

## Possible Emergence of New Geminiviruses by Frequent Recombination

Malla Padidam,\* Stanley Sawyer,† and Claude M. Fauquet‡<sup>1</sup>

\*Rohm and Haas Company, 727 Norristown Road, Spring House, Pennsylvania 19477; †Department of Mathematics and Department of Genetics, Washington University, St. Louis, Missouri 63130; and ‡ILTAB/Danforth Plant Science Center, University of Missouri St Louis, 8001 Natural Bridge Rd., St. Louis, Missouri 63121-4499

Received June 15, 1999; returned to author for revision July 8, 1999; accepted October 15, 1999

Although exchange of genetic information by recombination plays a role in the evolution of viruses, the extent to which it generates diversity is not clear. We analyzed genomes of geminiviruses for recombination using a new statistical procedure developed to detect gene conversions. Geminiviruses (family, Geminiviridae) are a group of plant viruses characterized by a genome of circular single-stranded DNA (~2700 nucleotides in length) encapsidated in twinned quasi-isometric particles. Complete nucleotide sequences of geminiviruses were aligned, and recombination events were detected by searching pairs of viruses for sequences that are significantly more similar than expected based on random distribution of polymorphic sites. The analyses revealed that recombination is very frequent and occurs between species and within and across genera. Tests identified 420 statistically significant recombinant fragments distributed across the genome. The results suggest that recombination is a significant contributor to geminivirus evolution. The high rate of recombination may be contributing to the recent emergence of new geminivirus diseases. © 1999 Academic Press

### INTRODUCTION

Diversity in viruses is generated by mutation, recombination, reassortment, and de novo gene acquisition. The relative contribution of these processes to genetic variation within species, within genera, or within families varies. Recombination can provide selective advantage in evolution of viruses within species, genera, and family levels (Keese and Gibbs, 1993; Morse, 1994; Gibbs *et al.*, 1995; Holland, 1998). While homologous recombination between two defective strains may produce a viable virus, non-homologous recombination facilitates rearrangement of viral genes and insertion of cellular or other viral genes. Recombination has been documented to occur in animal and plant viruses with RNA and DNA genomes (Lai, 1992; Chenault and Melcher, 1994; Simon and Bujarski, 1994; Gibbs *et al.*, 1995; Bujarski, 1996; Revers *et al.*, 1996; Burke, 1997; Roossinck, 1997). However, point mutations and small insertions or deletions are considered to be the main source of diversity. Such mutations occur during replication of RNA viruses due to the high error rates in RNA synthesis (Drake, 1993; Domingo and Holland, 1994). Published reports on recombination in DNA viruses do not indicate that recombination is more significant in the evolution of DNA viruses than of RNA viruses. We and others have recently identified recombinant geminiviruses that infect cassava in Africa (Deng *et al.*, 1997; Zhou *et al.*, 1997), cotton in

Pakistan (Zhou *et al.*, 1998; Bashir *et al.*, manuscript in preparation), pepper in Mexico (Torres-Pacheco *et al.*, 1993), and tomato in central America (Umaharan *et al.*, 1998). These observations on recombination in geminiviruses prompted us to study the frequency and role of recombination in geminivirus evolution and emergence.

Geminiviruses (family, Geminiviridae) are plant pathogenic viruses characterized by twinned quasi-isometric (geminete) particles, encapsidating a genome of circular single-stranded DNA. They are grouped into three genera based on insect vector, host range, and genome organization (Van Regenmortel *et al.*, 1997). Members of the genus *Begomovirus* are transmitted by whiteflies, have single or bipartite component genomes, and infect dicotyledonous plants. The genus *Curtovirus* comprises viruses with monopartite genomes that are transmitted by leafhoppers or treehoppers to dicotyledonous plants. Members of the genus *Mastrevirus* are transmitted by leafhoppers and have single-component genomes, and most infect monocotyledonous plants. Genome components of geminiviruses are 2500–3100 nucleotides (nts) in length and encode functions required for viral replication, encapsidation, and local and systemic movement in their hosts (Lazarowitz, 1992; Timmermans *et al.*, 1994). In viruses with two (A and B) component genomes, the B component encodes functions required for viral movement.

Here we provide evidence that interspecies recombination has resulted in significant diversity among geminiviruses. This recombination could play a major role in the recent emergence of new geminiviral diseases.

<sup>1</sup>To whom reprint requests should be addressed. Fax: (314) 516-4582. E-mail: ILTAB@danforthcenter.org.

## RESULTS AND DISCUSSION

Complete DNA sequences are now available for 95 geminiviruses that represent 64 distinct virus species and strains. Sequences that were >95% identical in nt sequence to these 64 viruses were not included in the analysis. A phylogenetic tree obtained from the alignment of complete DNA sequences is shown in Fig. 1. The tree has distinct branches for each of the three genera with the exception of Tomato pseudo curly top virus (TPCTV, *Curtovirus*), which forms a separate branch (a proposal for a fourth genus is now being accepted by ICTV). The members of the genus *Begomovirus* form clusters in the tree according to their geographical origin with distinct branches for viruses from the Americas, Asia, and Africa. However, the relative positions of viruses within a group often vary when different parts of the genome (i.e., genes, parts of genes, or blocks of 100 nts) were used as the basis for analysis (data not shown). This observation suggested to us the possibility of significant recombination among geminiviruses.

To search for recombination events among geminiviruses, we used a statistical technique for detecting gene conversion based on a new program (GENECONV; see Materials and Methods). Recombined genome segments containing relatively few mutations may be detectable in this way. Our procedure differs from other score-based methods in that we only consider homologous segments (which increases the statistical power for detecting homologous segments) and we do not introduce further gaps or indels. Other techniques for finding recombination events include the use of probabilities of runs (Stephens, 1985; Takahata, 1994), inconsistencies in estimated phylogenies (Hudson and Kaplan, 1985; DuBose *et al.*, 1988; Hein, 1993; Grassly and Holmes, 1997; McGuire *et al.*, 1997; Maynard Smith and Smith, 1998), and other techniques (Jakobson *et al.*, 1977; Maynard Smith, 1992; Weiller, 1998). A more detailed discussion of the GENECONV procedure will appear elsewhere.

Tests for recombination among viruses within *Begomovirus* groups of Asia, Africa, and Americas, *Curtovirus*, and *Mastrevirus* identified 129 statistically significant recombinant fragments distributed across the genome (Figs. 2A–2C, 2F, and 2H). The lengths of the recombinant fragments varied from 32 to 2391 nts, and recombinations appeared to occur throughout the entire genome. Among the whitefly-transmitted geminiviruses (*Begomovirus*), more examples of putative recombination were observed among viruses isolated from Asia (16 viruses, 49 recombinations) than among viruses from Africa (12 viruses, 13 recombinations) or the Americas (19 viruses, 36 recombinations). Twelve examples of recombination were detected among the 5 *Curtovirus* species, whereas 15 recombinations were detected among the 12 *Mastrevirus*. Possible recombinant fragments observed are

shorter (35–165 nts) among the *Mastrevirus* than in the *Begomovirus* and *Curtovirus*.

Some representative recombinant fragments are summarized in Table 1. The presence of recombinations is illustrated here using several examples. In the cases of East African cassava mosaic virus from Cameroon (EACMV-CM) and EACMV from Tanzania (EACMV-TZ), approximately two-thirds of the genome are similar to each other and the rest of the genome is different from each other. Similarly, Tomato yellow leaf curl viruses from Sardinia (TYLCV-Sar) and Sicily (TYLCV-Sic) are recombinants throughout most of their genomes (Table 1). Other recombinant pairs include Tobacco leaf curl virus from China (TbLCV-CN) and TYLCV from Thailand (TYLCV-TH), Potato yellow mosaic virus from Venezuela (PYMV-VE), and Tomato leaf curl virus from Panama (ToLCV-PA), and Beet curly top viruses from USA (BCTV-Cal, BCTV-Cfh) (Table 1).

The homoplasy test of Maynard Smith and Smith (1998) was applied to five geminivirus alignments as a second procedure to test for recombination: (1) Asian *Begomovirus* ( $n = 16$ ), (2) African *Begomovirus* ( $n = 12$ ), (3) New World *Begomovirus* ( $n = 19$ ), (4) *Mastrevirus* ( $n = 12$ ), and (5) *Curtovirus* ( $n = 4$ , excluding TPCTV; see Fig. 1). Sites with indels or missing values were excluded from the analysis.  $P$  values in this test are based on a proportion of simulations with an equal or larger number of homoplasies, which requires estimates of both the observed number of homoplasies and of the effective number of mutable sites ( $S_e$ ). Maynard Smith and Smith recommend the use of an estimate of  $S_e$  using an outgroup sequence, providing that this estimate lies between 60 and 100% of the number of sites. Otherwise, they recommend using either 60 or 100% of the number of sites. The *Curtovirus* BCTV-Cfh was used as the outgroup except for the *Curtovirus* alignment, for which the *Mastrevirus* MiSV was used (Fig. 1). Except for the *Curtovirus*, the estimated  $S_e$  lay between 60 and 100% of the true number of sites. Maynard Smith and Smith also suggest that their procedure be used only if in-group divergences average  $\leq 5\%$ . Average ingroup divergences varied between 20 and 40% for the geminivirus alignments, but the behavior of the  $S_e$  estimates suggests that the test is still valid.

The homoplasy test detected recombination in the three *Begomovirus* alignments, consistent with Table 1 ( $P = 0.0000$  in all three cases, based on 10,000 simulations). Recombination was not detected for the *Mastrevirus* ( $P = 0.9986$ ) nor for the four *Curtovirus* ( $P = 1.000$ ). The result for the *Mastrevirus* is consistent with the less-significant  $P$  values for the *Mastrevirus* fragments in Table 1 and a possible lower power of the homoplasy test due to the relatively high between-sequence divergences. The lack of significance for the four *Curtovirus* may be due to the small number of sequences in the alignment. Maynard Smith and Smith

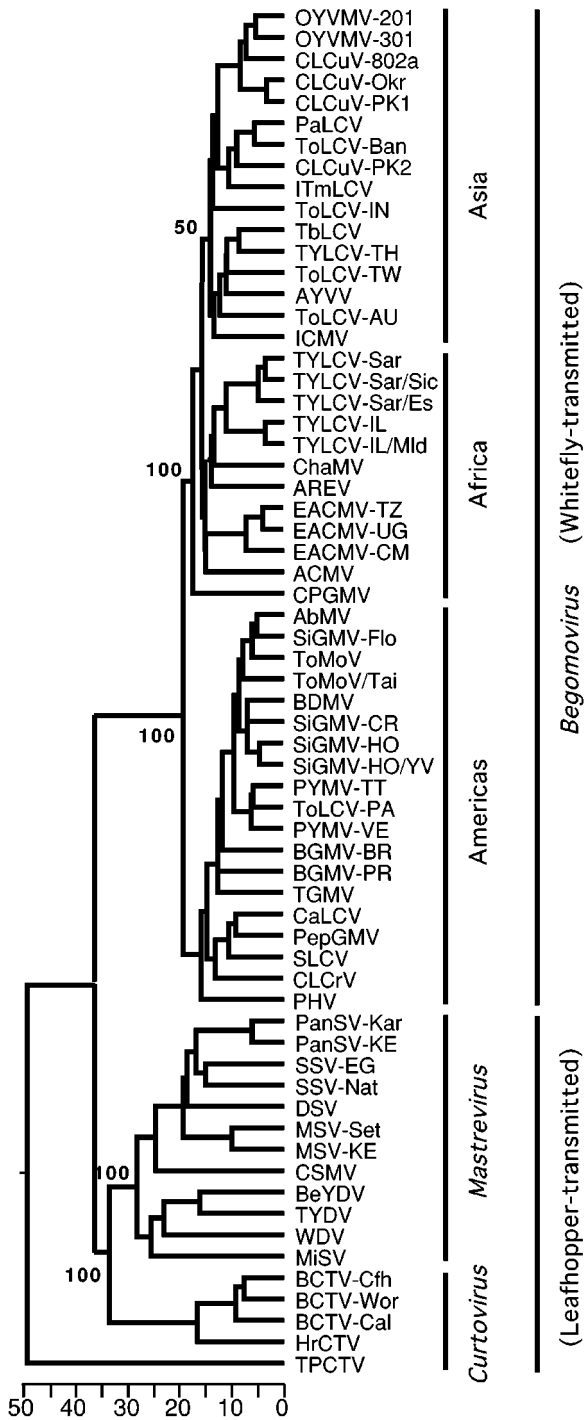


FIG. 1. Phylogenetic tree of 64 geminivirus DNA sequences. Complete DNA sequences of monopartite geminiviruses and the A component of bipartite geminiviruses were aligned. The tree was generated using the MegAlign program available with the DNASTAR package. The three genera (begomoviruses, curtoviruses, and mastreviruses), and New World begomoviruses within all begomoviruses, were supported in 1000 of 1000 bootstrapped trees. TPCTV was excluded from the curtoviruses for the bootstrap analysis. Asian begomoviruses were monophyletic within Old World begomoviruses in ~50% of bootstrapped trees. Vertical distances are arbitrary, and the scale below the tree measures the distance between sequences. The tree is unrooted. The virus names and GenBank Accession Numbers are available on our laboratory web site (<http://www.danforthcenter.org/fauquet/iltab/index.htm>). Viruses isolated in the Mediterranean region are also referred to as viruses from Africa.

suggest that their procedure be used only for alignments with at least six sequences.

Unlike the recombinations among viruses within geographically defined groups, the apparent recombinations between viruses from Asia and Africa, between the Americas and the Old World (Asia+Africa), and between begomoviruses and curtoviruses were localized to the N-terminal region of the replication associated protein (Rep) (Figs. 2D, 2E, and 2G). These putative recombination events between viruses from different groups probably represent older events as they presumably occurred before the geographical isolation. The high number of recombination events (total of 391) between members of different groups appeared to be due to less divergence in the 5' part of the Rep gene after earlier recombinations. No examples of recombination were detected between curtoviruses and mastreviruses or between begomoviruses and mastreviruses. The curtoviruses have a hybrid genome; the coat protein gene is similar to mastreviruses and the rest of the genome is similar to begomoviruses. It was proposed that the curtoviruses may have evolved as leafhopper-transmitted viruses after a whitefly-transmitted virus (*Begomovirus*) acquired the coat protein sequence from a leafhopper-transmitted virus (*Mastrevirus*) by recombination (Stanley *et al.*, 1986). Subsequently the curtoviruses evolved as a distinct group. These viruses did not show up as recombinants between mastreviruses and begomoviruses in our tests possibly because the initial recombination event was very old.

Most begomoviruses have two (A and B) component genomes. The two components have little sequence homology except for ~170 nts sequence in the intergenic region that is termed the common region (CR). The CR contains an origin of replication, where the Rep protein (encoded on the A component) binds to initiate replication of each component. When B components of 6 viruses from the Old World and 20 viruses from the Americas were analyzed for recombination, two recombination events were detected within the Old World viruses (data not shown), 10 recombinations were detected within the American viruses (Fig. 2I), whereas none were detected between the viruses from the Old World and the Americas.

Establishment of a new recombinant A component of a bipartite virus involving the CR will require a subsequent B component recombination. A B component with the same origin of replication as the recombinant A component can be created by recombination between recombinant A component and another B component in the intergenic region. We observed that A components of PYMV from Venezuela, PYMV from Trinidad and Tobago, and ToLCV from Panama are recombinants that differ only for a portion of the Rep gene and the CR sequence. Their B components are very similar to each other except

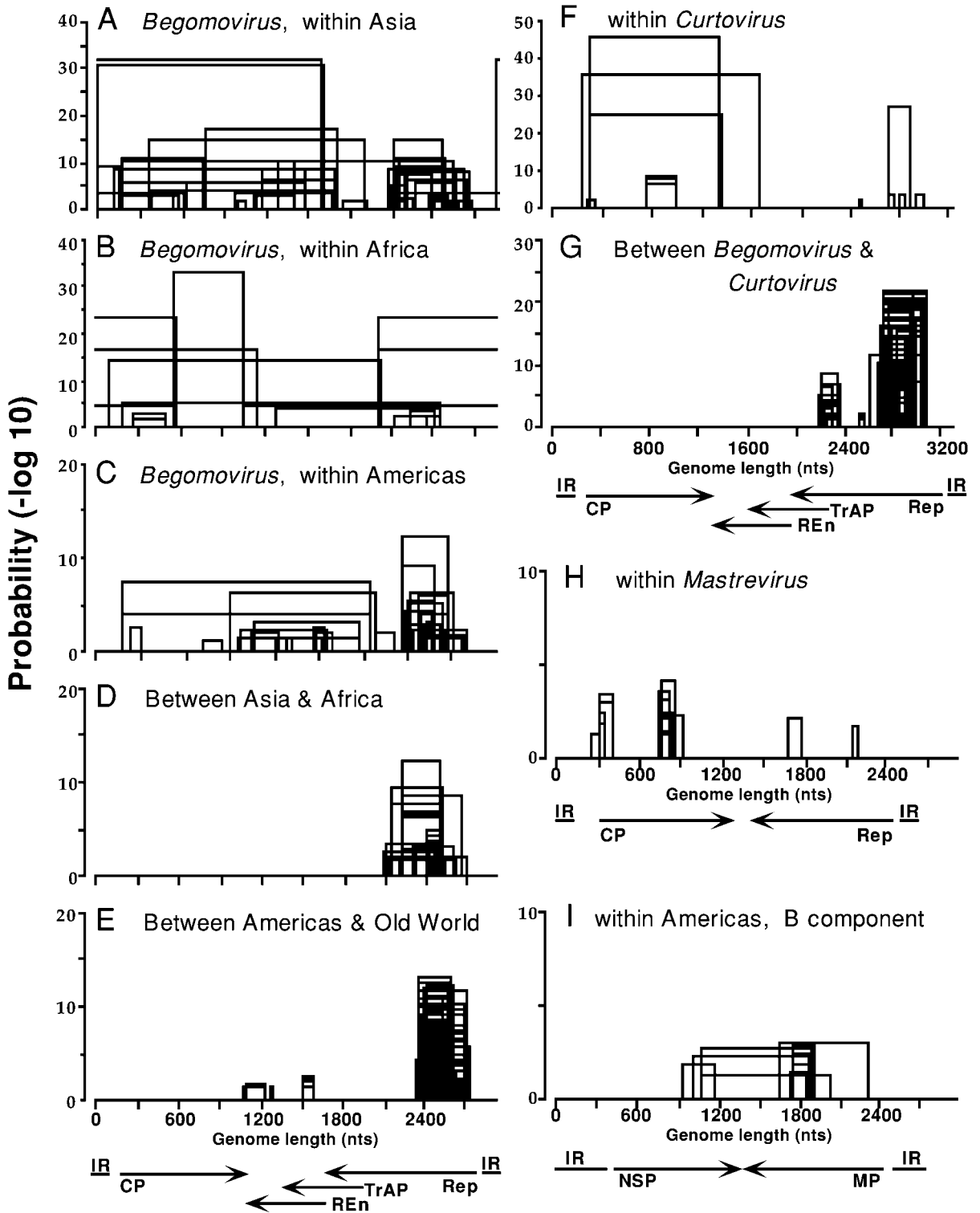


FIG. 2. Graphic representation of recombination in geminiviruses. For all parts in the figure, the beginning and the end of a fragment that underwent recombination are connected by a line to form an open rectangle. The height of the open rectangle is proportional to  $-\log_{10}$  of the probability. The probability values shown on the y axis are corrected for multiple comparisons. Only the fragments with  $P$  values that are  $\leq 0.05$  ( $-\log_{10} \leq 1.301$ ) are shown. The x axis shows genome length (in nucleotide numbers of consensus sequence), and the first nucleotide T in the invariable nonnucleotide TAATATTAC present in all geminiviruses is set as nucleotide #1. The approximate locations of genes coding for the coat protein (CP), replication associated protein (Rep), transactivator protein (TrAP), replication enhancer (REn), nuclear shuttle protein (NSP), and movement protein (MP) are shown below the graphs. IR, intergenic region. (A–I) Recombinant fragments among viruses from Asia (A), among viruses from Africa (B), among viruses from Americas (C), between viruses from Asia and Africa (D), between viruses from Americas and Old World (E), among viruses of genus *Curtovirus* (F), between viruses of genera *Begomovirus* and *Curtovirus* (G), among viruses of genus *Mastrevirus* (H), and among B components of viruses from Americas (I).

TABLE 1  
Partial List of Recombinant Fragments Identified among Geminiviruses

Virus pair	Fragment length (nts)	Beginning nucleotide # in		Probability
		Virus 1	Virus 2	
Within <i>Begomovirus</i>				
Within Asia				
CLCuV-802a;OYVMV-201	1494	2732	2727	$7.30 \times 10^{-33}$
ITmLCV;ToLCV-Ban	933	686	689	$9.70 \times 10^{-18}$
TbLCV;TYLCV-TH	1195	1190	1191	$2.53 \times 10^{-11}$
AYV;ICMV	125	2286	2366	$1.53 \times 10^{-3}$
Within Africa				
EACMV-CM;EACMV-TZ	1949	1973	1972	$1.71 \times 10^{-17}$
TYLCV-IL;TYLCV-IL,Mid 1916	67	67		$7.95 \times 10^{-15}$
TYLCV-Sar;TYLCV-Sic	2241	143	143	$6.90 \times 10^{-6}$
ChaMV;TYLCV-IL	224	254	224	$1.94 \times 10^{-3}$
Within Americas				
BDMV;PHV	326	1998	2023	$6.15 \times 10^{-13}$
PYMV-VE;ToLCV-PA	1625	133	134	$2.95 \times 10^{-8}$
CLCrV;PYMV-TT	690	1025	997	$5.89 \times 10^{-4}$
SiGMV-HN;SiGMV-Yv	556	941	949	$3.73 \times 10^{-3}$
Within Americas, B component				
ToMoV;ToMoV-Tai	664	1341	1372	$8.99 \times 10^{-4}$
SiGMV-Yv;PYMV-TT	164	1459	1474	$1.25 \times 10^{-3}$
SiGMV-HN;SiGMV-CR	720	879	905	$5.07 \times 10^{-3}$
ToMoV;PYMV-VE	120	1403	1442	$3.55 \times 10^{-2}$
Between Asia and Africa				
AREV;OYVMV-301	299	298	2080	$4.96 \times 10^{-13}$
AREV;ToLCV-AU	316	2098	2097	$1.60 \times 10^{-7}$
ICMV;TYLCV-Sar	104	2366	2292	$1.45 \times 10^{-5}$
ITmLCV;TYLCV-IL	200	2194	2213	$4.17 \times 10^{-4}$
Between New World and Old World				
CLCuV-Ok;SiGMV-Yv	221	2218	2109	$6.74 \times 10^{-14}$
BDMV;TYLCV-SP	191	2154	2297	$5.29 \times 10^{-13}$
AbMV;ICMV	176	2160	2360	$7.39 \times 10^{-12}$
AYV;PYMV-TT	161	2283	2132	$4.43 \times 10^{-8}$
Within <i>Curtovirus</i>				
BCTV-Cal;BCTV-Wor	1042	290	243	$2.04 \times 10^{-46}$
BCTV-Cal;BCTV-Cfh	1395	227	196	$3.51 \times 10^{-36}$
BCTV-Cal;TPCTV	176	2503	2368	$5.94 \times 10^{-28}$
BCTV-Wor;HrCTV	236	705	625	$1.07 \times 10^{-9}$
Within <i>Mastrevirus</i>				
PanSV-Kar;SSV-EG	92	720	704	$7.19 \times 10^{-5}$
MSV-KE;PanSV-KE	88	273	293	$3.28 \times 10^{-4}$
BeYDV;TYDV	106	1469	1502	$7.31 \times 10^{-3}$
DSV;MSV-Set	47	257	277	$1.36 \times 10^{-2}$
Between <i>Begomovirus</i> and <i>Curtovirus</i>				
ToMoV;TPCTV	359	2061	2372	$1.20 \times 10^{-22}$
ICMV;TPCTV	186	2359	2434	$1.59 \times 10^{-16}$
BCTV-Cal;BGMV-PR	211	2479	2071	$6.99 \times 10^{-10}$
HrCTV;SLCV	98	2593	2114	$1.70 \times 10^{-9}$

Note. Probabilities listed are multiple comparison-corrected values.

that the CRs are similar to their respective A components.

In addition to the natural recombination events described here, homologous and nonhomologous recombination events in geminiviruses have been frequently observed under experimental conditions. Bisaro (1994) listed the following: reversion of deletion mutants to wild-type genome size, deletion of foreign sequences

from geminivirus vectors, synthesis of wild-type recombinant from two mutants, synthesis of subgenomic defective DNA molecules, and release of infectious viral DNA from recombinant plasmids containing monomeric genome inserts. Evidence for the integration of geminivirus DNA into host plant nuclear genome has also been observed (Bejarano *et al.*, 1996). The results of studies on experimental recombination and our results on natural

recombination indicate that recombination is a significant strategy contributing to geminivirus evolution. Previous studies of other viruses also indicate the importance of recombination in viral evolution but the recombinations observed in these studies were predominantly between strains of a given virus species (Lai, 1992; Chenault and Melcher, 1994; Simon and Bujarski, 1994; Gibbs *et al.*, 1995; Bujarski, 1996; Revers *et al.*, 1996; Burke, 1997; Roossinck, 1997). For example, recombination is frequent in Human immunodeficiency virus and  $\leq 10\%$  of all isolates are recombinants (Fulz *et al.*, 1997). However, these mosaic genomes are generated by recombination between viruses of the same or different subtypes. Thus the extent of interspecies recombination observed in geminiviruses appears to be unique among viruses.

The extent of recombination among geminiviruses may be a fraction of what is present or possible in nature. More than 170 geminiviruses are known to infect plants (Bridson and Markham, 1995; Padidam *et al.*, 1997), and many more remain to be characterized, particularly from weed species, which can act as reservoirs of virus populations. Why do geminiviruses apparently have such a high propensity for recombination? Plant virus evolution is affected by various factors including virus-vector and virus-host interactions. Three factors could contribute significantly to recombination, namely mixed infections, high levels of viral replication, and increased host range of the vector. First, mixed infections, a prerequisite for recombination, are common in geminivirus diseases (Padidam *et al.*, 1995; Umaharan *et al.*, 1998). Second, geminiviruses replicate via a double-stranded replicative form and achieve very high copy numbers (Kanevski *et al.*, 1992; Accotto *et al.*, 1993). Third, with the emergence of the B biotype whitefly that can feed on hundreds of species (Bedford *et al.*, 1994), the host range of geminiviruses has expanded greatly. As a result, there is greater opportunity for a recombinant virus to emerge by infecting a new host.

The geographical separation that previously prevented recombinations between New and Old World viruses is now reduced due to modern agricultural practices. For example, Tomato yellow leaf curl virus from Israel (TYLCV-IL) was apparently imported into the Dominican Republic a few years ago along with tomato seedlings, and it has been reported in other Caribbean countries and in Florida (Polston *et al.*, 1994; Polston, 1998). As a result, there is great potential for recombination between TYLCV-IL and viruses from the Americas that may result in a virus with new pathological properties.

Has recombination in geminiviruses contributed to the recent emergence of geminiviruses? Cotton leaf curl disease in Pakistan (CLCuV) became severe during the past 6 years and is estimated to cause damage worth \$1.2 billion to cotton production in Pakistan (Hameed *et al.*, 1994; Harrison *et al.*, 1997; Bashir *et al.*, manuscript in

preparation). In Trinidad and Tobago, a geminivirus disease on tomato (PYMV-TT) observed in 1989 has recently spread throughout the country (Umaharan *et al.*, 1998). A new cassava mosaic virus (EACMV-UG) has devastated cassava production in Uganda (Deng *et al.*, 1997; Zhou *et al.*, 1997). Tomato production in Spain and Italy, both under field and greenhouse conditions, is severely constrained by TYLCV-Sar (Noris *et al.*, 1994). All of these viruses are recombinants. Virus disease emergence and spread is influenced by many factors including changing agricultural practices, but the hypothesis that recombination among geminiviruses is contributing to the increased emergence of geminiviruses is testable. Such studies will facilitate increased understanding of how viruses evolve in response to changes in the ecosystem.

## MATERIALS AND METHODS

The names and GenBank Accession Numbers of 64 geminivirus DNA sequences used in the analysis are available on our laboratory Web site (<http://www.danforthcenter.org/fauquet/iltab/index.htm>). The sequences were aligned using the Clustal V method of aligning multiple sequences (Higgins and Sharp, 1989) 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 generated using the neighbor-joining method (Saito and Nei, 1987) with the UPGMA distance matrix (MegAlign program). An identical consensus tree was also obtained when phylogenetic analysis was done by a cladistic parsimony method using the computer program PAUP (version 3.0, Illinois Natural History Survey, Champaign, IL). One thousand bootstrap replications were performed to place confidence estimates on major groups contained in the tree.

To detect recombination events among geminiviruses, we used a new program (GENECONV) based on an earlier statistical approach for detecting gene conversion (Sawyer, 1989). We use the term "fragment" to denote an aligned or homologous pair of segments in an alignment. The highest-scoring fragments in an alignment are listed and 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 (Altschul *et al.*, 1990; Karlin and Altschul, 1990). 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 (Sawyer, 1989). 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 (Altschul, 1993; Karlin and Altschul, 1993).

Given a *g* scale value, the program finds all nonoverlapping maximal fragments that have Bonferroni-corrected (Miller, 1981) *P* values of  $\leq 0.05$ . 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 (Miller, 1981). 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 all analyses. A program GENECONV that finds, sorts, and scores all significant fragments in an alignment can be found on Sawyer's Web site ([www.math.wustl.edu/~sawyer](http://www.math.wustl.edu/~sawyer)). GENECONV was applied to several geminivirus alignments by essentially entering the command line "geneconv <alignment\_file\_name>/c/n0/mkg0.05/g1/f" in a command-line window. Here the argument /c indicates circular DNA, /n0 means to suppress the calculation of permutation-test *P* values, /mkg0.05 means to list all fragments whose Bonferroni-corrected BLAST-like *P* value is  $\leq 0.05$ , /g1 means to set *g* scale = 1, and /f provides more detailed output. Each run of GENECONV writes a spreadsheet-like file in which the significant fragments are listed in decreasing order by Bonferroni-corrected *P* value. Documentation for using GENECONV can be found on Sawyer's Web site.

We also tested geminiviral alignments for recombination by applying the Homoplasmy Test of Maynard Smith and Smith (1998). This procedure constructs an approximate maximum parsimony pedigree and tests for an excess of homoplasies (reverse or parallel mutations). Such an excess would be suggestive of recombination. This yields a *P* value for recombination but gives no information about the location of the recombination events whose presence is inferred.

## ACKNOWLEDGMENTS

We thank R. N. Beachy, J. D. E. Holland, S. Lazarowitz, and M. Van Regenmortel for reviewing the manuscript. This work was supported by financial assistance from Maharashtra Hybrid Seeds Company, India (#5-98378), and ORSTOM, Paris.

## REFERENCES

- Accotto, G. P., Mullineaux, P. M., Brown, S. C., and Marie, D. (1993). Digitaria streak geminivirus replicative forms are abundant in S-phase nuclei of infected cells. *Virology* **195**, 257–259.
- 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.
- Bedford, I. D., Briddon, R. W., Brown, J. K., Rosell, R. C., and Markham, P. G. (1994). Geminivirus transmission and biological characterization of *Bemisia tabaci* (Gennadius) biotypes from different geographical regions. *Ann. Appl. Biol.* **125**, 311–325.
- Bejarano, E. R., Khashoggi, A., Witty, M., and Lichtenstein, C. (1996). Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. *Proc. Natl. Acad. Sci. USA* **93**, 759–764.
- Bisaro, D. M., Ed. (1994). Recombination in the geminiviruses: Mechanisms for maintaining genome size and generating genomic diversity. In "Homologous Recombination and Gene Silencing in Plants" (J. Paszkowski, Ed.), pp. 39–60. Kluwer, Boston.
- Briddon, R. W., and Markham, P. (1995). *Geminiviridae*. In "Virus Taxonomy, Sixth Report of International Committee on Taxonomy of Viruses" (F. A. Murphy, et al. Eds.). New York, Springer-Verlag.
- Bujarski, J. J. (1996). "Seminars in Virology issue devoted to RNA recombination." *Semin. Virol.* **7**(6).
- Burke, D. S. (1997). Recombination in HIV: An important viral evolutionary strategy. *Emerg. Infect. Dis.* **3**, 253–259.
- Chenault, K. D., and Melcher, U. (1994). Phylogenetic relationships reveal recombination among isolates of cauliflower mosaic virus. *J. Mol. Evol.* **39**, 496–505.
- 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.
- Domingo, D., and Holland, J. J., Eds. (1994). Mutation rates and rapid evolution of RNA viruses. In "The Evolutionary Biology of Viruses" (S. S. Morse, Ed.). Raven Press, New York.
- Drake, J. W. (1993). Rates of spontaneous mutations among RNA viruses. *Proc. Natl. Acad. Sci. USA* **90**, 4171–4175.
- DuBose, R., Dykhuizen, D. E., and Hartl, D. L. (1988). Genetic exchange among natural isolates of bacteria: recombination within the *phoA* locus of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **85**, 7036–7040.
- Fulz, P. N., Yue, L., Wei, Q., and Girard, M. (1997). Human immunodeficiency virus type 1 intersubtype (B/E) recombination in a superinfected chimpanzee. *J. Virol.* **71**, 7990–7995.
- Gibbs, A., Calisher, C. H., and Garcia-Arenal, F. (1995). "Molecular Basis of Viral Evolution." Cambridge University Press, Cambridge.
- Grassly, N. C., and Holmes, E. C. (1997). A likelihood method for the detection of selection and recombination using sequence data. *Mol. Biol. Evol.* **14**, 239–247.
- Hameed, S., Khalid, S., Ehsan Ul, H., and Hashmi, A.A. (1994). Cotton leaf curl disease in Pakistan caused by a whitefly transmitted geminivirus. *Plant Dis.* **78**, 529.
- 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.
- Hein, J. (1993). A heuristic method to reconstruct the history of sequences subject to recombination. *J. Mol. Evol.* **36**, 396–405.
- Higgins, D. G., and Sharp, P. M. (1989). Fast and sensitive multiple sequence alignments on a microcomputer. *Comput. Appl. Biosci.* **5**, 151–153.
- Holland, J. D. E. (1998). Origin and evolution of viruses. *Virus Genes* **16**, 13–21.
- Hudson, R. R., and Kaplan N. L. (1985). Statistical properties in the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.
- Jakobsen, I. B., Wilson S. R., and Easteal S. (1997). The partition matrix: exploring variable phylogenetic signals along nucleotide sequence alignments. *Mol. Biol. Evol.* **14**, 474–484.
- Kanevski, I. F., Thakur, S., Cosowsky, L., Sunter, G., Brough, C., Bisaro, D., and Maliga, P. (1992). Tobacco lines with high copy number of replicating recombinant geminivirus vectors after biolistic DNA delivery. *Plant J.* **2**, 457–463.
- 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.
- Keese, P., and Gibbs, A. (1993). Plant viruses: master explorers of evolutionary space. *Curr. Opin. Genet. Dev.* **3**, 873–877.
- Lai, M. M. C. (1992). RNA recombination in animal and plant viruses. *Microbiol. Rev.* **56**, 61–79.
- Lazarowitz, S. G. (1992). Geminiviruses: Genomes structure and gene function. *Critical Rev. Plant Sci.* **11**, 327–349.
- Maynard Smith, J. (1992). Analyzing the mosaic structure of genes. *J. Mol. Evol.* **34**, 126–129.
- Maynard Smith, J., and Smith, N. H. (1998). Detecting recombination from gene trees. *Mol. Biol. Evol.* **15**, 590–599.
- McGuire, G. F., Wright, F., and Prentice, M. J. (1997). A graphical method for detecting recombination in phylogenetic data sets. *Mol. Biol. Evol.* **14**, 1125–1131.
- Miller, R. G. (1981). "Simultaneous Statistical Inference," 2nd ed. Springer-Verlag, New York.
- Morse, S. S. (1994). "The Evolutionary Biology of Viruses." Raven Press, New York.
- Noris, E., Hidalgo, E., Accotto, G. P., and Moriones, E. (1994). High similarity among the tomato yellow leaf curl virus isolates from the West Mediterranean Basin: The nucleotide sequence of an infectious clone from Spain. *Arch. Virol.* **135**, 165–170.
- Padidam, M., Beachy, R. N., and Fauquet, C. M. (1995). Tomato leaf curl geminivirus from India has a bipartite genome and coat protein is not essential for infectivity. *J. Gen. Virol.* **76**, 25–35.
- Padidam, M., Maxwell, D. P., and Fauquet, C. M. (1997). A proposal for naming geminiviruses. *Arch. Virol.* **142**, 2553–2562.
- Polston, J. E., (1998). First report of a tomato yellow leaf curl-like geminivirus in Florida. International Symposium on Whiteflies and Geminiviruses, Puerto-Rico, June 3–5, 1998.
- Polston, J. E., Bois, D., Serra, C. A., and Concepcion, S. (1994). First report of a tomato yellow leaf curl-like geminivirus in the western hemisphere. *Plant Dis.* **78**, 831.
- Revers, F., Le Gall, O., Candresse, T., Le Romancer, M., and Dunez, J. (1996). Frequent occurrence of recombinant potyvirus isolates. *J. Gen. Virol.* **77**, 1953–1965.
- Roossinck, M. J. (1997). Mechanisms of plant virus evolution. *Annu. Rev. Phytopath.* **35**, 191–209.
- Saito, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Sawyer, S. (1989). Statistical tests for detecting gene concertions. *Mol. Biol. Evol.* **6**, 526–538.
- Simon, A. E., and Bujarski, J. J. (1994). RNA-RNA recombination and evolution in virus-infected plants. *Annu. Rev. Phytopathol.* **32**, 337–362.
- Stanley, J., Markham, P. G., Callis, R. J., and Pinner, M. S. (1986). The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. *EMBO J.* **5**, 1761–1767.
- Stephens, J. C. (1985). Statistical methods of DNA sequence analysis: Detection of intragenic recombination or gene conversion. *Mol. Biol. Evol.* **2**, 539–556.
- Takahata, N. (1994). Comments on the detection of reciprocal recombination or gene conversion. *Immunogenetics* **39**, 146–149.
- Timmermans, M. C. P., Das, O. P., and Messing, J. (1994). Geminiviruses and their uses as extrachromosomal replicons. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 79–112.
- Torres-Pacheco, I., Garzon, - T., J. A., Herrera-Estrella, L., and Rivera-Bustamante, R. F. (1993). Complete nucleotide sequence of pepper huasteco virus: Analysis and comparison with bipartite geminiviruses. *J. Gen. Virol.* **74**, 2225–2231.
- Umaharan, P., Padidam, M., Phelps, R. H., Beachy, R. N., and Fauquet, C. M. (1998). Distribution and diversity of geminiviruses in Trinidad and Tobago. *Phytopathology* **88**, 1262–1268.
- Van Regenmortel, M. H. V., Bishop, D. H. L., Fauquet, C. M., Mayo, M. A., Maniloff, J., and Calisher, C. H. (1997). Guidelines to the demarcation of virus species. *Arch. Virol.* **142**, 1505–1518.
- Weiller, G. F. (1998). Phylogenetic profiles: A graphical method for detecting recombinations in homologous sequences. *Mol. Biol. Evol.* **15**, 326–335.
- 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.
- Zhou, X., Liu, Y., Robinson, D. J., and Harrison, B. D. (1998). Four DNA-A variants among Pakistani isolates of cotton leaf curl virus and their affinities to DNA-A of geminivirus isolates from okra. *J. Gen. Virol.* **79**, 915–925.