JOURNAL OF VIROLOGY, Jan. 2008, p. 957–965 Vol. 82, No. 2
0022-538X/08/$08.00+0 doi:10.1128/JVI.01929-07

Phylogenetic Evidence for Rapid Rates of Molecular Evolution in the Single-Stranded DNA Begomovirus *Tomato Yellow Leaf Curl Virus*▽†

Siobain Duffy[1]* and Edward C. Holmes[1,2]

Center for Infectious Disease Dynamics, Department of Biology, Mueller Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802,[1] and Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892[2]

Received 4 September 2007/Accepted 22 October 2007

Geminiviruses are devastating viruses of plants that possess single-stranded DNA (ssDNA) DNA genomes. Despite the importance of this class of phytopathogen, there have been no estimates of the rate of nucleotide substitution in the geminiviruses. We report here the evolutionary rate of the tomato yellow leaf curl disease-causing viruses, an intensively studied group of monopartite begomoviruses. Sequences from GenBank, isolated from diseased plants between 1988 and 2006, were analyzed using Bayesian coalescent methods. The mean genomic substitution rate was estimated to be 2.88×10^{-4} nucleotide substitutions per site per year (subs/site/year), although this rate could be confounded by frequent recombination within *Tomato yellow leaf curl virus* genomes. A recombinant-free data set comprising the coat protein (V1) gene in isolation yielded a similar mean rate (4.63×10^{-4} subs/site/year), validating the order of magnitude of genomic substitution rate for protein-coding regions. The intergenic region, which is known to be more variable, was found to evolve even more rapidly, with a mean substitution rate of $\sim 1.56 \times 10^{-3}$ subs/site/year. Notably, these substitution rates, the first reported for a plant DNA virus, are in line with those estimated previously for mammalian ssDNA viruses and RNA viruses. Our results therefore suggest that the high evolutionary rate of the geminiviruses is not primarily due to frequent recombination and may explain their ability to emerge in novel hosts.

While most research on plant virus evolution has concentrated on RNA viruses (27, 73), the biggest emerging threat to crops worldwide are the single-stranded DNA (ssDNA) geminiviruses, especially those of the dicot-infecting, whitefly-transmitted *Begomoviridae* (47, 78). One reason for the emphasis on plant RNA viruses has been the desire to assess the extent and structure of genetic variation created by their use of error-prone RNA polymerase, which in turn is necessary to understand their molecular epidemiology (54) and which can complicate the creation of resistant crop lines (35). However, multiple genera of geminiviruses also display high levels of within-host genetic variation (37, 74, 84), suggesting that plant ssDNA viruses might exhibit genetic diversity similar to that seen in plant RNA viruses, even though they utilize host DNA polymerases (8). More recently, a controlled experimental study on within-host variability of the begomovirus *Tomato yellow leaf curl China virus* found an average mutation frequency of 3.5×10^{-4} per base in 60 days of infection on a tobacco plant (30), a value similar to those found in plant RNA viruses (76, 77). Together, these results imply that the mutation rate (and, by extension, long-term rates of nucleotide [nt] substitution) of the geminiviruses might be more similar to those of plant RNA viruses (as high as $\sim 10^{-5}$ nt/generation) (37, 46, 60) or to the ssDNA bacteriophage ϕX174 ($\sim 7 \times 10^{-6}$ nt/generation) (70) than to double-stranded DNA viruses ($\sim 10^{-8}$ nt/generation) (17).

The idea that ssDNA viruses exhibit substitution rates similar to those seen in RNA viruses (between 10^{-3} and 10^{-5} substitutions per site per year) (38) has been previously proposed. Most notably, studies of two mammalian parvoviruses, canine parvovirus and B19 erythovirus, found substitution rates similar to those of RNA viruses, at $\sim 10^{-4}$ substitutions per site per year (subs/site/year) (79, 80). However, to date, no such equivalent analyses have been conducted on any plant ssDNA virus.

Rather than considering substitution dynamics, most work on evolutionary mechanisms in the geminiviruses has focused on the causes and consequences of recombination (34). Geminiviruses can replicate by rolling-circle and recombination-dependent mechanisms (69) and show evidence of both homologous and heterologous recombination. Strikingly, recombinant strains occur frequently (21, 26), emerge often (4, 52, 62), and occasionally yield virulent strains that cause pandemics (95). However, although recombination is undoubtedly an important evolutionary mechanism in these viruses, it cannot create genetic variation de novo, so background mutation pressure must also be a key element of geminivirus evolution (30, 37).

To explore the dynamics of nt substitution in the geminiviruses, we undertook an analysis of the rate of molecular evolution in one of the most devastating and well-studied tomato pathogens, *Tomato yellow leaf curl virus* (TYLCV) (65), for which a relatively abundant data set of whole-genome sequences exists. This virus was first identified in 1939 (65). It caused the complete destruction of the tomato crop in the Jordan River valley in 1959 (10) and a near-complete loss in

* Corresponding author. Mailing address: Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802. Phone: (814) 863-6471. Fax: (814) 865-9131. E-mail: smd16@psu.edu.

† Supplemental material for this article may be found at http://jvi.asm.org/.

▽ Published ahead of print on 31 October 2007.

Hispaniola in the early 1990s (66). It was also the first begomovirus described with an unusual monopartite genome—analogous to the DNA-A strand—that is in some, but not all instances, accompanied by a DNA-β or a typical DNA-B (30, 71).

We used these data to estimate the rate of nt substitution and its determinants in the whole DNA-A-like genome of tomato yellow leaf curl disease (TYLCD)-causing viruses. The DNA-A codes for two proteins in the sense strand: a pre-coat (V2) and coat protein (V1), the latter of which interacts with the whitefly vector (65) and is shown to be conserved among geminiviruses (61). The complement of the DNA-A codes for four proteins: the replication-associated protein (Rep), a transcriptional activator of the coat protein (TrAP), a replication enhancer (REn), and the C4 protein that assists in intercell movement and which is entirely encoded in a different reading within the Rep gene (65). To increase the rigor of our analysis, we also examined the substitution rate in the coat protein-coding gene, *V1*, and in the "hypervariable" intergenic region (30, 61).

MATERIALS AND METHODS

Complete genome sequences of viruses casing TYLCD were downloaded from GenBank (Table 1). When the viruses were bipartite or were accompanied by a DNA-β, only the DNA-A was considered in the analysis. Viruses were abbreviated as most recently codified (22) or as suggested in publications referencing the sequence. For the viruses submitted directly to GenBank without accompanying publications, we named the sequences in accordance with current conventions and according to the isolate designation information contained in the GenBank file.

Since we were interested in the rate of molecular evolution, our analyses only considered sequences that could be dated. We therefore collected the dates (years) of isolation for as many of these TYLCD-causing viruses as possible from the GenBank files, from related publications, or by contacting the authors directly (Table 1). Sequences from viruses that had been reared for a substantial period of time on plants in the laboratory were excluded from analysis, since they were not necessarily reflective of evolutionary dynamics in nature. We also excluded the distantly related *Tomato yellow leaf curl Indonesia virus* and *Tomato yellow leaf curl Aragua virus*.

All genome sequences were organized to begin at the nick site in the invariant nonanucleotide sequence in the origin of replication (TAATATT-3′/5′-AC [61]). The data sets were then aligned by using MUSCLE (http://www.drive5.com/muscle/ [20]) and manually adjusted by using SE-AL (http://evolve.zoo.ox.ac.uk). In the coat protein-coding region, two sequence regions were excised from the alignment because they were present only in the TYLCV-[IL] isolate and would have generated three missense residues and the insertion of two residues in the coat protein-coding region (6 nt in total, nt numbers 1109 to 1114 and 1123 in GenBank accession X15656). Since TYLCV-[IL] was the first TYLCV sequenced, we considered it likely that these "mutations" were sequencing errors and so excluded them from our study. The final full genome alignment was 2,830 nt in length and consisted of 56 taxa (Table 2).

Subgenomic data sets. In addition to the full genome, the 56 genome data set was trimmed to create a smaller data set (780 nt), corresponding to the coding region for the coat protein (V1 or CP). We did not conduct an equivalent analysis for the gene encoding the replication-associated protein (Rep) because the length of the Rep protein was highly variable.

We also created two data sets to explore the rate of evolution in the rapidly evolving intergenic region (61) for two clades of TYLCV: the severe phenotype of TYLCV sensu stricto and the mild phenotype of TYLCV sensu stricto. We supplemented the dated full-length sequences with additional partial genome sequences of TYLCV with the published dates of isolation. For the severe TYLCV sensu stricto, 19 additional "intergenic region" sequences from Italy and Spain sampled between 2000 and 2003 (DQ121479 to DQ121485, DQ317727 to DQ317771) were added to the data set (excluding two highly divergent isolates: TYLCV-Gez and TYLCV-IR). The resultant 312 nt, 71-taxa alignment was trimmed to exclude all coding regions. The mild clade of TYLCV sensu stricto included all 12 whole-genome sequences and nine additional intergenic sequences that have dates of isolation between 2000 and 2003 (DQ121474 to DQ121478, DQ317772 to DQ317775). Trimming these to just the intergenic region yielded a 319-nt alignment.

Detection of recombination. Putative recombinant genomes were identified by using the RDP3 package (http://darwin.uvigo.es/rdp/rdp.html), which contains six recombination detection programs: RDP, GENECONV, MaxChi, Chimeara, Bootscan, and SiScan (49). The default detection thresholds were used in all cases.

Phylogenetic and coalescent analyses. Maximum-likelihood (ML) phylogenetic trees for each data set were estimated by using PAUP* 4.0 (86) with tree-bisection-recombination branch-swapping in each case, with the best-fit model of nt substitution for each data set determined using MODELTEST (68) (Table 2). To assess the support for individual nodes on the phylogenies, we undertook a bootstrap analysis utilizing 1,000 replicate neighbor-joining trees under the ML substitution model in each case. This analysis was also conducted in PAUP*.

To estimate rates of nt substitution per site and the "times to most recent common ancestor" (TMRCA), we used the Bayesian Markov chain Monte Carlo (MCMC) method available in the BEAST package (http://evolve.zoo.ox.ac.uk). Both strict and relaxed (uncorrelated exponential and uncorrelated lognormal) molecular clocks were utilized for each data set (19), as well as five demographic models (constant population size, expansion, exponential and logistic population growth, and a piecewise Bayesian skyline plot), which were used as coalescent priors. The best-fitting models were determined by a comparison of posterior probabilities and are shown in Table 2. MCMC chains were run for sufficient length to ensure convergence, with 10% of the MCMC chains discarded as burn-in, and statistical uncertainty in the estimates provided by the 95% highest probability density (HPD) values (Table 2).

Additional sequence analyses. The ratio of nonsynonymous (d_N) to synonymous (d_S) nt substitutions per site (ratio d_N/d_S) in the coat protein alignment was estimated by using the ML single-likelihood ancestor counting method available through the DATAMONKEY web server (44), with the nt substitution model chosen by MODELTEST and the input phylogenetic tree inferred by using neighbor joining.

Finally, we used a phylogenetic approach to assess the extent of mutational bias in the full genome data set. These data were first aligned with their probable closest outgroup (*African cassava mosaic virus*, Kenya isolate, GenBank accession number J02057 [see, for example, references 45 and 63]), and an ML phylogeny, rooted with this outgroup, was inferred by using the procedure described above (data not shown, but available from the authors on request). The number of unambiguous nt changes of each type (e.g., A→C and A→G), as well as the putative ancestral sequence, were then inferred by mapping all substitutions onto the ML phylogeny using the parsimony method available in PAUP*. These observed proportions were then compared, using χ^2 tests of significance, to those expected by summing the forward and reverse of each substitution type (i.e., the sum of the number of A→C and C→A transversions), assuming they were equally likely by dividing that sum in half, and then adjusting that value by the relative frequency of each base in the inferred ancestral sequence (i.e., frequency of C/[frequency of C + frequency of A]).

RESULTS

Frequent recombination in TYLCV genomes. For the full genome data set, a phylogenetic signal compatible with recombination (i.e., topological incongruence) was detected in nearly all of the isolates analyzed. Hence, removing recombinants would have left too few sequences for a viable estimation of substitution rates. However, while the occurrence of recombination violates the assumptions of coalescent analysis, this process is most likely to increase genetic variation and hence elevate both the average apparent substitution rate and its variability.

Significant phylogenetic evidence for recombination ($P < 0.05$) was also detected in the coat protein gene (*V1*) alignment, but in only in 2 of the 56 taxa: TYLCTHV-SaNa by RDP, Geneconv, and Bootscan and TYLCTHV-ChMai by Geneconv, Bootscan, MaxChi, Chimaera, and SciScan. Removing these two taxa was sufficient to remove detectable recombinants, producing a data set of 54 taxa for subsequent analysis.

TABLE 1. TYLCD-causing virus sequences used in this study

Isolate	Accession no.	Yr isolated	Sequence source or reference	Other source or reference[a]
Tomato yellow leaf curl virus				
Severe phenotype				
TYLCV-[IL]	X15656	1988	58	12
TYLCV-[DO]	AF024715	1994	56	
TYLCV-Gez	AY044138	1996	36	Most recent year in reference 36
TYLCV-IR	AJ132711	1997	4	K. Bananej
TYLCV-[CU]	AJ223505	1997	GenBank	Most recent year suggested by E. Bejarano
TYLCV-[Flo]	AY530931	1997	67, 92	
TYLCV-[Omu]	AB116630	1999	87	S. Ueda
TYLCV-[Alm]	AJ489258	1999	GenBank	E. Bejarano
TYLCV-[Mis]	AB116631	2000	87	S. Ueda
TYLCV-[EG:Ism]	AY594174	2000	1	M. Abhary
TYLCV-[Miy]	AB116629	2001	87	S. Ueda
TYLCV-[PR]	AY134494	2001	6	
TYLCV-[MA]	EF060196	2002	GenBank	N. Boukhatem
TYLCV-[TR:Mer1:04]	AJ812277	2004	43	43
TYLCV-[Tosa](H)	AB192966	2004	88	
TYLCV-[Tosa]	AB192965	2004	88	
TYLCV-[RE4]	AM409201	2004	14	J. M. Lett
TYLCV-[Sic]	DQ144621	2004	13	
TYLCV-TN	EF101929	2004	32	S. Gharsallah Chouchane
TYLCV-[JO]	EF054893	2006	GenBank	G. Anfoka
TYLCV-[MX:Cul]	DQ631892	2006	7	Most recent year in reference 7
Mild phenotype				
TYLCV-Mld[PT]	AF105975	1995	57	
TYLCV-Mld[Aic]	AB014347	1996	39	
TYLCV-Mld[Shi]	AB014346	1996	39	
TYLCV-Mld[ES7297]	AF071228	1997	57	
TYLCV-Mld[ES]	AJ519441	1998	GenBank	E. Bejarano
TYLCV-Mld[Sz:Yai]	AB116632	2001	87	S. Ueda
TYLCV-Mld[Atu]	AB116633	2002	87	S. Ueda
TYLCV-Mld[Kis]	AB116634	2002	87	S. Ueda
TYLCV-Mld[Sz:Dai]	AB116635	2002	87	S. Ueda
TYLCV-Mld[Sz:Osu]	AB116636	2002	87	S. Ueda
TYLCV-Mld[RE]	AJ865337	2002	15	J.-M. Lett
TYLCV-Mld[JO]	EF054894	2006	GenBank	G. Anfoka
Tomato yellow leaf curl Axarquia virus				
TYLCAxV-[Alg]	AY227892	2000	GenBank	E. Moriones
Tomato yellow leaf curl Malaga virus				
TYLCMalV	AF271234	1999	51	
Tomato yellow leaf curl Mali virus				
TYLCMLV	AY502934	2003	GenBank	T. Kon
TYLCMLV-[ET]	DQ358913	2003	81	
Tomato yellow leaf curl Sardinia virus				
TYLCSV	X61153	1988	41	
TYLCSV-[Sic]	Z28390	1991	11	G. Accotto
TYLCSV-[ES1]	Z25751	1992	59	
TYLCSV-[ES2]	L27708	1993	GenBank	E. R. Bejarano
TYLCSV-[MA]	AY702650	2002	31	GenBank
TYLCSV-[TN]	AY736854	2002	GenBank	GenBank
Tomato yellow leaf curl China virus				
TYLCCNV	AF311734	1989	93	Y. Hong
TYLCCNV-[Tb:Y25]	AJ457985	2000	45	
TYLCCNV-YM	DQ256460	2004	16	
Tomato yellow leaf curl Kanchanaburi virus				
TYLCKaV-[TH:Kan1]	AF511529	2001	GenBank	GenBank
TYLCKaV-[TH:Kan2]	AF511530	2001	GenBank	GenBank
TYLCKaV-[VN]	DQ169054	2001	33; GenBank	S. K. Green
Tomato yellow leaf curl Thailand virus				
TYLCTHV-[1]	X63015	1988	71	W. Kositratana
TYLCTHV-[2]	AF141922	1989	3	P. Chiemsombat
TYLCTHV-[MM]	AF206674	1999	33	GenBank
TYLCTHV-[Y72]	AJ495812	2002	45	
TYLCTHV-[ChMai]	AY514630	2003	75	P. Chiemsombat
TYLCTHV-[NoK]	AY514631	2003	75	P. Chiemsombat
TYLCTHV-[SaNa]	AY514632	2003	75	P. Chiemsombat

[a] That is, if not obtained from the sequence reference.

TABLE 2. Details of sequence alignments and parameter estimates

Parameter[a]	Full genome	Coat protein (V1) gene	TYLCV intergenic region	
			Severe phenotype	Mild phenotype
Recombination detected	Yes	No	No	No
Best-fit substitution model	GTR+I+Γ	GTR+Γ	TrN+Γ[b]	HKY+Γ
Best-fit molecular clock model	Relaxed (lognormal)	Relaxed (exponential)	Relaxed (exponential)	Relaxed (exponential)
Best-fit population growth model	Constant	Exponential	Bayesian skyline	Exponential
Sequence length (nt)	2,830	780	312	319
No. of sequences	56	54	71	21
Date range of sequences	1988–2006	1988–2006	1994–2006	1995–2006
Chain length (in millions)	60	20	40	20
Mean substitution rate	2.88×10^{-4}	4.63×10^{-4}	1.75×10^{-3}	1.37×10^{-3}
HPD substitution rate	1.34×10^{-6} to 6.06×10^{-4}	6.62×10^{-5} to 8.85×10^{-4}	1.03×10^{-3} to 2.59×10^{-3}	3.60×10^{-4} to 2.19×10^{-3}
Mean age (yr)	9,304	1,308	28	20
HPD age (yr)	290–40,211	221–3,319	19–44	13–35

[a] yr, years prior to 2007; HPD, 95% high probability density.
[b] GTR+ Γ used for BEAST.

No recombination was also detected in either of the intergenic data sets.

Viral species are supported as monophyletic groups. The ML phylogenetic tree estimated for the full genome alignment is shown in Fig. 1. Despite the high levels of recombination detected in this data set, the monophyly of nearly all TYLCD-causing viral species was strongly supported (the placement of TYLCV-Gez, a known interspecific recombinant [36] disrupted the monophyly of TYLCV sensu stricto). Further, the phylogenies estimated from the whole-genome alignment and the coat protein (*V1*) alignment were qualitatively congruent (Fig. 1 and 2). One notable exception was the grouping of the three Asian species: TYLCCNV, TYLCTHV, and TYLCKaV. The *V1* tree strongly suggests that the coat protein of these Asian species is most closely related to that of TYLCSV and its recombinant relatives (Fig. 2), which are predominantly found in the Mediterranean (25), whereas there is no bootstrap support for this in the full genome phylogeny (Fig. 1). Support for a clade containing TYLCSV and the Asian TYLCV has been previously observed with a smaller coat protein data set (53) but is unsupported in all full genome (DNA-A) phylogenies published to date (see, for example, references 52 and 64). The phylogenies for the two intergenic alignments are provided in the supplemental material.

Mutational bias in the TYLCV genome. To further reveal the processes shaping TYLCD-causing virus evolution, we determined whether specific kinds of mutational changes are overrepresented relative to others in the full-genome alignment. After adjusting for the initial nt frequency, 8 of the 12 substitution patterns were significantly more (C→A, C→T, G→A, and G→T) or less (A→C, A→G, T→C, and T→G) likely than expected (Table 3). The most significant of these were C→T ($P = 6.9 \times 10^{-4}$) and G→A ($P = 1.1 \times 10^{-6}$). Therefore, these results indicate there are major biases in the evolution of the TYLCD-causing viruses, even within the same substitution types. These biases also favor a shift from CG→AT genome content, most notably in a bias in C→T and G→A transitions.

Estimation of nt substitution rates. Although a variety of different demographic models were supported, in all cases the relaxed molecular clock was a better fit than the strict molecular clock to the data analyzed here (Table 2). Notably, for all four data sets, qualitatively (and, for the mean values, quantitatively) similar values were found with all demographic models and for both uncorrelated lognormal and exponential relaxed molecular clock models. The mean rate of nt substitution estimated for the full genome was high, at 2.88×10^{-4} subs/site/year (Table 2) and hence within the range documented in RNA viruses (38). As we had expected given recombination, the 95% HPD values around that average estimate were extensive, representing the widest HPD calculated for any of the alignments in the present study.

Despite the occurrence of recombination, that similarly high rates were observed in the coat protein gene data set where recombinants were excluded reveals that this process has not significantly biased estimates. Indeed, the mean substitution rate in the coat protein gene data set was higher than that of the whole genome, at 4.63×10^{-4} subs/site/year. It was also notable that the high substitution rate for the coat protein gene was largely due to synonymous substitutions ($d_N/d_S = 0.12$). This low d_N/d_S ratio is typical of the coat proteins of vector-borne plant RNA viruses and most likely reflects strong purifying selection resulting from the highly specific interaction between the virus and the insect vector (27, 78). Indeed, in our data set, 44% of the codons (114 of 259) were determined to be under statistically significant negative selection, with no sites under positive selection ($P < 0.05$), so that most fixations are likely to be neutral.

Substitution rates were even higher for the noncoding intergenic regions (severe, 1.75×10^{-3} subs/site/year; mild, 1.37×10^{-3} subs/site/year; Table 2). Notably, the severe intergenic region evolved significantly faster than *V1* such that the 95% HPD values around the average estimates did not overlap. The mild TYLCV analysis had wider HPD values, presumably because there were fewer sequences in this data set compared to the TYLCV severe intergenic analysis (21 compared to 71; Table 2). Others have also noted the increased genetic variability of the intergenic region relative to genes in geminiviruses, likely due to relaxed purifying selection since there are fewer critical functions coded in this region (30, 61).

Age of TYLCV species. Full genome sequences from the eight divergent TYLCD-causing virus species (Table 1) were estimated to have a mean TMRCA of approximately 10,000 years ago, with a very wide range of 95% HPD values of between 290 and 40,211 years ago (Table 2). A more recent mean TMRCA of 1,308 years was estimated for the coat pro-

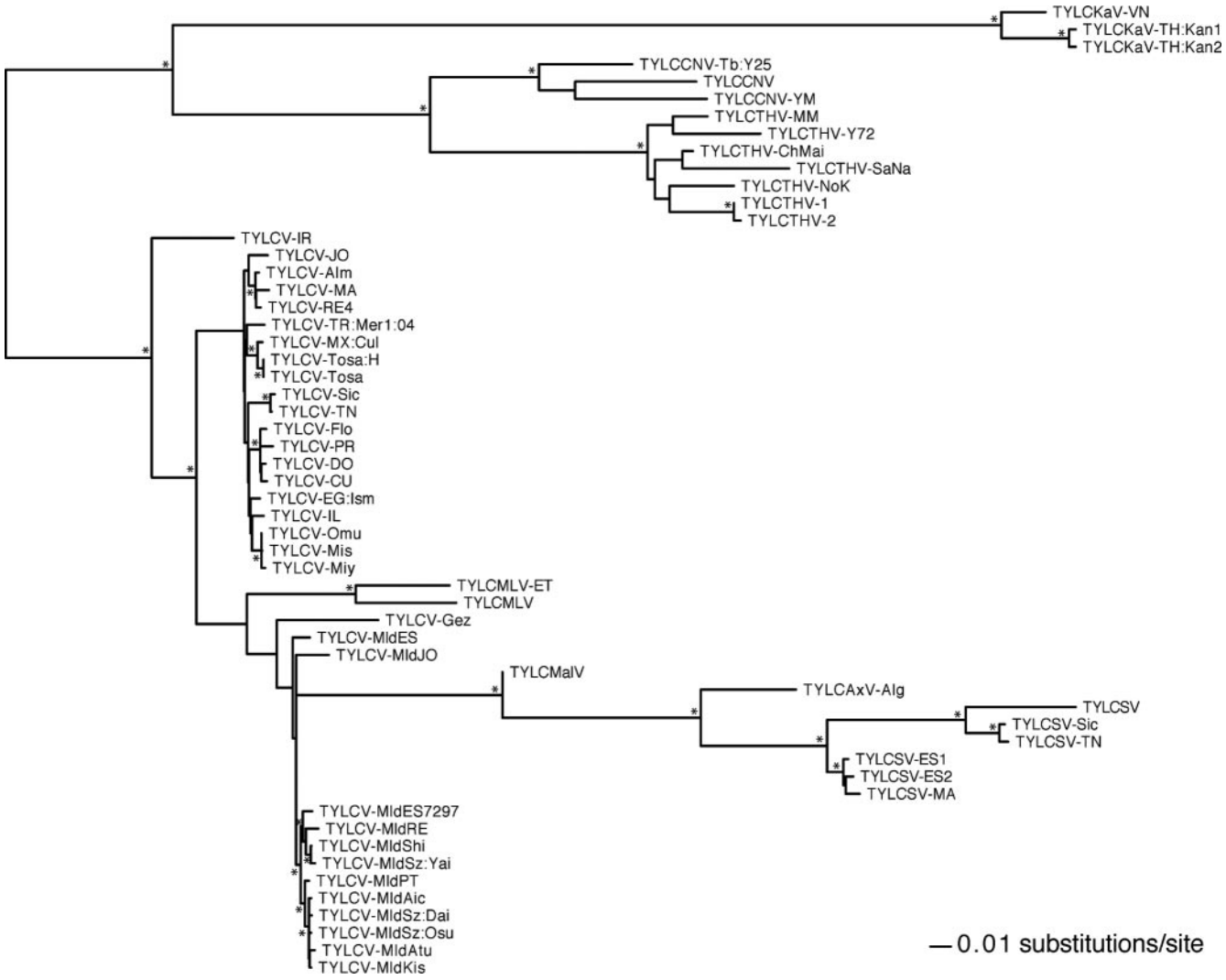

FIG. 1. ML phylogeny of TYLCD-causing viral genomes. Branch lengths are scaled according to the number of nt substitutions per site, and the tree is mid-point rooted for clarity only. Asterisks denote all nodes with >80% bootstrap support.

tein data set, although this falls within the HPD values estimated for the genomic data set (Table 2). The TMRCAs estimated for the severe and mild phenotypes from the intergenic regions were even more recent, at approximately 30 and 20 years ago, respectively (Table 2). Despite this wide range of date estimates, our analyses reveal that the TYLCD-causing viruses diversified from a common ancestor long before TYLCD became a significant agricultural problem, since the symptoms associated with this virus were not noted until 1939 (65). On the basis of the intergenic region, we infer that the severe phenotype TYLCV sensu stricto isolates shared a common ancestor within the last 44 years (upper HPD value). This suggests that the diversification of TYLCV sensu stricto occurred after the first major outbreak of TYLCD in the Jordan River valley in 1959 (65).

The mild TYLCV clade was well supported in the full genome phylogeny, and all but two mild isolates formed a well-supported clade in the *V1* phylogeny. The mild phenotype of TYLCV is due to incomplete base pairing in a stem-loop motif in the intergenic region (57), and recombination could have led to "mild" isolates with different V1 genes (as is likely the case for TYLCV-Mld[RE] and TYLCV-Mld[JO]; Fig. 2). Our data suggest that the intergenic region of the mild phenotype TYLCV evolved once, and the TMRCA of the isolates included in the present study is within the last 35 years (Table 2). Since the first mild phenotype TYLCV was isolated from a mixed infection maintained in the laboratory from an initial isolation in the late 1960s (2), it could be that the mild genotype existed in mixed infection with severe TYLCV for some time before the virus spread worldwide.

DISCUSSION

We were able to elucidate the evolutionary history and dynamics of the TYLCD-causing viruses by using dated genomic sequences. Most notably, we observed that these geminiviruses evolve as rapidly as some RNA viruses. Also of note was that the mild phenotype of TYLCV appears to have evolved only

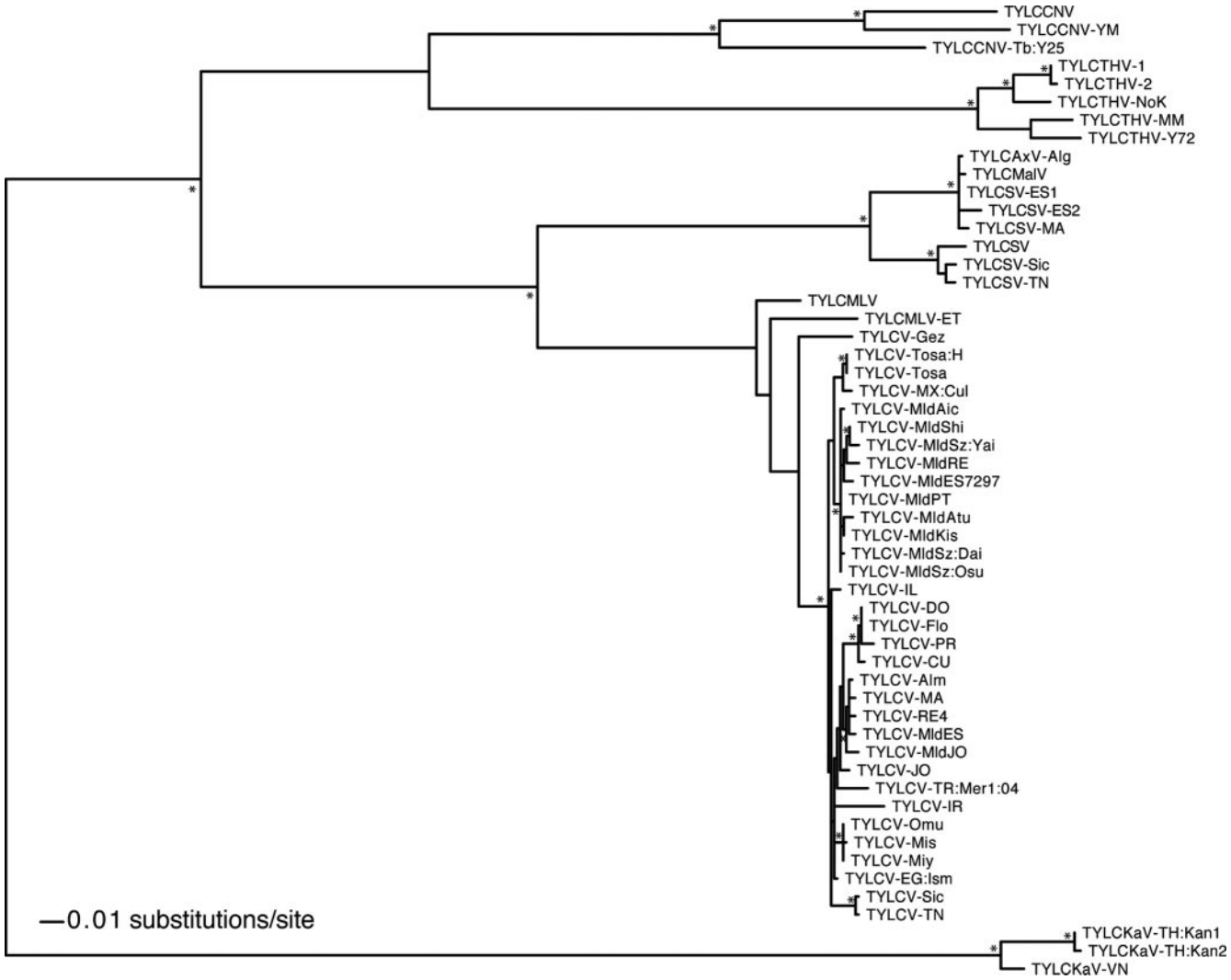

FIG. 2. ML phylogeny of the TYLCD-causing virus coat protein-coding gene *V1*. Branch lengths are scaled according to the number of nt substitutions per site, and the tree is mid-point rooted for clarity only. Asterisks denote all nodes with >80% bootstrap support.

once from severe TYLCV (13, 21, 36, 43, 52, 57, 88) and within the last 35 years has spread to places as distant as Japan (88), Reunion Island (15), and Venezuela (94).

While there have been few estimates of substitution rates in plant viruses, there is some evidence that plant RNA viruses are more genetically stable than animal RNA viruses, leading to lower rates of evolution (24, 27, 28) attributable to a lack of immune selection on plant viruses (27). Indeed, there are prior suggestions that geminiviruses evolve faster than some plant RNA viruses (60), including tobamoviruses (40, 42, 72) and criniviruses (48). We present here the first evidence that this might be the case: the coat protein gene evolves at least as rapidly as plant RNA viruses (40, 85), as quickly as animal ssDNA viruses (79, 80, 89), and as fast as some animal RNA viruses (38). Most notably, the substitution rates inferred here are many orders of magnitude greater than those observed in double-stranded DNA viruses (5, 50). It is also likely that the high substitution rate of the begomovirus TYLCV is typical of

TABLE 3. nt substitutions inferred from the ML phylogeny of the TYLCD-causing viruses, evaluated with χ^2 tests of significance[a]

Original base	No. of nt substitutions (expected no. of nt substitutions), χ^2, P			
	A	C	G	T
A		123 (202.8), 6.7, 9.6×10^{-3}	312 (507.5), 18.4, 1.8×10^{-5}	309 (317.4), 0.05, 0.80
C	140 (85.4), 9.7, 1.9×10^{-3}		115 (97.2), 0.84, 0.36	473 (360.1), 11.5, 6.9×10^{-4}
G	395 (243.8), 23.7, 1.1×10^{-6}	109 (129), 0.75, 0.39		270 (205.3), 6.0, 1.4×10^{-2}
T	337 (328.1), 0.06, 0.81	551 (723.7), 7.5, 6.1×10^{-3}	280 (382), 4.4, 3.6×10^{-2}	

[a] The expected number of substitutions (see Materials and Methods) is shown in parentheses. The significance (P value) of the chi-square value is also shown.

geminiviruses as a whole. High levels of genetic variation have been observed in the curtovirus *Beet curly top virus* (84), and a high mutation frequency has been observed in the mastrevirus *Maize streak virus* (37).

Hence, we propose that there is a specific feature of geminivirus genomes, and perhaps of ssDNA viruses as a whole, that leads to these high, RNA virus-like rates of evolution. Although positive selection may greatly elevate substitution rates compared to background (neutral) mutation rates, the very high evolutionary rates observed in the intergenic regions and the low d_N/d_S ratios in the coat protein gene strongly suggest that the high substitution rates observed in TYLCV are a reflection of rapid mutational dynamics rather than frequent adaptive evolution.

It is currently unclear what mechanisms lead to the elevated substitution rates in ssDNA viruses (79). Unlike RNA viruses that use their own error-prone polymerases, geminiviruses use the cellular machinery of their host plants, implying that they should have the same replication fidelity as their host. However, it is possible that geminivirus genomes cannot be repaired by host exonucleases because they do not possess the proper methylation patterns, which will increase mutation rates during replication of viral DNA (74). Alternatively, even if geminivirus genomes are correctly methylated, base-excision repair might not proceed on replicating geminivirus genomes because they are only transiently double stranded during rolling-circle replication. It is also possible that geminivirus genomes recruit a more error-prone polymerase (such as one lacking a base-excision repair mechanism altogether) from the host's nucleus for its own replication (29).

The biased pattern of mutation, favoring C→T and G→A transitions, suggests another potential mechanism for rapid mutation. Both of the most significant transition biases commonly occur due to deamination of the bases: from cytosine to uracil and from guanine to xanthine (which base pairs to thymine [9]). Base deamination can occur spontaneously and is more likely to occur when the nt is not based paired, as for instance when DNA spends a prolonged period of time in a single-stranded state (9, 23, 91). Biases consistent with deaminated nt are often observed in viral genomes, such as human immunodeficiency virus, because animals have specific enzymes to deaminate ssDNA and ssRNA, such as the APOBEC family (90). Deaminating enzymes have been identified in plant nuclei (82) that could be responsible for the biased substitutions observed in TYLCD-causing viruses, in a process unrelated to mutations due to DNA polymerase error.

Consequently, deamination could represent a significant component of mutation in TYLCD-causing viruses, particularly since a recent mutation frequency study in TYLCCNV revealed the same transition biases as our genomic analysis (30). Deamination of bases, whether by enzymatic or spontaneous reactions, is a mutational mechanism that occurs before the template replicates; it is therefore separate from the polymerase-induced mutation that is measured by mutation rate assays. This additional source of mutation likely experienced by ssDNA viruses may help explain how geminiviruses generate variation while using fidelitous plant DNA polymerases. However, without direct experimental tests, we cannot be certain of the role that the significant base deamination found in TYLCD-causing virus genomes plays in shaping mutation rates.

RNA viruses mutate quickly (18) and usually have high substitution rates (38). While these traits are not necessarily adaptive, they are thought to facilitate viral emergence on novel hosts (55, 83). The frequent emergence of geminiviruses on novel hosts has previously been explained by the fact that they are more vector specific than host specific (78) and by the creation of genetic variation by recombination (62). We demonstrate here that the geminivirus TYLCV also has a high, RNA-like rate of nt substitution, independent of its frequent recombination, and this could be an important factor in its emergence on novel crops.

ACKNOWLEDGMENTS

This study was supported by National Science Foundation grant DBI-0630707 to S.D.

We thank Laura Shackelton for technical assistance. We thank the geminivirus research community for supplying dates of isolation for many TYLCD-causing virus sequences and an anonymous reviewer for the suggestion that host error correction mechanisms might not function during rolling-circle replication.

REFERENCES

1. **Abhary, M. K., G. H. Anfoka, M. K. Nakhla, and D. P. Maxwell.** 2006. Post-transcriptional gene silencing in controlling viruses of the *Tomato yellow leaf curl virus* complex. Arch. Virol. **151:**2349–2363.
2. **Antignus, Y., and S. Cohen.** 1994. Complete nucleotide-sequence of an infectious clone of a mild isolate of *Tomato yellow leaf curl virus* (TYLCV). Phytopathology **84:**707–712.
3. **Attathom, S., P. Chiemsombat, W. Kositratana, and N. Sae-Ung.** 1994. Complete nucleotide sequence and genome analysis of bipartite *Tomato yellow leaf curl virus* in Thailand. Kasetsart J. (Nat. Sci.) **28:**632–639.
4. **Bananej, K., A. Kheyr-Pour, G. H. Salekdeh, and A. Ahoonmanesh.** 2004. Complete nucleotide sequence of Iranian *tomato yellow leaf curl virus* isolate: further evidence for natural recombination amongst begomoviruses. Arch. Virol. **149:**1435–1443.
5. **Bernard, H. U.** 1994. Coevolution of papillomaviruses with human populations. Trends Microbiol. **2:**140–143.
6. **Bird, J., A. M. Idris, D. Rogan, and J. K. Brown.** 2001. Introduction of the exotic *Tomato yellow leaf curl virus-Israel* in tomato to Puerto Rico. Plant Dis. **85:**1028.
7. **Brown, J. K., and A. M. Idris.** 2006. Introduction of the exotic monopartite *Tomato yellow leaf curl virus* into west coast Mexico. Plant Dis. **90:**1360–1362.
8. **Carrillo-Tripp, J., H. Shimada-Beltran, and R. Rivera-Bustamante.** 2006. Use of geminiviral vectors for functional genomics. Curr. Opin. Plant Biol. **9:**209–215.
9. **Caulfield, J. L., J. S. Wishnok, and S. R. Tannenbaum.** 1998. Nitric oxide-induced deamination of cytosine and guanine in deoxynucleosides and oligonucleotides. J. Biol. Chem. **273:**12689–12695.
10. **Cohen, S., and Y. Antignus.** 1994. *Tomato yellow leaf curl virus* (TYLCV), a whitefly-borne geminivirus of tomatoes. Adv. Dis. Vector Res. **10:**259–288.
11. **Crespi, S., E. Noris, A. M. Vaira, and G. P. Accotto.** 1995. Molecular characterization of cloned DNA from a *Tomato yellow leaf curl virus* isolate from Sicily. Phytopathol. Mediterr. **34:**93–99.
12. **Czosnek, H., and H. Laterrot.** 1997. A worldwide survey of *Tomato yellow leaf curl viruses*. Arch. Virol. **142:**1391–1406.
13. **Davino, S., C. Napoli, M. Davino, and G. P. Accotto.** 2006. Spread of *Tomato yellow leaf curl virus* in Sicily: partial displacement of another geminivirus originally present. Eur. J. Plant Pathol. **114:**293–299.
14. **Delatte, H., H. Holota, F. Naze, M. Peterschmitt, B. Reynaud, and J. M. Lett.** 2005. The presence of both recombinant and nonrecombinant strains of *Tomato yellow leaf curl virus* on tomato in Reunion Island. Plant Pathol. **54:**262–264.
15. **Delatte, H., D. P. Martin, F. Naze, R. Goldbach, B. Reynaud, M. Peterschmitt, and J. M. Lett.** 2005. South West Indian Ocean islands tomato begomovirus populations represent a new major monopartite begomovirus group. J. Gen. Virol. **86:**1533–1542.
16. **Dong, J. H., Y. Q. Luo, M. Ding, Z. K. Zhang, and C. K. Yang.** 2007. First report of *Tomato yellow leaf curl China virus* infecting kidney bean in China. Plant Pathol. **56:**342.
17. **Drake, J. W.** 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88:**7160–7164.
18. **Drake, J. W.** 1993. Rates of spontaneous mutation among RNA viruses. Proc. Natl. Acad. Sci. USA **90:**4171–4175.

19. **Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut.** 2006. Relaxed phylogenetics and dating with confidence. PLOS Biol. **4:**699–710.
20. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792–1797.
21. **Fauquet, C. M., S. Sawyer, A. M. Idris, and J. K. Brown.** 2005. Sequence analysis and classification of apparent recombinant begomoviruses infecting tomato in the Nile and Mediterranean basins. Phytopathology **95:**549–555.
22. **Fauquet, C. M., and J. Stanley.** 2005. Revising the way we conceive and name viruses below the species level: a review of geminivirus taxonomy calls for new standardized isolate descriptors. Arch. Virol. **150:**2151–2179.
23. **Frederico, L. A., T. A. Kunkel, and B. R. Shaw.** 1990. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation-energy. Biochemistry **29:**2532–2537.
24. **French, R., and D. C. Stenger.** 2003. Evolution of *Wheat streak mosaic virus*: dynamics of population growth within plants may explain limited variation. Annu. Rev. Phytopathol. **41:**199–214.
25. **García-Andrés, S., G. P. Accotto, J. Navas-Castillo, and E. Moriones.** 2007. Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. Virology **359:**302–312.
26. **García-Andrés, S., D. M. Tomás, S. Sánchez-Campos, J. Navas-Castillo, and E. Moriones.** 2007. Frequent occurrence of recombinants in mixed infections of tomato yellow leaf curl disease-associated begomoviruses. Virology **365:** 210–219.
27. **García-Arenal, F., A. Fraile, and J. M. Malpica.** 2001. Variability and genetic structure of plant virus populations. Annu. Rev. Phytopathol. **39:**157–186.
28. **García-Arenal, F., A. Fraile, and J. M. Malpica.** 2003. Variation and evolution of plant virus populations. Int. Microbiol. **6:**225–232.
29. **Garcia-Diaz, M., and K. Bebenek.** 2007. Multiple functions of DNA polymerases. Crit. Rev. Plant Sci. **26:**105–122.
30. **Ge, L. M., J. T. Zhang, X. P. Zhou, and H. Y. Li.** 2007. Genetic structure and population variability of *Tomato yellow leaf curl China virus*. J. Virol. **81:** 5902–5907.
31. **Gharsallah Chouchane, S., F. Gorsane, M. K. Nakhla, D. P. Maxwell, M. Marrakchi, and H. Fakhfakh.** 2006. Complete nucleotide sequence and construction of an infectious clone of a Tunisian isolate of *Tomato yellow leaf curl Sardinia virus*. J. Phytopathol. **154:**626–631.
32. **Gharsallah Chouchane, S., F. Gorsane, M. K. Nakhla, D. P. Maxwell, M. Marrakchi, and H. Fakhfakh.** 2007. First report of *Tomato yellow leaf curl virus Israel* (TYLCV-[IL]) species infecting tomato, pepper, and bean in Tunisia. J. Phytopathol. **155:**236–240.
33. **Green, S. K., W. S. Tsai, S. L. Shih, L. L. Black, A. Rezaian, M. H. Rashid, M. M. Roff, Y. Y. Myint, and L. T. A. Hong.** 2001. Molecular characterization of begomoviruses associated with leafcurl diseases of tomato in Bangladesh, Laos, Malaysia, Myanmar, and Vietnam. Plant Dis. **85:**1286.
34. **Harrison, B. D., and D. J. Robinson.** 1999. Natural genomic and antigenic variation in whitefly-transmitted geminiviruses (begomoviruses). Annu. Rev. Phytopathol. **37:**369–398.
35. **Herrero, S., A. K. Culbreath, A. S. Csinos, H. R. Pappu, R. C. Rufty, and M. E. Daub.** 2000. Nucleocapsid gene-mediated transgenic resistance provides protection against *Tomato spotted wilt virus* epidemics in the field. Phytopathology **90:**139–147.
36. **Idris, A. M., and J. K. Brown.** 2005. Evidence for interspecific-recombination for three monopartite begomoviral genomes associated with the tomato leaf curl disease from central Sudan. Arch. Virol. **150:**1003–1012.
37. **Isnard, M., M. Granier, R. Frutos, B. Reynaud, and M. Peterschmitt.** 1998. Quasispecies nature of three *Maize streak virus* isolates obtained through different modes of selection from a population used to assess response to infection of maize cultivars. J. Gen. Virol. **79:**3091–3099.
38. **Jenkins, G. M., A. Rambaut, O. G. Pybus, and E. C. Holmes.** 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. **54:**156–165.
39. **Kato, K., M. Onuki, S. Fuji, and K. Hanada.** 1998. The first occurrence of *Tomato yellow leaf curl virus* in tomato (*Lycopersicon esculentum* Mill.) in Japan. Ann. Phytopathol. Soc. Jpn. **64:**552–559.
40. **Kearney, C. M., M. J. Thomson, and K. E. Roland.** 1999. Genome evolution of *Tobacco mosaic virus* populations during long-term passaging in a diverse range of hosts. Arch. Virol. **144:**1513–1526.
41. **Kheyr-Pour, A., M. Bendahmane, V. Matzeit, G. P. Accotto, S. Crespi, and B. Gronenborn.** 1991. *Tomato yellow leaf curl virus* from Sardinia is a whitefly-transmitted monopartite geminivirus. Nucleic Acids Res. **19:**6763–6769.
42. **Kim, T., M. Y. Youn, B. E. Min, S. H. Choi, M. Kim, and K. H. Ryu.** 2005. Molecular analysis of quasispecies of *Kyuri green mottle mosaic virus*. Virus Res. **110:**161–167.
43. **Koklu, G., A. Rojas, and A. Kvarnheden.** 2006. Molecular identification and the complete nucleotide sequence of a *Tomato yellow leaf curl virus* isolate from Turkey. J. Plant Pathol. **88:**61–66.
44. **Kosakovsky Pond, S. L., and S. D. W. Frost.** 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics **21:**2531–2533.
45. **Li, Z. H., X. P. Zhou, X. Zhang, and Y. Xie.** 2004. Molecular characterization of tomato-infecting begomoviruses in Yunnan, China. Arch. Virol. **149:** 1721–1732.
46. **Malpica, J. M., A. Fraile, I. Moreno, C. I. Obies, J. W. Drake, and F. Garcia-Arenal.** 2002. The rate and character of spontaneous mutation in an RNA virus. Genetics **162:**1505–1511.
47. **Mansoor, S., R. W. Briddon, S. E. Bull, I. D. Bedford, A. Bashir, M. Hussain, M. Saeed, Y. Zafar, K. A. Malik, C. Fauquet, and P. G. Markham.** 2003. Cotton leaf curl disease is associated with multiple monopartite begomoviruses supported by single DNA beta. Arch. Virol. **148:**1969–1986.
48. **Marco, C. F., and M. A. Aranda.** 2005. Genetic diversity of a natural population of *Cucurbit yellow stunting disorder virus*. J. Gen. Virol. **86:**815–822.
49. **Martin, D. P., C. Williamson, and D. Posada.** 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics **21:**260–262.
50. **McGeoch, D. J., and D. Gatherer.** 2005. Integrating reptilian herpesviruses into the family herpesviridae. J. Virol. **79:**725–731.
51. **Monci, F., J. Navas-Castillo, and E. Moriones.** 2001. Evidence of a naturally occurring recombinant between *Tomato yellow leaf curl virus* and *Tomato yellow leaf curl Sardinia virus* in Spain. Plant Dis. **85:**1289.
52. **Monci, F., S. Sanchez-Campos, J. Navas-Castillo, and E. Moriones.** 2002. A natural recombinant between the geminiviruses *Tomato yellow leaf curl Sardinia virus* and *Tomato yellow leaf curl virus* exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. Virology **303:** 317–326.
53. **Moriones, E., and J. Navas-Castillo.** 2000. *Tomato yellow leaf curl virus*, an emerging virus complex causing epidemics worldwide. Virus Res. **71:**123–134.
54. **Moury, B., C. Desbiez, M. Jacquernond, and H. Lecoq.** 2006. Genetic diversity of plant virus populations: towards hypothesis testing in molecular epidemiology. Plant Virus Epidemiol. **67:**49–87.
55. **Moya, A., E. C. Holmes, and F. Gonzalez-Candelas.** 2004. The population genetics and evolutionary epidemiology of RNA viruses. Nat. Rev. Microbiol. **2:**279–288.
56. **Nakhla, M. K., D. P. Maxwell, R. T. Martinez, M. G. Carvalho, and R. L. Gilbertson.** 1994. Widespread occurrence of the eastern Mediterranean strain of *Tomato yellow leaf curl geminivirus* in tomatoes in the Dominican Republic. Plant Dis. **78:**926.
57. **Navas-Castillo, J., S. Sanchez-Campos, E. Noris, D. Louro, G. P. Accotto, and E. Moriones.** 2000. Natural recombination between *Tomato yellow leaf curl virus-Is* and *Tomato leaf curl virus*. J. Gen. Virol. **81:**2797–2801.
58. **Navot, N., E. Pichersky, M. Zeidan, D. Zamir, and H. Czosnek.** 1991. *Tomato yellow leaf curl virus*: a whitefly-transmitted geminivirus with a single genomic component. Virology **185:**151–161.
59. **Noris, E., E. Hidalgo, G. P. Accotto, and E. Moriones.** 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.
60. **Ooi, K., S. Ohshita, I. Ishii, and T. Yahara.** 1997. Molecular phylogeny of geminivirus infecting wild plants in Japan. J. Plant Res. **110:**247–257.
61. **Padidam, M., R. N. Beachy, and C. M. Fauquet.** 1995. Classification and identification of geminiviruses using sequence comparisons. J. Gen. Virol. **76:**249–263.
62. **Padidam, M., S. Sawyer, and C. M. Fauquet.** 1999. Possible emergence of new geminiviruses by frequent recombination. Virology **265:**218–225.
63. **Paximadis, M., A. M. Idris, I. Torres-Jerez, A. Villarreal, M. E. C. Rey, and J. K. Brown.** 1999. Characterization of tobacco geminiviruses in the Old and New World. Arch. Virol. **144:**703–717.
64. **Paximadis, M., and M. E. C. Rey.** 2001. Genome organization of *Tobacco leaf curl Zimbabwe virus*, a new, distinct monopartite begomovirus associated with subgenomic defective DNA molecules. J. Gen. Virol. **82:**3091–3097.
65. **Picó, B., M. J. Díez, and F. Nuez.** 1996. Viral diseases causing the greatest economic losses to the tomato crop. 2. The Tomato yellow leaf curl virus: a review. Sci. Hort. **67:**151–196.
66. **Polston, J. E., D. Bois, and C. A. Serra.** 1994. First report of a tomato yellow leaf curl-like geminivirus in the western hemisphere. Plant Dis. **78:**831.
67. **Polston, J. E., R. J. McGovern, and L. G. Brown.** 1999. Introduction of *Tomato yellow leaf curl virus* in Florida and implications for the spread of this and other geminiviruses of tomato. Plant Dis. **83:**984–988.
68. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818.
69. **Preiss, W., and H. Jeske.** 2003. Multitasking in replication is common among geminiviruses. J. Virol. **77:**2972–2980.
70. **Raney, J. L., R. R. Delongchamp, and C. R. Valentine.** 2004. Spontaneous mutant frequency and mutation spectrum for gene A of Phi X174 grown in *Escherichia coli*. Environ. Mol. Mutag. **44:**119–127.
71. **Rochester, D. E., J. J. Depaulo, C. M. Fauquet, and R. N. Beachy.** 1994. Complete nucleotide sequence of the geminivirus *Tomato yellow leaf curl virus*, Thailand isolate. J. Gen. Virol. **75:**477–485.
72. **Rodríguez-Cerezo, E., S. F. Elena, A. Moya, and F. García-Arenal.** 1991. High genetic stability in natural populations of the plant RNA virus *Tobacco mild green mosaic virus*. J. Mol. Evol. **32:**328–332.

73. **Roossinck, M. J.** 1997. Mechanisms of plant virus evolution. Annu. Rev. Phytopathol. **35:**191–209.
74. **Sanz, A. I., A. Fraile, J. M. Gallego, J. M. Malpica, and F. Garcia-Arenal.** 1999. Genetic variability of natural populations of *Cotton leaf curl geminivirus*, a single-stranded DNA virus. J. Mol. Evol. **49:**672–681.
75. **Sawangjit, S., O. Chatchawankanphanich, P. Chiemsombat, T. Attathom, J. Dale, and S. Attathom.** 2005. Molecular characterization of tomato-infecting begomoviruses in Thailand. Virus Res. **109:**1–8.
76. **Schneider, W. L., and M. J. Roossinck.** 2000. Evolutionarily related Sindbis-like plant viruses maintain different levels of population diversity in a common host. J. Virol. **74:**3130–3134.
77. **Schneider, W. L., and M. J. Roossinck.** 2001. Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. J. Virol. **75:**6566–6571.
78. **Seal, S. E., F. vandenBosch, and M. J. Jeger.** 2006. Factors influencing begomovirus evolution and their increasing global significance: implications for sustainable control. Crit. Rev. Plant Sci. **25:**23–46.
79. **Shackelton, L. A., and E. C. Holmes.** 2006. Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. J. Virol. **80:**3666–3669.
80. **Shackelton, L. A., C. R. Parrish, U. Truyen, and E. C. Holmes.** 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc. Natl. Acad. Sci. USA **102:**379–384.
81. **Shih, S. L., S. K. Green, W. S. Tsai, L. M. Lee, and J. T. Wang.** 2006. First report of a begomovirus associated with Tomato yellow leaf curl disease in Ethiopia. Plant Dis. **90:**974.
82. **Stasolla, C., R. Katahira, T. A. Thorpe, and H. Ashihara.** 2003. Purine and pyrimidine nucleotide metabolism in higher plants. J. Plant Phys. **160:**1271–1295.
83. **Steinhauer, D. A., and J. J. Holland.** 1987. Rapid evolution of RNA viruses. Annu. Rev. Microbiol. **41:**409–433.
84. **Stenger, D. C., and C. L. McMahon.** 1997. Genotypic diversity of *Beet curly top virus* populations in the western United States. Phytopathology **87:**737–744.
85. **Stenger, D. C., D. L. Seifers, and R. French.** 2002. Patterns of polymorphism in *Wheat streak mosaic virus*: sequence space explored by a clade of closely related viral genotypes rivals that between the most divergent strains. Virology **302:**58–70.
86. **Swofford, D. L.** 2003. PAUP*: phylogenetic analysis using parsimony (* and other methods), 4.0b8 ed. Sinauer Associates, Sunderland, MA.
87. **Ueda, S., T. Kimura, M. Onuki, K. Hanada, and T. Iwanami.** 2004. Three distinct groups of isolates of *Tomato yellow leaf curl virus* in Japan and construction of an infectious clone. J. Gen. Plant Pathol. **70:**232–238.
88. **Ueda, S., S. Takeuchi, M. Okabayashi, K. Hanada, K. Tomimura, and T. Iwanami.** 2005. Evidence of a new *Tomato yellow leaf curl virus* in Japan and its detection using PCR. J. Gen. Plant Pathol. **71:**319–325.
89. **Umemura, T., Y. Tanaka, K. Kiyosawa, H. J. Alter, and J. W.-K. Shih.** 2002. Observation of positive selection within hypervariable regions of a newly identified DNA virus (SEN virus). FEBS Lett. **510:**171–174.
90. **Walsh, C. P., and G. L. Xu.** 2006. Cytosine methylation and DNA repair. Curr. Top. Microbiol. Immunol. **301:**283–315.
91. **Xia, X., and K. Y. Yuen.** 2005. Differential selection and mutation between dsDNA and ssDNA phages shape the evolution of their genomic AT percentage. BMC Genet. **6:**20.
92. **Yang, Y., T. A. Sherwood, C. P. Patte, E. Hiebert, and J. E. Polston.** 2004. Use of *Tomato yellow leaf curl virus* (TYLCV) Rep gene sequences to engineer TYLCV resistance in tomato. Phytopathology **94:**490–496.
93. **Yin, Q. Y., H. Y. Yang, Q. H. Gong, H. Y. Wang, Y. L. Liu, Y. G. Hong, and P. Tien.** 2001. *Tomato yellow leaf curl China virus*: monopartite genome organization and agroinfection of plants. Virus Res. **81:**69–76.
94. **Zambrano, K., O. Carbaloo, F. Geraud, D. Chirinos, C. Fernández, and E. Marys.** 2007. First report of *Tomato yellow leaf curl virus* in Venezuela. Plant Dis. **91:**768.
95. **Zhou, X. P., Y. L. Liu, L. Calvert, C. Munoz, G. W. OtimNape, D. J. Robinson, and B. D. Harrison.** 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.