pairwise and Bonferroni-corrected *P* values are conservative. A g scale value of 1 was used in the analyses. GeneConv, a program that finds, sorts, and scores all significant fragments in an alignment is available from S. Sawyer (available online from Washington University Mathematics Department, St. Louis). Each GeneConv analysis writes a spreadsheet-like file in which the significant fragments are listed in decreasing order using a Bonferroni-corrected *P* value.

Pairwise analysis. According to the guidelines established by the ICTV Geminiviridae Study Group, two geminivirus sequences sharing more than 89% identity of their A component sequences are considered strains or isolates of the same species, otherwise

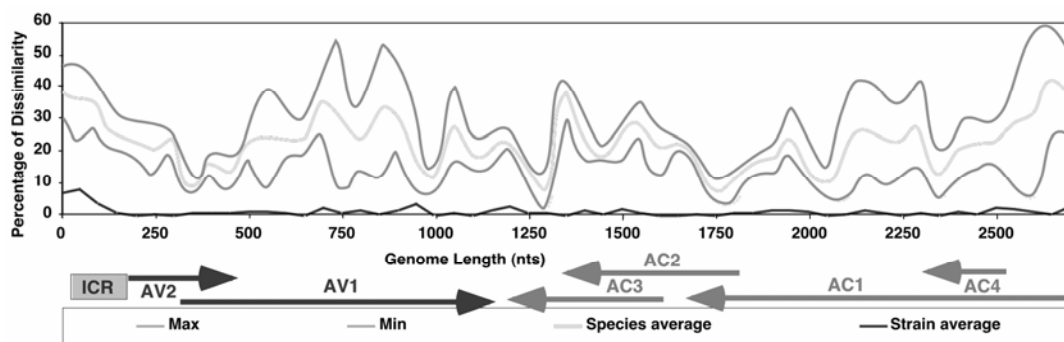


Fig. 1. Pairwise sequence comparison profiles of begomoviruses of the same clade. The horizontal axis represents the genome linearized at the nick site associated with replication (position 1) and the vertical axis represents the percentage of dissimilarity of two sequences calculated in segments of 50 nucleotides along the genome. The white curve represents the species average of the chosen clade; the gray curves represent the minimum (Min) and maximum (Max) of the average values (± 2 standard deviations), respectively. The black curve represents the “strain average” in the chosen clade. The begomovirus genome organization is represented under the *x* axis, using the conventional code system for genes and noncoding sequences.

they are considered different species (9). However, viruses that share between 80 and 90% nucleotide identities are often found to be recombinants (10); therefore, we propose that recombinant begomoviruses, which share greater than 80% nucleotide identity, should be considered the same species. Pairwise sequence comparison profiles were carried out between sequences of different species and of different strains and an average profile for the considered cluster of viruses was calculated for these two categories

with increments of 50 nts. A standard deviation value for each segment was calculated and a minimum and maximum value corresponding to 2 standard deviation values was also calculated (Fig. 1). Each chosen pairwise analysis for putative recombinant sequences was then compared with the species average profile, and the pertaining of each 50-nt fragment to this category was examined. Segments differing by more than 2 standard deviation values were considered to be putative recombinant fragments. For each pairwise analysis, a putative recombination percentage for the genome was calculated and a corresponding map can be drawn. It has been verified (a posteriori) that the particular species and strains selected for the species and strain average curve are 100% nonrecombinant.

RESULTS

Phylogenetic analysis. All of the branches of the phylogenetic tree reconstructed using maximum parsimony had significant bootstrap values at 70% indicating that all branches are very stable (Fig. 2). Only TYLCMaIV, which is a hybrid between TYLCSV and TYLCV, had a marginal bootstrap value at 56%. TYLCV isolates separated into two major clades, one that contains the “prototype” TYLCV from Israel, and the second that contains the “mild” strain of the TYLCV, also described from Israel. The newly described TYLCV isolate from Sudan, TYLCV-[SD], clusters with the TYLCV-Mild clade, while TYLCV-[IR] from Iran clusters with the TYLCV clade. ToLCSDV-[Gez] and ToLCSDV-[Sha] cluster together and form a separate branch located between the TYLCV-Mild and TYLCSV clades, of which the latter relationship supports the proposal that these viruses belong to a new species (18).

The nucleotide distance matrix for these viruses (Fig. 3) indicated that the nucleotide identity percentages range from 70 to

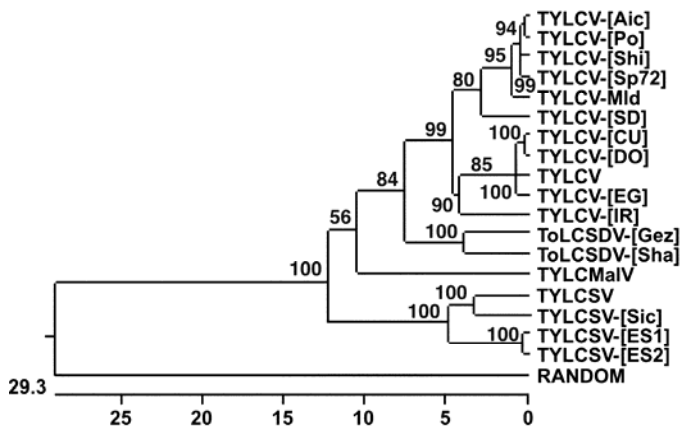


Fig. 2. Phylogenetic tree of 19 begomovirus DNA sequences. Complete DNA sequences of monopartite begomoviruses infecting tomato in the Nile and Mediterranean Basins were aligned. The tree was generated using the MegAlign program available with the DNASTAR package (Madison, WI). Vertical distances are arbitrary and the scale below the tree indicates the distance between sequences. The tree is unrooted. A bootstrap analysis at 70% confidence with 500 repetitions was performed using the PAUP 4.0 software, and the bootstrap values of each branch were reported at each major node on the phylogenetic tree.

	TYLCV-[Aic]	TYLCV-[PO]	TYLCV-[Shi]	TYLCV-[Sp72]	TYLCV-Mild	TYLCV-[SD]	TYLCV-[CU]	TYLCV-[DR]	TYLCV	TYLCV-[EG]	TYLCV-[IR]	TYLCSDV-[Gez]	TYLCSDV-[Sha]	TYLCMaIV	TYLCSV	TYLCSV-[Sic]	TYLCSV-[ES1]	TYLCSV-[ES2]
TYLCV-[Aic]	***	99	98	98	97	93	90	90	90	90	87	83	83	89	73	73	75	75
TYLCV-[PO]		***	99	99	97	93	91	91	90	90	88	83	83	89	73	73	74	74
TYLCV-[Shi]			***	99	97	93	91	91	90	91	88	84	83	89	73	73	74	73
TYLCV-[Sp72]				***	97	93	89	91	90	90	88	83	83	89	73	73	74	73
TYLCV-Mild					***	92	90	90	89	90	88	83	83	88	73	73	74	74
TYLCV-[SD]						***	87	87	87	87	86	83	83	86	72	73	75	74
TYLCV-[CU]							***	99	98	98	90	81	78	81	71	71	72	71
TYLCV-[DR]								***	98	98	90	81	78	81	71	71	72	71
TYLCV									***	97	89	79	78	80	70	70	71	71
TYLCV-[EG]										***	89	80	78	81	72	71	72	71
TYLCV-[IR]											***	79	78	79	72	72	73	72
TYLCSDV-[Gez]												***	90	80	70	71	72	71
TYLCSDV-[Sha]													***	81	71	71	72	71
TYLCMaIV														***	79	80	85	85
TYLCSV															***	91	86	86
TYLCSV-[Sic]																***	90	90
TYLCSV-[ES1]																	***	99
TYLCSV-[ES2]																		***

Fig. 3. Matrix of percentage of identity between genomic sequences of begomoviruses infecting tomatoes in the Nile and Mediterranean Basins. The numbers represent the percentage of identity between a pair of sequences. The squares represent clusters of viruses that follow the species demarcation of 89% identity.

100%. The first evidence is the clear partitioning between TYLCV isolates and TYLCSV isolates that share only 70 to 75% identity. The TYLCSV isolates form two different clades that share less than 89%, the newly recognized threshold between species and strains, which possibly suggests their separation into two species. The recombinant TYLCMaIV shares 89% identity with TYLCV isolates of the TYLCV-Mld clade but only 79 to 86% nucleotide identity with all the other TYLCV and TYLCSV isolates.

Recombination analyses. *GeneConv analysis.* The GeneConv analysis performed on the 19 sequences found 104 inner fragments and 168,562 overlapping fragments that were discarded. Among the 104 fragments, we have listed 33 in Table 1, and Figure 4B shows all 33 fragments involving 1 of the 10 viruses best represented in the phylogenetic study. A total of nine sequences representing close isolates of these 10 viruses were discarded for simplification. The average length of the fragments was 1,101 nts and the average *P* value was 1.14×10^{-3} . Nine fragments had a *P* value inferior to 10^{-20} and seven of those involved the hybrid TYLCMaIV, which is a very recent recombinant, hence the low *P* values. The newly described species ToLCSDV was involved 15 times, while the new strain TYLCV-[SD] was involved six times.

A frequency distribution of the recombination sites for this tomato virus population revealed that they are not distributed at random along the genome organization (Fig. 4A), but have privileged positions. We can identify six major locations at positions 100, 550, 1150, 1300, 2000, and 2800. These positions are identical for a population of 19 viruses with 104 recombinant fragments or for 10 viruses with 33 recombinant fragments (data not shown). These locations correspond in the begomovirus genome

to the intergenic region (IR) (100 and 2800); the pre-coat protein open reading frame (ORF), AV2 (550); the site between the coat protein (CP) (AV1) and the replication enhancer protein, AC3 ORF (1150 and 1300); and a specific region of the replication-associated protein (REP), AC1 ORF (2000).

These recombination sites define several relatively homogeneous zones in the viral genomes (Fig. 4B), namely (i) the AV1 and AV2 domains, (ii) the AC3 ORF, (iii) the AC2 ORF and a portion of the AC1 (REP) domain, and (iv) the N-terminal AC1-IR domain. The later is the most heterogeneous domain with numerous recombination sites spanning the entire region.

Pairwise analysis. The results of the pairwise analysis performed on tomato-infecting begomoviral sequences are presented in Figure 5. The analysis delineated two major populations of viruses, those with a TYLCV backbone and those with a TYLCSV backbone. In addition, TYLCMaIV is a perfect hybrid between the latter two types, with about 40% of its genome being typical of TYLCSV (23). The map also reveals a number of fragments of different sizes in different locations, but only four domains were recognized herein with confidence for these viruses. The first domain encompasses the AV2 ORF, the second one the CP gene, the third one the AC3, AC2, and C-term of the AC1 ORFs, and the fourth one the N-term of the AC1 ORF and the IR. The most heterogeneous domain is the later, which has numerous smaller fragments belonging to various tomato-infecting viruses examined here.

DISCUSSION

The two methods for recombination analysis used in this study revealed that most of the recombined fragments are similarly

TABLE 1. List of putative recombinant fragments among 10 viruses infecting tomato in the Nile and Mediterranean Basins^a

Sequence names		Karlin-Altschul <i>P</i> value	Aligned offsets		
Seq1	Seq2		Begin	End	Length
TYLCSV-[ES1]	TYLCMaIV	1.24E-41	2,802	1,154	1,176
TYLCSV-[ES2]	TYLCMaIV	1.17E-40	2,802	1,154	1,176
TYLCMaIV	TYLCV-Mld	2.98E-31	1,139	2,738	1,600
TYLCMaIV	TYLCV-[IR]	3.66E-28	1,139	2,164	1,026
TYLCMaIV	TYLCV	2.02E-27	1,139	1,999	861
TYLCV-[IR]	TYLCV-Mld	3.29E-24	62	2,144	2,083
ToLCSDV-[Sha]	TYLCMaIV	2.98E-23	1,300	22	1,546
ToLCSDV-[Gez]	TYLCMaIV	7.01E-22	1,139	2,260	1,122
TYLCV	TYLCV-Mld	1.53E-20	70	1,992	1,923
TYLCMaIV	TYLCV-[SD]	4.19E-18	1,333	22	1,513
TYLCV	TYLCV-[IR]	2.84E-16	2,806	1,987	2,005
TYLCV	TYLCV-[SD]	1.28E-14	95	1,987	1,893
ToLCSDV-[Gez]	TYLCV-Mld	3.67E-14	1,138	2,260	1,123
ToLCSDV-[Gez]	TYLCV-[SD]	5.69E-14	1,153	2,177	1,025
ToLCSDV-[Gez]	TYLCV-[IR]	8.76E-13	1,090	2,144	1,055
ToLCSDV-[Sha]	TYLCV-Mld	1.56E-12	1,297	2,745	1,449
TYLCV-[IR]	TYLCV-[SD]	6.25E-12	250	2,005	1,756
ToLCSDV-[Sha]	TYLCV-[SD]	1.44E-11	1,235	2,677	1,443
TYLCSV-[Sic]	TYLCMaIV	8.53E-11	301	1,111	811
ToLCSDV-[Gez]	TYLCV	1.31E-10	1,138	1,987	850
ToLCSDV-[Gez]	TYLCV	2.77E-10	72	544	473
ToLCSDV-[Sha]	TYLCV-[IR]	2.90E-10	1,297	1,963	667
ToLCSDV-[Sha]	TYLCV	2.51E-08	1,332	1,963	632
ToLCSDV-[Gez]	TYLCV-[SD]	2.97E-07	95	550	456
TYLCSV	TYLCSV-[Sic]	1.11E-06	2,780	2,413	2,457
TYLCSV	TYLCMaIV	1.31E-06	301	1,111	811
ToLCSDV-[Gez]	TYLCV-Mld	8.29E-06	72	544	473
TYLCSV	TYLCSV-[ES1]	4.70E-05	1,962	2,413	452
ToLCSDV-[Gez]	ToLCSDV-[Sha]	6.56E-05	477	1,127	651
TYLCMaIV	TYLCV	2.14E-03	2,454	2,572	119
ToLCSDV-[Sha]	TYLCSV-[ES1]	9.29E-03	2,615	2,698	84
TYLCSV-[ES1]	TYLCSV-[Sic]	2.48E-02	2,184	2,710	527

^a The first two columns indicate the first and second viruses of a pair: *Tomato yellow leaf curl Sardinia virus* (TYLCSV), TYLCSV from Spain (TYLCSV-[ES1]), TYLCSV-[ES2], *Tomato yellow leaf curl Malaga virus* from Spain (TYLCMaIV), TYLCV from Iran (TYLCV-[IR]), *Tomato leaf curl Sudan virus* (ToLCSDV), Gezira isolate (ToLCSDV-[Gez]), Shambat isolate (ToLCSDV-[Sha]), TYLCSV from Sicily (TYLCSV-[Sic]), TYLCV-mild from Israel (TYLCV-Mld), and TYLCV from Sudan (TYLCV-[SD]). The third column shows the *P* value of the recombinant fragments, the fourth and fifth columns indicate the location of the beginning and the end of the recombination (in the multiple alignment), and the last column indicates the length of the recombinant fragment in nucleotides.

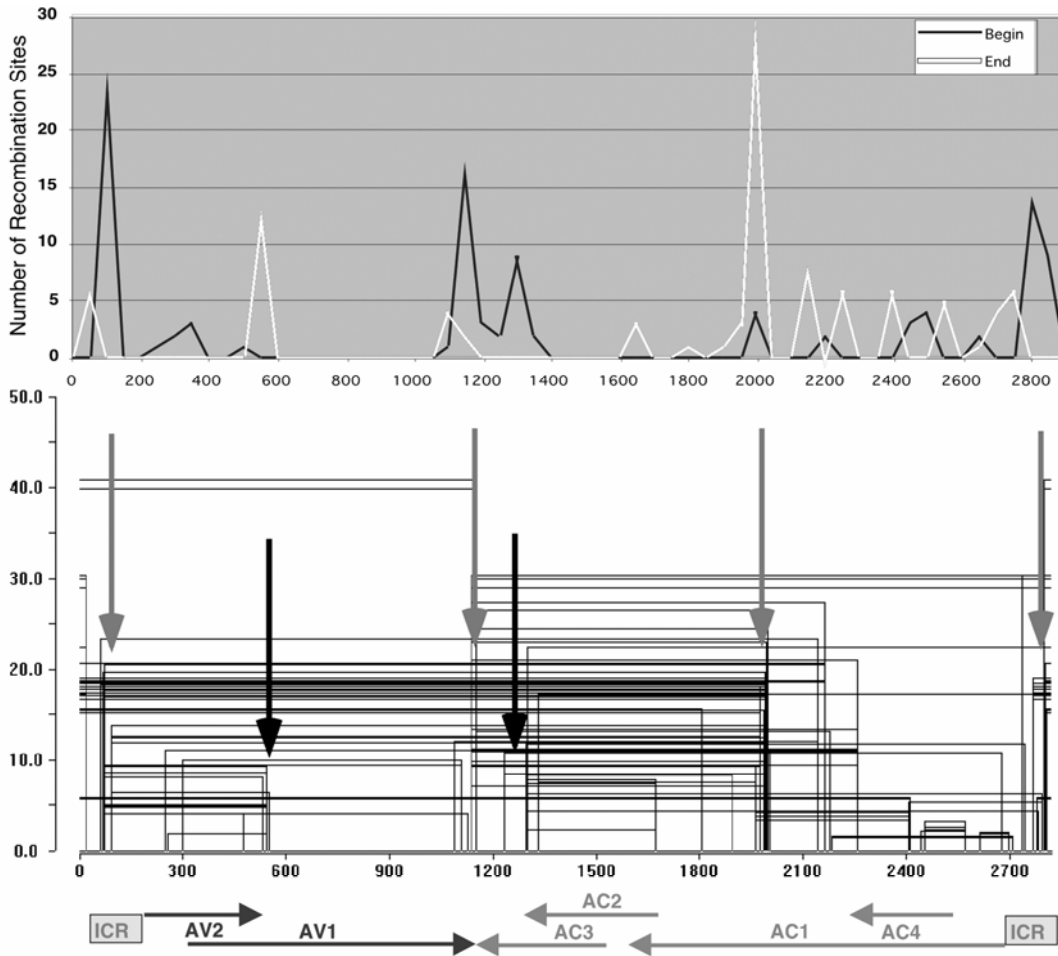


Fig. 4. Diagram representing the **A**, position and **B**, P value of the recombinant fragments along the virus genome. The top of the diagram (**A**) represents the number of recombination sites along the virus genome. The bottom of the diagram (**B**) represents the P value (y axis) and the position of the recombinant fragments (squares). The virus genome organization is represented under the diagram, positioning the different viral genes named according to the begomovirus convention.

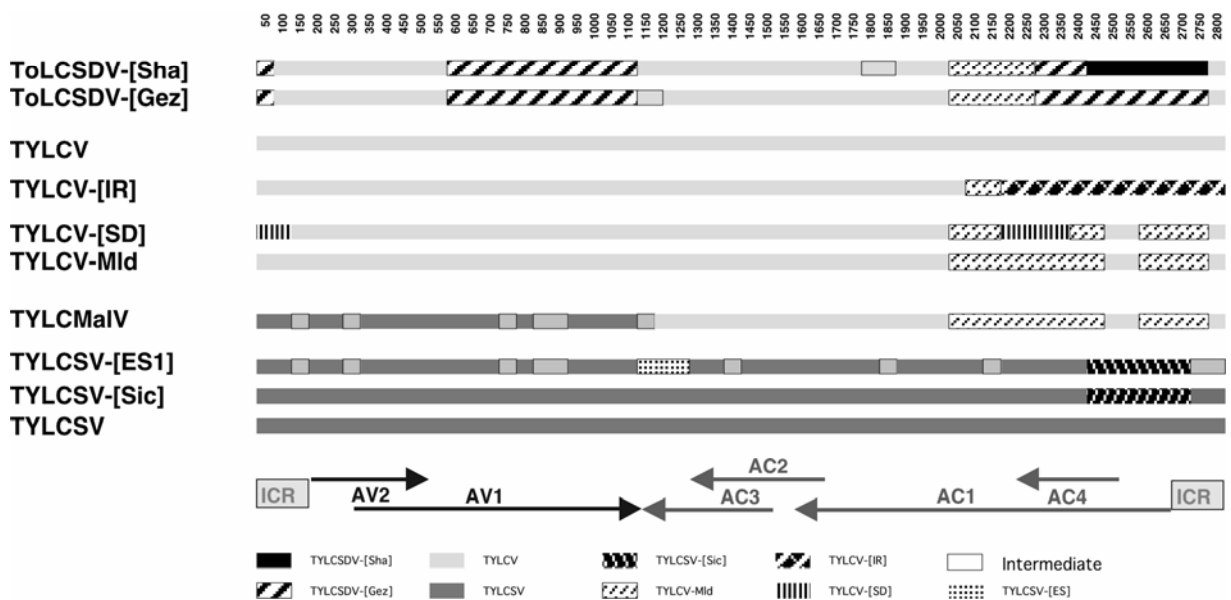


Fig. 5. Diagram representing the recombination map of the recombinant fragments for the 10 tomato begomoviruses from the Nile and Mediterranean Basins. Only fragments longer than 100 nts were considered with the exception of the *Tomato yellow leaf curl Sardinia virus* (TYLCSV) hybrid *Tomato yellow leaf curl Malaga virus* from Spain (TYLCMaIV). Each color/pattern represents a sequence specific of a virus. The virus genome organization is represented under the diagram, positioning the different viral genes named according to the begomovirus convention.

positioned in all viral genomes (Fig. 6, center panel). The differences that do not support such an observation are seen for several fragments that occur in highly variable zones of the begomoviral genomes (AC1-IR domain). The GeneConv analysis failed to detect small recombinant fragments (<500 nts), probably because of their relatively high heterogeneity or divergence. It is interesting to note that the *P* values for the recombinant fragments of the TYLCMaV are extremely small, at 10^{-18} to 10^{-41} , confirming that this particular recombination event probably occurred relatively recently (23). Therefore, used together, the two methods provide complementary information that contributes to a better estimation of the frequency of recombination for each virus population or isolate that is considered.

The recombinant sites were not distributed randomly among the populations considered here, and notably are located in definitive regions of the genome (Fig. 6, top panel). Such a non-random distribution could have come about through one and/or two mechanisms. First, specific sequences may be more likely to promote or to be conducive to recombination. Second, it is possible that only certain exchanges are contextually tolerated in begomoviral genomes, in general. For example, the AC2 and AC3 ORFs have been shown to be relatively nonviral (species)-specific proteins (29) and can be functional in trans for other species, and consequently would seem to be more easily exchanged. These recombination sites were not located in zones of the genome that are among the most conserved, but instead were typically found in the more sequence variable regions of the genome (Fig. 6, top panel). The six most frequent sites of recombination were in zones ranging from 60 to 92% shared identity and occurred in the viruses examined at nucleotide positions 100, 550, 1150, 1300, 2000, and 2800 (Fig. 6, top panel).

The recently described tomato begomoviruses from Sudan, TYLCV-[SD], ToLCSDV-[Gez], ToLCSDV-[Sha] (18), exhibited

a substantial number of recombinant fragments, some of them with very low *P* values, which may possibly indicate that recombination occurred relatively recently. From this study, it is clear that TYLCV-[SD] is a strain of TYLCV or TYLCV-Mid, while ToLCSDV-[Gez] and ToLCSDV-[Sha] are both strains of a newly described species. Both viruses share common sequences with TYLCV-Mid, but they also have specific sequences not present in any other extant virus (of which all known to date are represented here) (Fig. 6, center panel).

These results illustrate the difficulty in demarcating begomovirus species, particularly when several or numerous recombination events have occurred, and also underscore the potential for confounding phylogenetic analysis owing to homoplasy. The phylogenetic tree reconstructed with the multiple genome sequence alignment (Fig. 6, left panel) reflects both the similarity of the sequences and the percentage of shared sequences occurring through recombination. It is clear that there are three clusters of sequences, the TYLCV-like, the TYLCSV-like, and the hybrid TYLCMaV-like genomes. Within the TYLCV cluster, ToLCSDV strains can be differentiated from the TYLCV strains, and within the latter sequences there are three types of sequences (the TYLCV-like, the TYLCV-[IR]-like, and the TYLCV-Mid-like sequences) that can be differentiated.

Until now, the ICTV classification has only differentiated viral species according to the nomenclature used herein, but the newly proposed guidelines for begomovirus species demarcation from the ICTV Study Group would call for 89% shared nucleotide identity between the DNA-A component or monopartite genome as a threshold, and to date, TYLCV-[IR] and TYLCV-[SD] would be the only exceptions to that rule. Because there is as of yet no correlation between a specific recombinant fragment and a particular biological feature, the same weight can be applied to all of the genome zones. This permits the estimation of a general per-

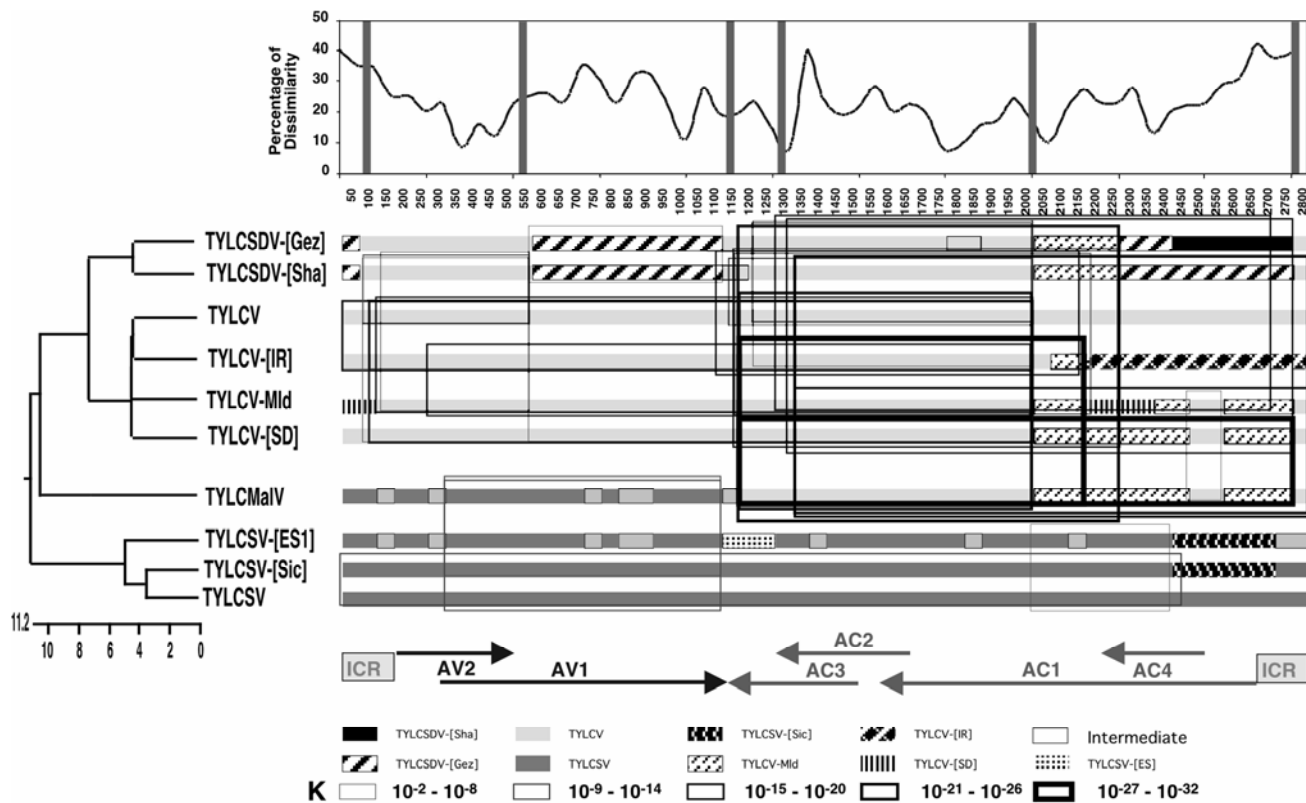


Fig. 6. Diagram representing the phylogenetic tree (left), the sequence consensus along the virus genome (top), the recombination map of the recombinant fragments (center) with the *P* values of the recombination recombinatorial fragments (center), and the genome organization (bottom) for 10 tomato begomoviruses from the Nile and Mediterranean Basins. Each color/pattern represents a sequence specific of a virus (legend at the bottom). The virus genome organization is represented under the diagram, positioning the different viral genes named according to the geminivirus convention.

centage of identity, and together with the frequency of recombination, to be considered for classification purposes, all other biological criteria being similar.

Tomato-infecting begomoviruses, with more than 23 recognized species worldwide (GeminiNet: available online from the Danforth Center), are among the most numerous within the genus and they all cause similar symptoms, albeit with mostly minor differences. It is therefore impossible to identify and classify these viruses based on their phenotypes. Certainly, DNA sequencing and recombination analyses are very powerful and useful tools for that task and permit a concise description of the molecular diversity of these viruses.

Recombination is now known to occur widely among begomoviruses (16,24,25,27); however, to date, begomoviruses that infect tomato contain evidence of the greatest number of recombination events when compared with begomoviruses, overall (24). This observation may be a function of the abundance of sequences available for tomato-infecting viruses compared with the number of available sequences for viruses associated with other plant species.

LITERATURE CITED

- Altschul, S. F. 1993. A protein alignment scoring system sensitive at all evolutionary distances. *J. Mol. Evol.* 36:290-300.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
- Bisaro, D. M. 1994. Recombination in geminiviruses: Mechanisms for maintaining genome size and generating genomic diversity. Pages 39-60 in: *Homologous Recombination and Gene Silencing in Plants*. J. Paszkowski, ed. Kluwer Academic, Dordrecht, the Netherlands.
- Brown, J. K. 2000. Molecular markers for the identification and global tracking of whitefly vector-begomovirus complexes. *Virus Res.* 71:233-260.
- Brown, J. K. 2001. The molecular epidemiology of begomoviruses. Pages 279-316 in: *Trends in Plant Virology*. J. A. Khan and J. Dykstra, eds. The Haworth Press, New York.
- Brown, J. K., Ostrow, K. M., Idris, A. M., and Stenger, D. C. 1999. Biotic, molecular, and phylogenetic characterization of bean calico mosaic virus, a distinct *Begomovirus* species with affiliation in the Squash leaf curl virus cluster. *Phytopathology* 89:273-280.
- Deng, D., Otim-Nape, W. G., Sangare, A., Ogwal, S., Beachy, R. N., and Fauquet, C. M. 1997. Presence of a new virus closely related to East African cassava mosaic geminivirus, associated with cassava mosaic outbreak in Uganda. *African J. Root Tuber Crops* 2:23-28.
- Fauquet, C. M., Bisaro, D. M., Briddon, R. W., Brown, J. K., Harrison, B. D., Rybicki, E. P., Stenger, D. C., and Stanley, J. 2003. Revision of taxonomic criteria for species demarcation in the family *Geminiviridae*, and an updated list of begomovirus species. *Arch. Virol.* 14:405-421.
- Fauquet, C. M., and Stanley, J. 2003. Geminivirus classification and nomenclature: Progress and problems. *Ann. Appl. Biol.* 142:165-189.
- Fondong, V. N., Pita, J. S., Rey, M. E. C., de Kochko, A., Beachy, R. N., and Fauquet, C. M. 2000. Evidence of synergism between *African cassava mosaic virus* and a new double-recombinant geminivirus infecting cassava in Cameroon. *J. Gen. Virol.* 81:287-297.
- Frischmuth, T., Engel, M., Lauster, S., and Jeske, H. 1997. Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted, *Sida*-infecting bipartite geminiviruses in Central America. *J. Gen. Virol.* 78:2675-2682.
- Harrison, B. D., Liu, Y. L., Khalid, S., Hameed, S., Otim-Nape, G. W., and Robinson, D. J. 1997. Detection and relationships of cotton leaf curl virus and allied whitefly-transmitted geminiviruses occurring in Pakistan. *Ann. Appl. Biol.* 130:61-75.
- Higgins, D. G., and Sharp, P. M. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *Comput. Appl. Biosci.* 5:151-153.
- Hou, Y.-M., and Gilbertson, R. L. 1996. Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. *J. Virol.* 70:5430-5436.
- Idris, A. M., and Brown, J. K. 2002. Molecular analysis of *Cotton leaf curl virus*-Sudan reveals an evolutionary history of recombination. *Virus Genes* 24:249-256.
- Idris, A. M., and Brown, J. K. 2004. *Cotton leaf crumple virus* is a distinct Western Hemisphere begomovirus species with complex evolutionary relationships indicative of recombination and reassortment. *Phytopathology* 94:1068-1074.
- Idris, A. M., and Brown, J. K. 2005. Evidence for interspecies-recombination for three begomovirus genomes associated with tomato leaf curl disease from central Sudan. *Arch. Virol.* (In press.)
- Idris, A. M., Hiebert, E., Bird, J., and Brown, J. K. 2003. Two newly described begomoviruses of *Macropitium lathyroides* and common bean. *Phytopathology* 93:774-783.
- Karlin, S., and Altschul, S. F. 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA* 87:2264-2268.
- Karlin, S., and Altschul, S. F. 1993. Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc. Natl. Acad. Sci. USA* 90:5873-5877.
- Miller, R. G. 1981. *Simultaneous Statistical Inference*. 2nd ed. Springer-Verlag, New York.
- Monci, F., Sanchez-Campos, S., Navas-Castillo, J., and Moriones, E. 2002. A natural recombinant between the geminiviruses *Tomato yellow leaf curl Sardinia virus* and *Tomato yellow leaf curl virus* exhibit a novel pathogenic phenotype and is becoming prevalent in Spanish populations. *Virology* 303:317-326.
- Padidam, M., Sawyer, S., and Fauquet, C. M. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* 265:218-225.
- Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S., and Fauquet, C. M. 2001. Recombination, pseudorecombination and synergism of geminiviruses are determinants keys to the epidemic of severe cassava mosaic disease in Uganda. *J. Gen. Virol.* 82:655-665.
- Saito, N., and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-425.
- Saunders, K., Bedford, I. D., and Stanley, J. 2001. Pathogenicity of a natural recombinant associated with *Ageratum* yellow vein disease: Implications for geminivirus evolution disease etiology. *Virology* 282:38-47.
- Sawyer, S. 1989. Statistical tests for detecting gene conversions. *Mol. Biol. Evol.* 6:526-538.
- Sunter, G., Stenger, D. C., and Bisaro, D. M. 1994. Heterologous complementation by geminivirus AL2 and AL3 genes. *Virology* 203:203-210.
- Swofford, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, MA.
- Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G. W., Robinson, D. J., and Harrison, B. D. 1997. Evidence that DNA A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *J. Gen. Virol.* 78:2101-2111.