

Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts

Alison T. M. Lima,¹ Roberto R. Sobrinho,¹ Jorge González-Aguilera,^{1†} Carolina S. Rocha,^{1‡} Sarah J. C. Silva,^{1§} César A. D. Xavier,¹ Fábio N. Silva,¹ Siobain Duffy² and F. Murilo Zerbini¹

Correspondence

Siobain Duffy
duffy@aesop.rutgers.edu
Francisco Murilo Zerbini
zerbini@ufv.br

¹Departamento de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG 36570-000, Brazil

²Department of Ecology, Evolution and Natural Resources, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA

Begomoviruses are ssDNA plant viruses that cause serious epidemics in economically important crops worldwide. Non-cultivated plants also harbour many begomoviruses, and it is believed that these hosts may act as reservoirs and as mixing vessels where recombination may occur. Begomoviruses are notoriously recombination-prone, and also display nucleotide substitution rates equivalent to those of RNA viruses. In Brazil, several indigenous begomoviruses have been described infecting tomatoes following the introduction of a novel biotype of the whitefly vector in the mid-1990s. More recently, a number of viruses from non-cultivated hosts have also been described. Previous work has suggested that viruses infecting non-cultivated hosts have a higher degree of genetic variability compared with crop-infecting viruses. We intensively sampled cultivated and non-cultivated plants in similarly sized geographical areas known to harbour either the weed-infecting *Macrotium yellow spot virus* (MaYSV) or the crop-infecting *Tomato severe rugose virus* (ToSRV), and compared the molecular evolution and population genetics of these two distantly related begomoviruses. The results reinforce the assertion that infection of non-cultivated plant species leads to higher levels of standing genetic variability, and indicate that recombination, not adaptive selection, explains the higher begomovirus variability in non-cultivated hosts.

Received 17 August 2012
Accepted 5 November 2012

INTRODUCTION

Single-stranded DNA begomoviruses (whitefly transmitted members of the family *Geminiviridae*) have become an important factor limiting crop production in tropical and subtropical regions (Morales & Anderson, 2001; Rojas *et al.*, 2005; Seal *et al.*, 2006). Begomoviruses causing cassava mosaic disease (CMD) are the major biotic constraint to cassava cultivation in Africa (Legg & Fauquet, 2004; Legg & Thresh, 2000; Ndunguru *et al.*, 2005; Were *et al.*, 2004). A

complex of at least six different begomovirus species is responsible for the devastating tomato yellow leaf curl disease (TYLCD) (Moriones & Navas-Castillo, 2000). In the Americas, diseases caused by begomoviruses have significantly impacted tomato and bean production since the 1980s (Blair *et al.*, 1995; Brown & Bird, 1992; Gilbertson *et al.*, 1993; Morales & Jones, 2004; Polston & Anderson, 1997).

Previous studies have indicated a high intra- and interspecific diversity of begomoviruses, which can facilitate adaptation to new climates and novel hosts (Monci *et al.*, 2002). Recombination is the most intensively studied population genetic process in begomoviruses, and has been considered more significant than mutation by many researchers (Lefevre *et al.*, 2007a, 2009; Martin *et al.*, 2005, 2011; Monci *et al.*, 2002; Padidam *et al.*, 1999; Pita *et al.*, 2001). It appears to heavily contribute to begomovirus genetic diversity, increasing the evolutionary potential and local adaptation of strains (Berrie *et al.*, 2001; Graham *et al.*, 2010; Harrison & Robinson, 1999; Monci

[†]Present address: Embrapa Trigo, Rodovia BR 285 km 294, CP 451, Passo Fundo, RS, 99001-970, Brazil.

[‡]Present address: FuturaGene Brasil, Avenida José Lembo 1010, Itapetininga, SP, 18210-780, Brazil.

[§]Present address: Dep. de Fitossanidade/CECA, Universidade Federal de Alagoas, Rio Largo, AL, 57100-000, Brazil.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are KC004091–KC004134, KC004068–KC004090 and JX865615–JX865650.

Supplementary material is available with the online version of this paper.

et al., 2002; Padidam *et al.*, 1999). There is ample opportunity for recombination because multiple begomovirus species are often found co-infecting the same plant (Davino *et al.*, 2009; García-Andrés *et al.*, 2006; Harrison *et al.*, 1997; Pita *et al.*, 2001; Ribeiro *et al.*, 2003; Sanz *et al.*, 2000; Torres-Pacheco *et al.*, 1996), and more than one virus can simultaneously replicate in the same nucleus (Morilla *et al.*, 2004). Additionally, the high recombination frequency observed for begomoviruses may be explained by a theoretical recombination-dependent replication mechanism (RDR) (Jeske *et al.*, 2001), in addition to the well-documented rolling circle replication (RCR) (Saunders *et al.*, 2001). However, recent studies have also indicated that begomoviruses can evolve by mutation alone as quickly as RNA viruses (Duffy & Holmes, 2008, 2009), and positive selection – on mutations or the products of recombination events – may also play a role in begomovirus evolutionary dynamics (Monci *et al.*, 2002; Pita *et al.*, 2001; Zhou *et al.*, 1997).

Additionally, host use may play an important role in the standing genetic variability of begomovirus populations (Seal *et al.*, 2006). Several species of non-cultivated plants, especially of the families Malvaceae, Euphorbiaceae, Fabaceae and Solanaceae, are known hosts of begomoviruses (Morales & Anderson, 2001). These weed/wild hosts can serve as reservoirs for infection of nearby crops (Alabi *et al.*, 2008; Barbosa *et al.*, 2009; Bedford *et al.*, 1998; García-Andrés *et al.*, 2006), as overwintering refugia (Alabi *et al.*, 2007, 2008; García-Andrés *et al.*, 2006) and as ‘mixing vessels’ for interspecific coinfection and recombination (García-Andrés *et al.*, 2006; Monde *et al.*, 2010; Silva *et al.*, 2012). Increased host use and diminished bottlenecks would both potentially increase the effective population size of begomovirus populations (Power, 2000; Seal *et al.*, 2006). Although there is limited data on the variability of begomovirus populations in non-cultivated hosts, such data suggest that it is higher than that observed in crop-infecting begomoviruses (Fiallo-Olivé *et al.*, 2012; Silva *et al.*, 2011, 2012; Wyant *et al.*, 2011).

To gather data on the factors affecting genetic variability in begomovirus populations and shed light on whether frequent infection of non-cultivated plants alters viral evolutionary dynamics, we contrasted two populations of distantly related begomoviruses. We intensively sampled crops and non-cultivated plants in similarly sized geographical areas known to harbour either *Macropodium yellow spot virus* (MaYSV) or *Tomato severe rugose virus* (ToSRV). MaYSV is a recently isolated species that was previously only reported in non-cultivated hosts in north-eastern Brazil (Silva *et al.*, 2012), whereas ToSRV is the most widespread tomato-infecting begomovirus in the country (Fernandes *et al.*, 2008; Rocha, 2011; Zerbini *et al.*, 2005). ToSRV can naturally infect other important crops such as chili pepper (Bezerra-Agasie *et al.*, 2006) and potato (Souza-Dias *et al.*, 2008) and is only rarely found in non-cultivated plants such as *Nicandra physaloides* (family Solanaceae) (Barbosa *et al.*, 2009). Over a 3 year period we obtained more than 50 full-length DNA-A

sequences of each of these viruses isolated from a mixture of crops and non-cultivated plants. We compared the molecular evolution and population genetics of the mostly weed-infecting MaYSV to the predominantly tomato-infecting ToSRV. Our results bolster the assertion that infection of indigenous and non-cultivated plant species leads to higher levels of standing genetic variability, apparently driven by higher levels of detectable recombination.

RESULTS

Natural infection of the crop plant *Phaseolus vulgaris* by MaYSV and of the wild host *Sida* spp. by ToSRV

MaYSV was originally described infecting leguminous non-cultivated hosts (*Macropodium lathyroides*, *Calopogonium mucunoides* and *Canavalia* sp.) (Silva *et al.*, 2012). However, in 2011 MaYSV was readily isolated from both *M. lathyroides* and a crop plant, *P. vulgaris* (common bean), in Alagoas. Forty-four DNA-A components were cloned from these samples (Table S1, available in JGV Online). Pairwise comparisons of the DNA-A sequences (data not shown) revealed that only one begomovirus was present, with 90–99 % identity to MaYSV. Therefore, based on the criteria established by the *Geminiviridae* Study Group of the ICTV (Brown *et al.*, 2012) these genomic components represent additional isolates of this species. MaYSV appeared to be the prevalent begomovirus infecting common bean in this portion of Alagoas, since *Bean golden mosaic virus* was not detected in *P. vulgaris* in our study. Overall, 56 whole genomes were included in the MaYSV dataset, with 23 (41 %) isolated from non-cultivated hosts.

Despite sampling many non-cultivated species neighbouring infected tomato plants, we were only able to find ToSRV in two samples of the malvaceous non-cultivated host, *Sida* sp. Therefore, 53 out of 55 whole ToSRV genomes were isolated from tomatoes and only two from non-cultivated hosts.

MaYSV population is more variable than the ToSRV population

Although the ToSRV and MaYSV populations were sampled in similarly sized geographical areas, they differed profoundly in their levels of standing genetic variability. While the average pairwise number of nucleotide differences (π) for the full-length DNA-A of the ToSRV population was 0.0084 (Table 1), this same statistic was about eightfold higher for the MaYSV population ($\pi=0.0658$). Interestingly, the variability within the MaYSV population was not evenly distributed throughout the genome. The MaYSV Rep gene was 3.8 times more variable than its CP gene, and more than 10-fold higher than the ToSRV Rep (Table 1). Within the MaYSV Rep, the N-terminal half was much more variable than the C-terminal half – again, a distribution of variation

Table 1. Genetic variability of the begomoviruses MaYSV and ToSRV

H_d =Haplotype diversity, π =pairwise, per-site nucleotide diversity.

Population	No. sequences	DNA-A H_d	DNA-A π	CP π	Rep π
ToSRV (total)	55	0.999 (± 0.004)	0.00844 (± 0.00136)	0.00963 (± 0.00238)	0.00743 (± 0.00140)
MaYSV (total)	56	1.000 (± 0.003)	0.06580 (± 0.00231)	0.02852 (± 0.00322)	0.10889 (± 0.00456)
<i>P. vulgaris</i>	33	1.000 (± 0.007)	0.06264 (± 0.00300)	0.02189 (± 0.00102)	0.11175 (± 0.00599)
Non-cultivated hosts	23	1.000 (± 0.013)	0.06702 (± 0.00425)	0.03616 (± 0.00670)	0.10547 (± 0.00882)

not shared by the ToSRV Rep (Fig. 1). The nucleotide diversity of the ToSRV CP gene was similar to that of MaYSV.

We also compared the variability between MaYSV isolates sampled from crops and non-cultivated hosts. Each MaYSV subset was still markedly more variable than the whole ToSRV population, with isolates sampled from *P. vulgaris* and non-cultivated hosts showing similar levels of standing genetic variability ($\pi=0.0626$, $SD=0.003$ and $\pi=0.0670$, $SD=0.0043$, respectively).

Phylogenetic analysis

The ToSRV CP and Rep ML phylogenetic trees are not congruent (Fig. 2), but the topological differences between them were due to a strongly supported recombination event in the CP gene of ToSRV-[BR:Vic20:10] ($P=2.18 \times 10^{-10}$, Table 2). Without this isolate, the clustering of isolates in the CP and Rep trees are identical (Fig. S1), mimicking the clustering found in the Rep tree on Fig. 2. Both ML trees showed strong support for isolates from Florestal being a separate population from those from Carandaí, Jaíba and Viçosa (Fig. 2). There is rough clustering of isolates from each sampling site together, with Jaíba being nested within isolates from Viçosa. However, the lack of bootstrap support for geographical structure among these three sites (and the weakly supported clustering of Vic06 and Vic07 with isolates from Carandaí) suggests that migration may occur between

these subpopulations. Consistent with our phylogenetic analyses, Wright's fixation index F , based on the ToSRV CP dataset, indicated genetic differentiation amongst isolates sampled from different geographical locations ($F_{ST}=0.51066$).

The MaYSV CP and Rep trees were highly incongruent (Fig. 3). The MaYSV Rep tree shows four clades with significant genetic distance between them (two with 100% bootstrap support), and the CP shows fewer well-supported clades and less genetic variability among isolates. Both trees showed little evidence of geographical structuring. In fact, Wright's fixation index F based on the MaYSV CP dataset indicated less evidence of genetic differentiation ($F_{ST}=0.11428$) than the ToSRV population.

MaYSV isolates have a mixture of different recombinant patterns

While ToSRV showed little evidence that recombination has significantly contributed to its evolution (Table 2), MaYSV showed both inter- and intraspecific recombination events. Corroborating the conflicting MaYSV CP and Rep phylogenies, we identified a total of six unique potential recombination events in the MaYSV population. Most events involved breakpoints located inside the Rep gene and in the common region (events 1, 2, 3, 4 and 6; Table 2), though one event had breakpoints within the CP gene (event 5; Table 2). Isolates that showed evidence of a

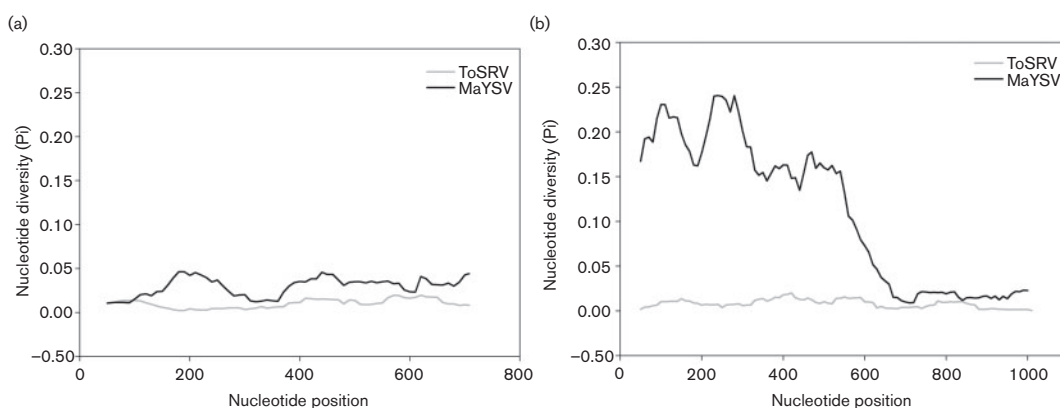


Fig. 1. Mean pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated on a sliding window across the CP (a) and Rep (b) sequences from ToSRV (grey line) and MaYSV (black line) populations.

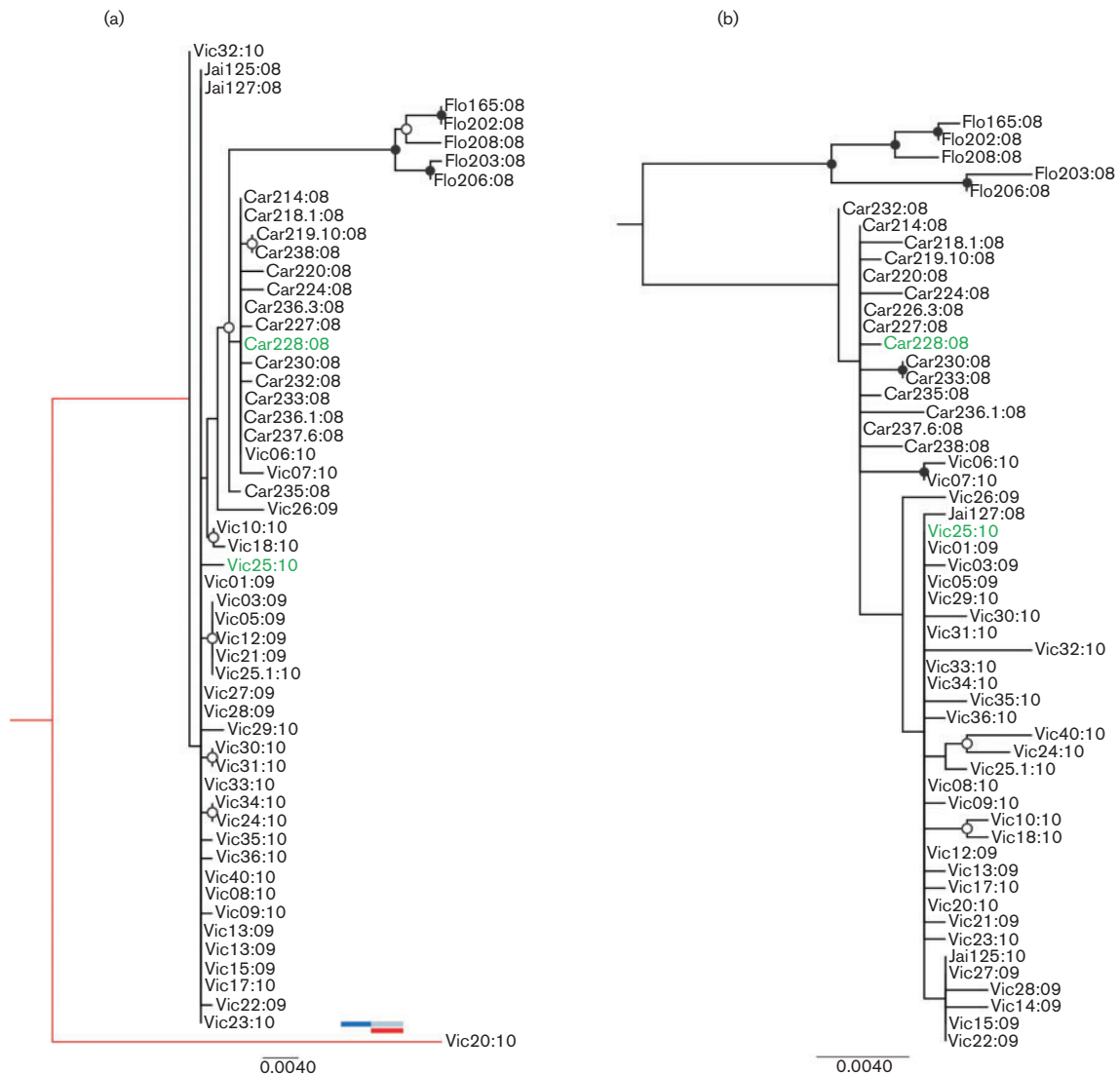


Fig. 2. Midpoint-rooted ML trees based on the CP (a) and Rep (b) nucleotide sequences of isolates from the ToSRV population. Nodes with bootstrap values equal or higher than 80% are indicated by filled circles, and those with values lower than 80 and higher than 50% by empty circles. The contribution of parental sequences to the unique recombination event detected within the CP coding sequence (event 1, isolate Vic20; Table 2) is shown as a diagram above the branch where the substitutions due to recombination were mapped (in red colour). Isolates sampled from non-cultivated hosts are shown in green.

shared recombination event in the Rep sequence were readily identified in well supported clades observed on the Rep ML phylogenetic tree (Fig. 3). There were no strong geographical patterns associated with recombination events. For instance, four out of five recombination events detected in MaYSV Rep (events 1, 2, 4 and 6) were detected in isolates sampled in Olho d'Água das Flores, but two of these events (events 1 and 6) were observed elsewhere.

Adaptive selection does not explain the higher begomovirus variability in non-cultivated hosts

We investigated the extent that positive and negative selection at the amino acid level had shaped the standing

genetic variability in the CP and Rep datasets (comprised of all isolates obtained from cultivated and non-cultivated hosts) of both viruses. All datasets showed d_N/d_S ratios (ω) lower than 1 (Table 3), indicating negative selection. However, the wide variation of the values indicated that each gene/population might be under different selective constraints. The ω value for the ToSRV CP dataset (0.446493) was considerably higher than that observed for the MaYSV CP (0.0514088). The high ω of the ToSRV CP was observed even with the elimination of the single recombinant isolate, Vic20 ($\omega=0.27438$). The ω value for the ToSRV Rep (0.268859) was only slightly higher than that observed for the MaYSV Rep dataset (0.208195).

Table 2. Recombination events detected by RDP in the DNA-A of ToSRV and MaYSV populations

Event	Recombinant	Recombination breakpoints		Parents		Methods†	P-value‡				
		Initial	Final	Major	Minor						
ToSRV 1	Vic20:10	537	1011	Unknown	HV25:10	GBMCS <u>3</u>	2.18×10^{-10}				
MaYSV 1	Oaf12:11	2654	1661	Unknown	Pdi1:10	GBMCS <u>3</u>	6.07×10^{-60}				
	Bat1:10			Unknown	Oaf1:10						
	Inp1:10			Unknown	Oaf2:10						
	Pir1:10			Unknown	Ced1:09						
	Crb8:11			Deg1:10	Ced1:09						
	Crb10:11			Inp2:10	Ced1:09						
	Crb13:11			Mac5:10	Ced1:09						
	Crb14:11			Crb9:11	Ced1:09						
	Crb15:11			Crb11:11	Ced1:09						
	Crb16:11			Crb12:11	Ced1:09						
	Oaf16:11			Crb17:11	Ced1:09						
	Oaf17:11			Crb18:11	Ced1:09						
	Oaf20:11			Crb19:11	Ced1:09						
	Oaf21:11			Oaf15:11	Ced1:09						
	Oaf22:11			Oaf18:11	Ced1:09						
	Oaf23:11			Oaf19:11	Ced1:09						
	Oaf1:11			Crb1:11	Ced1:09						
	Oaf2:11			Crb2:11	Ced1:09						
	Oaf7:11			Oaf6:11	Ced1:09						
	Oaf3:11			Oaf9:11	Ced1:09						
	Crb3:11			Crb5:11	Ced1:09						
	Oaf4:11, Oaf5:11, Oaf8:11, Crb4:11, Oaf10:11, Oaf11:11, Oaf14:11, Oaf24:11, Oaf25:11, Oaf25:11, Crb7:11			Oaf13:11	Ced1:09						
				Crb6:11	Ced1:09						
2	Oaf19:11			(?) 1662	2039			Pdi1:10	Oaf24:11	GBMS <u>3</u>	8.50×10^{-23}
	Oaf6:11							Oaf1:10	Bat1:10		
	Oaf9:11							Oaf2:10	Inp1:10		
3	Inp1:10	1984	2648	Oaf14:11	Crb8:11	BS <u>3</u>	4.54×10^{-35}				
4	Oaf25:11	1798	(?) 2653	Oaf24:11	Unknown	GBMCS <u>3</u>	3.51×10^{-22}				
	Oaf22:11			Bat1:10	Unknown						
	Oaf7:11			Pir1:10	Unknown						
	Oaf8:11			Crb8:11	Unknown						
	Oaf10:11			Crb10:11	Unknown						
	Oaf11:11			Crb13:11	Unknown						
	Oaf12:11			Crb14:11	Unknown						

Table 2. cont.

Event	Recombinant	Recombination breakpoints			Parents		Methods†	P-value‡
		Initial	Final	Major	Minor			
5	Oaf14:11			Crb15:11	Unknown	GBM3	1.09 × 10 ⁻⁰⁶	
	Mac1:10	415	1114	Oaf9:11	Unknown			
	Bas1:09			Oaf2:10	Unknown			
	Pir1:10	(?) 2159	(?) 2653	Oaf14:11	Unknown			
6	Bat1:10			Oaf22:11	Unknown	MCS	1.39 × 10 ⁻²⁰	
	Imp1:10			Oaf7:11	Unknown			
	Crb8:11			Oaf8:11	Unknown			
	Crb10:11			Oaf10:11	Unknown			
	Crb13:11			Oaf11:11	Unknown			
	Crb14:11			Oaf12:11	Unknown			
	Crb15:11, Crb16:11, Oaf16:11, Oaf17:11, Oaf20:11, Oaf21:11,			Oaf25:11	Unknown			
	Oaf23:11, Oaf1:11, Oaf2:11, Oaf3:11, Crb3:11, Oaf4:11, Oaf5:11,							
Crb4:11, Oaf24:11, Crb7:11								

*Numbering starts at the first nucleotide after the cleavage site at the origin of replication and increases clockwise. (?) Indicates that the breakpoint could not be precisely pinpointed.

†R, RDP; G, GeneConv; B, Bootscan; M, MaxChi; C, Chimaera; S, SisScan; 3, 3Seq.

‡The reported P-value is for the program in bold, underlined type and is the lowest P-value calculated for the event in question.

Although GARD was unable to detect the well-supported recombination event in the ToSRV CP dataset, no sites were under statistically significant negative selection, including or excluding Vic20 (Fig. 4 and Fig. S2). Excluding the recombinant Vic20 from the dataset, Random Effects Likelihood (REL) detected 20 sites under positive selection. However, the evidence for selection on individual sites is weak, as no positively selected sites were detected using the Single Likelihood Ancestor Counting (SLAC) or Partitioning for Robust Inference of Selection (PARRIS) methods. Two negatively selected sites were detected by the SLAC method in the ToSRV Rep dataset (Fig. 4). In addition, nine positively selected sites were detected by REL (Fig. S2), with six of them located between motif III in the catalytic domain and motif Walker A in the ATPase domain. PARRIS did not identify any sites under positive selection.

Recombinant partitions were readily detected by GARD in the MaYSV CP and Rep datasets. Nevertheless, a higher number of sites were under detectable selection in both MaYSV genes. In the CP dataset, all codons were identified as being under negative selection by REL (Fig. S2). In the MaYSV Rep, 106 sites were shown to be under negative selection by SLAC or REL, and 12 sites were under positive selection by REL (Fig. 4, Fig. S2). Many sites in the highly variable Rep N-terminal showed high synonymous substitutions rates with statistical evidence of negative selection (Fig. 4). Although GARD was able to detect recombination events in this dataset, none of the positively selected sites detected by REL correlated with the host from which the isolates were obtained (non-cultivated vs cultivated), but were clearly related to a given recombination event, indicating that they most probably represent a spurious selection signal due to recombination. Results obtained by PARRIS confirmed the absence of sites under positive selection.

Together, our results indicate that adaptive selection at the amino acid level is not driving the higher variability in the MaYSV population.

Relative contribution of mutation and recombination to the genetic variability of ToSRV and MaYSV populations

We mapped all substitutions over the branches in our midpoint-rooted ML trees and determined which branches were associated with well-supported recombination events. As there were no detectable recombination events in the ToSRV Rep dataset, all of the substitutions along the phylogeny are presumably due to mutation. Out of a total of 99 substitutions over the ToSRV CP phylogeny, the single recombination event detected (spanning 50 % of the CP gene) encompasses 42 substitutions. Therefore, 42.4 % of the standing genetic variability in the ToSRV CP could be attributed to recombination (Table 4). In the MaYSV CP tree, there was a total of 313 substitutions, and 49 substitutions were located inside a putatively recombinant

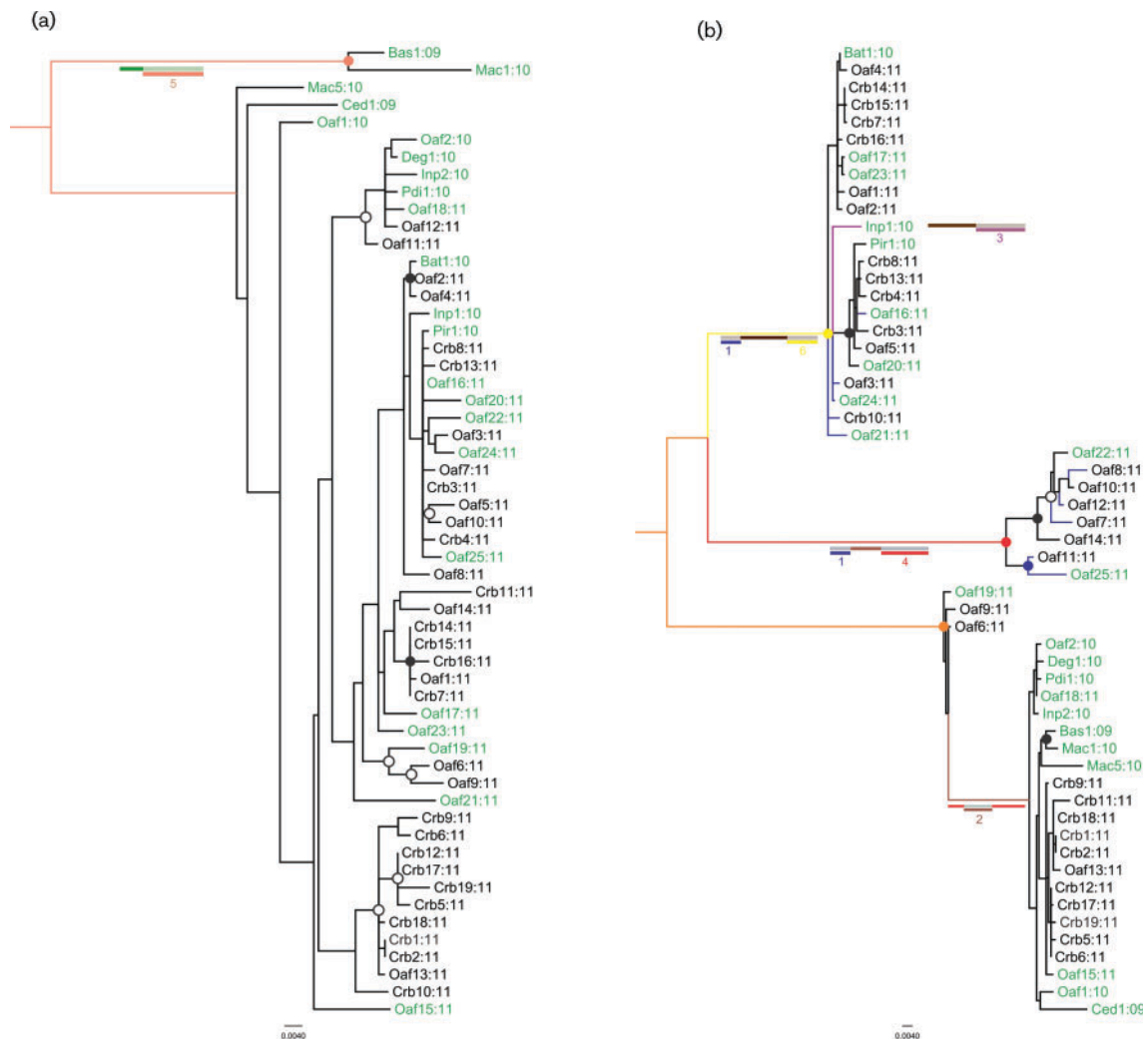


Fig. 3. Midpoint-rooted ML trees based on the CP (a) and Rep (b) nucleotide sequences of isolates from the MaYSV population. Nodes with bootstrap values equal or higher than 80 % are indicated by filled circles, and those with values lower than 80 and higher 50 % by empty circles. The unique recombination events detected within the CP (event 5) and Rep sequences (events 1–4 and 6, Table 2) are shown as diagrams close to the branches where the substitutions due to each recombination event were mapped. Branches in blue, brown, purple, red, pink and yellow colours were assigned to the events 1–6, respectively. Additional substitutions due to the recombination events 4 and 6 were mapped over the orange branch. Isolates sampled from non-cultivated hosts are shown in green.

Table 3. Mean values of non-synonymous to synonymous substitution ratios (d_N/d_S) for the CP and Rep genes of ToSRV and MaYSV

Population	CP	Rep
ToSRV	0.446493/0.27438*	0.268859
MaYSV	0.0514088	0.208195

*Excluding the recombinant isolate Vic20:10.

region (spanning about 75 % of the CP gene) on a single branch of the tree. In this considerably more variable dataset, recombination accounted for only 16 % of the substitutions (Table 4).

As expected, a considerably higher number of substitutions was observed for the MaYSV Rep ML tree: 553 substitutions. The unique recombination events 2, 4 and 6 were readily assigned to the long branches (the brown, red and yellow branches, respectively, in Fig. 3b), with another long branch leading to the clade containing events 4 and 6 (the orange branch; Fig. 3b). For events 2 (shared by isolates

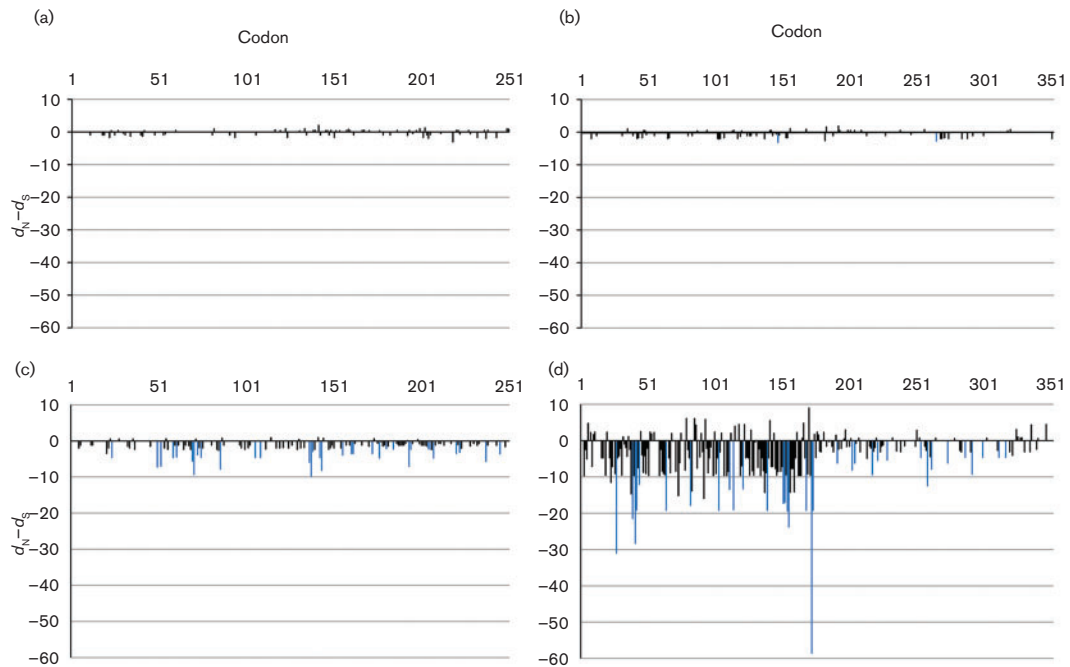


Fig. 4. Difference between the non-synonymous and synonymous substitution rates (d_N-d_S) calculated across the codons of the CP (a, c) and Rep sequences (b, d) of isolates from the ToSRV (a, b) and MaYSV (c, d) populations using the SLAC method. Sites with statistical evidence of negative selection are indicated by blue bars, while neutrally evolving sites are indicated by black bars.

Oaf6, Oaf9 and Oaf19) and 4 (shared by all isolates in clade II), 22 and 36 substitutions were counted as associated to them, respectively. Event 6 (shared by all 23 isolates in clade I) accounted for the largest individual contribution amongst the unique recombination events: 37 substitutions. Additionally, 16 changes on the branch leading to the clades associated with events 4 and 6 were either in the region of overlap between events 4 and 6, or within the larger event 6, and were also assigned to $\eta_{\text{recombination}}$ (η_r) (Table 4). Event 3 (assigned to the purple terminal branch) was exclusively detected in the Inp1 isolate and a total of 12 substitutions were counted in association with this event.

Although event 1 was well supported by the methods of analysis contained in RDP, the exact location of the breakpoint was variable and frequently ended in the intergenic region instead of the Rep gene. Only those substitutions in the terminal branches leading to sequences whose end breakpoint of this event occurred inside the Rep

gene (isolates Oaf3, Oaf7, Oaf8, Oaf11, Oaf12, Oaf16, Oaf21, Oaf24, Oaf25, Crb10 and Inp1) were added to the η_r . A total of five substitutions could be associated with this event.

Overall, 201 substitutions were counted as due to recombination over the MaYSV Rep phylogeny, which represented a relative contribution of 36.3% to the standing genetic variability (Table 4).

DISCUSSION

Most studies on the diversity of begomovirus populations have focused on species diversity (Ala-Poikela *et al.*, 2005; Bull *et al.*, 2006; Fernandes *et al.*, 2008; García-Andrés *et al.*, 2006; Lefeuvre *et al.*, 2007b; Lozano *et al.*, 2009; Ndunguru *et al.*, 2005; Reddy *et al.*, 2005; Ribeiro *et al.*, 2003; Rothenstein *et al.*, 2006; Sserubombwe *et al.*, 2008). In

Table 4. Relative contribution of mutation and recombination to the standing genetic variability of ToSRV and MaYSV populations

Population	CP			Rep		
	η_r	η_μ	η_{total}	η_r	η_μ	η_{total}
ToSRV	42 (42.4%)	57 (57.6%)	99	0	92 (100%)	92
MaYSV	49 (15.7%)	264 (84.3%)	313	201 (36.3%)	352 (63.7%)	553

contrast, little is known about the standing genetic variability within species, especially in begomovirus populations in non-cultivated hosts. Here, we present a study contrasting the molecular variability of begomovirus populations of the mostly weed-infecting MaYSV and the predominantly tomato-infecting ToSRV, and discuss the relative contribution of evolutionary processes on this variability. Our results, based on 111 viral genome sequences cloned from samples collected over a 3 year period, support the hypothesis that although the genetic structure of viral populations is modulated by the common processes of mutation, recombination, selection and genetic drift (García-Arenal *et al.*, 2003; Roossinck, 2003), the relative contribution of each process can vary widely among begomovirus populations.

The MaYSV population was notably more variable than the ToSRV population, largely due to a number of recombination events in the Rep gene. Even MaYSV isolates sampled from the crop *P. vulgaris* (collected from two locations, Craibas and Olho d'Água das Flores) were considerably more variable than all ToSRV isolates sampled from tomato plants (collected from four different locations). Previous studies have indicated that begomovirus populations infecting non-cultivated hosts are more variable than crop-infecting ones (Rocha, 2011; Silva *et al.*, 2012). The standing genetic variabilities of four tomato-infecting begomovirus populations were much lower than that of a begomovirus population from a non-cultivated host (*Blainvillea yellow spot virus*) sampled in south-eastern Brazil (although these results are based on different sample sizes) (Rocha, 2011). The genetic structure was also determined for a population of *Cleome leaf crumple virus* infecting an annual weed (*Cleome affinis*, family Capparaceae) often associated with leguminous crops in Brazil (Silva *et al.*, 2011). Several recombination events and a high molecular variability were estimated for this population.

Although an increasing body of evidence points to a higher genetic variability in begomovirus populations infecting non-cultivated hosts, all previous studies have detected whether various evolutionary processes have contributed to the population genetic variability without attempting to determine their relative contribution to the total variability in these populations. Our novel substitution counting method allowed us to assess the relative contribution of recombination and mutation to the standing genetic variability of two begomoviruses – a step towards disambiguating the effects of various evolutionary processes on viral genetic variation.

Recombination and the genetic structure of begomovirus populations

Recombination is the most studied evolutionary process acting on geminivirus populations and has greatly contributed to their evolution (Bridson *et al.*, 1996; García-Andrés *et al.*, 2007a; García-Andrés *et al.*, 2007b; Padidam *et al.*, 1999; Pita *et al.*, 2001; Schnippenkoetter *et al.*, 2001).

In begomoviruses, the non-random location of recombination breakpoints is conserved amongst mono- and bipartite genomes, with hot spots in the Rep N-terminal portion and in the 5'-end of the common region (Lefevre *et al.*, 2007a, b; Prasanna & Rai, 2007). Our analyses of the MaYSV population indicated that most unique recombination events involved the Rep gene, with at least one breakpoint located in the N-terminal portion and/or the 5'-end of the common region. In addition, we also observed that the N-terminal portion accounted for the largest fraction of the total variability in the Rep gene, suggesting that recombination may be the evolutionary process responsible for this uneven distribution of polymorphisms.

It has been shown that recombination events that preserve co-evolved intragenome interactions (protein–protein and/or protein–DNA) are favoured by selection (Martin *et al.*, 2011). In fact, the linkage between the Rep N-terminal portion and the 5' common region was not broken by recombination events detected in the MaYSV population. These portions were exchanged as blocks that preserve them as originating from the same genetic background. This could increase the frequency in which viable recombinants for the Rep sequence are maintained by preserving the interaction between the catalytic domains (in the Rep N-terminal portion) and their cognate iterative elements located in the 5' portion of the common region (Martin *et al.*, 2011). Although a reduced number of breakpoints occur within the CP-encoding sequence, most of them are located in the central portion of the gene (Lefevre *et al.*, 2007b). In agreement, two unique events (the single event detected in the ToSRV population and the MaYSV event 5) had initial breakpoints located in the central portion of the gene and spanned its entire C-terminal portion.

Purifying selection in viral genes

Although previous studies indicate that the main type of selection acting on the CP and Rep genes in begomovirus populations is purifying selection (García-Andrés *et al.*, 2007a; Sanz *et al.*, 1999; Silva *et al.*, 2011, 2012), we also assessed the possible contribution of adaptive selection in shaping the standing genetic variability in the ToSRV and MaYSV populations, since a small fraction of codons could be under different selective constraints. Interestingly, the d_N/d_S ratio estimated for the ToSRV CP was markedly high ($\omega=0.446$), even after the exclusion of the recombinant isolate Vic20 ($\omega=0.274$). These values are high even when compared with non-vector-borne ssRNA viruses such as *Prune dwarf virus* (family *Bromoviridae*; $\omega=0.222$) and similar to *Pepper mild mottle virus* (genus *Tobamovirus*; $\omega=0.301$) (Chare & Holmes, 2004). Despite the lower molecular variability, REL detected stronger evidence of adaptive selection in the CP (two sites) and especially in the ToSRV Rep (nine sites). Interestingly, out of nine positively selected sites in the ToSRV Rep, six were located between motif III in the catalytic domain and motif Walker

A in the ATPase domain. This region is putatively involved in Rep oligomerization and interaction with host factors (Hanley-Bowdoin *et al.*, 2004; Orozco *et al.*, 2000). Although REL is considerably less conservative than SLAC in detecting positively selected sites, it is important to note that the Rep ToSRV dataset was free of any detectable evidence of recombination that is known to affect the results of selection analyses.

In striking contrast, we did not detect any evidence of adaptive selection acting on both MaYSV genes, indicating that adaptive selection did not contribute to the higher molecular variability in the MaYSV population.

Although a higher variability was found in the Rep N-terminal portion (probably due to recombination) it was composed mostly of synonymous substitutions. The Rep catalytic domain is located in the N-terminal portion and includes three motifs which are conserved in proteins involved in RCR (motif I: FLTY; motif II: HxH; and motif III: YxxxV) (Ilyina & Koonin, 1992). Adjacent to these elements are the DNA-binding specificity determinants involved in high affinity DNA-binding (Londoño *et al.*, 2010). The integrity of these elements in terms of amino acid sequence and structure is important for viral replication and therefore suggests that although the nucleotide sequence in the 5' portion of the Rep gene may harbour a high variability, most of these substitutions would tend to preserve the amino acid sequence and consequently protein structure as well.

Host preference and recombination

The lower synonymous site variation in ToSRV population suggests that this population may have been recently introduced in the sampled geographical areas (either from imported tomatoes or from a wild host) and/or, less likely, experienced a recent genetic bottleneck. Low levels of molecular variability were also observed in genetically differentiated subpopulations of begomoviruses causing TYLCD in Spain and Italy. This scenario was consistent with the foundation of these subpopulations by few variants of limited genetic variability (García-Andrés *et al.*, 2007a). It is possible that the lack of alternative perennial hosts that would maintain high ToSRV effective population sizes between cultivation periods results in a strong reduction of variability. MaYSV is able to efficiently infect both *P. vulgaris* and perennial leguminous non-cultivated hosts. These latter hosts could represent an important epidemiological component for the MaYSV population for allowing that genetic variability is maintained between cultivation periods. Furthermore, by infecting new hosts these populations may be subject to environments inaccessible to viruses that have a narrow host range (Power, 2000). As a consequence, these viruses could experience high frequencies of interspecific recombination, resulting in a rapid evolution of these populations. Studies attempting to assess the genetic variability of begomovirus populations on *Solanum nigrum* (an indigenous weed in Spain) showed that this host is an

efficient reservoir of all previously characterized begomoviruses species and strains causing TYLCD (García-Andrés *et al.*, 2006). Additionally, an uncharacterized recombinant isolate was detected in this host. In fact, several novel species of begomoviruses causing TYLCD are actually interspecific recombinants (Monci *et al.*, 2002; Navas-Castillo *et al.*, 2000). Begomoviruses causing CMD are the major biotic constraint to cassava cultivation in Africa (Legg & Fauquet, 2004; Legg & Thresh, 2000; Ndunguru *et al.*, 2005; Were *et al.*, 2004), and multiple natural reservoirs for begomoviruses causing CMD have been described previously (Alabi *et al.*, 2007, 2008). Associated with the ability to infect many alternative hosts, these viruses appear to evolve largely by intra- and interspecific recombination (Berrie *et al.*, 2001; Fondong *et al.*, 2000; Pita *et al.*, 2001; Tiendrébogo *et al.*, 2012; Zhou *et al.*, 1997).

Mutation accounts for most begomovirus genetic variability

Our results suggest that recombination explains the higher molecular variability of the MaYSV population compared with the ToSRV population. However, both populations were dominated by mutational diversification. By mapping the recombination events onto ML trees, we were able to distinguish the individual contributions associated with these two mechanisms that create variability. Then, we took advantage of the additive character of η to express the individual contributions of these processes as fractions of the total variability ($\eta_{\text{total}} = \eta_r + \eta_\mu$). This simple statistic is not without its flaws, chiefly that recombination is known to reduce the accuracy of the phylogenetic trees upon which this method relies (Posada & Crandall, 2002). However, phylogeny independent measures of population variability, such as pairwise comparisons, are non-additive, which complicates partitioning variability due to recombination and mutation.

Some of the variation that our method attributes to mutation is undoubtedly due to recombination that was not statistically detectable. Therefore, our estimates should be interpreted as the minimal relative contribution of recombination. Furthermore, while our method attributes 100% of changes on specified branches within identified recombination breakpoints to recombination, this is not a liberal assumption. The reconstructed sequence of the recombinant region that finds each recombinant clade may not be the correct ancestral state, but it is unlikely to contain dramatically more substitutions than the actual first recombinant. All subsequent changes that occur in that region on branches to descendent nodes are counted as mutations.

Using this novel partitioning, we assessed recombination as the source of up to 50% of the variation in begomovirus populations. Importantly, our results indicate that its relative contribution to the total variability does not necessarily correlate with the number of detectable recombination events: the highest relative contribution of

recombination was in the ToSRV CP, which was the least variable dataset analysed. On the other hand, results from the MaYSV Rep indicate that in recombination hotspots, recombination may be at least as important as mutation in generating variability.

We have assessed the molecular variability of two Brazilian begomovirus populations, and found that in certain portions of the genome recombination could be responsible for a considerable fraction of the total variability in addition to mutation. Additionally, we also found evidence that the distinct evolutionary patterns might, at least in part, be related to the ability to infect non-cultivated, perennial hosts. Such ability would thus be an important factor in the rapid evolution of these populations.

METHODS

Sequence datasets. A total of 55 full-length DNA-A sequences corresponding to ToSRV and 56 corresponding to MaYSV were analysed [GenBank accession numbers: JN419005, JN419007, JN419009, JN419012–JN419016, JN419018–JN419020, JN419022 and KC004091–KC004134 (MaYSV), KC004068–KC004090 and JX865615–JX865650 (ToSRV)]. The ToSRV dataset comprised isolates obtained from tomato (53 isolates) and *Sida* spp. plants (two isolates) collected in three different municipalities in the state of Minas Gerais (MG) from 2008 to 2010 (Table S1; González-Aguilera *et al.*, 2012). Twelve of the MaYSV sequences were isolated from leguminous non-cultivated hosts collected at nine municipalities of the states of Alagoas (AL), Sergipe (SE) and Paraíba (PB) in 2009 and 2010 as described previously (Silva *et al.*, 2012), and 44 additional sequences were obtained from *M. lathyroides* and the crop *P. vulgaris* (common bean) samples showing typical symptoms of begomovirus infection collected in Olho d'Água das Flores and Craibas (AL) in 2011 (Table S1). The sampling areas in MG and AL/SE/PB have similar sizes (approximately 45 000 km²).

Total DNA was extracted from fresh tissue or herbarium-like (pressed and dried) samples as described by Doyle & Doyle (1987). Full-length viral genomes were amplified by rolling-circle amplification, according to the standard circular DNA cloning method (Inoue-Nagata *et al.*, 2004). Single genome-length fragments were excised from these concatamers with *Bam*HI and ligated into the pBLUESCRIPT-KS+ (Stratagene) plasmid vector, previously cleaved with the same enzyme. Viral inserts were commercially sequenced (Macrogen Inc.) by primer walking. All genome sequences were organized to begin at the nicking site in the invariant nonanucleotide at the origin of replication (5'-TAATATT//AC-3').

Multiple sequence alignments and phylogenetic analysis. Multiple sequence alignments were prepared for the full-length DNA-A and for the CP and Rep coding sequences of each viral species (ToSRV and MaYSV) using MUSCLE (Edgar, 2004). ML trees were inferred for CP and Rep sequences using PAUP* v. 4.0 (Swofford, 2003). These genes were chosen due to their essential role for insect transmission and viral replication, respectively. Additionally, they encompass about 70% of the full-length DNA-A genome. The best-fit model of nucleotide substitution was determined using MODELTEST (Posada & Crandall, 1998) by the Akaike Information Criterion (Table S2). The heuristic ML search was initiated with a neighbour-joining tree, with optimization using the tree-bisection-reconnection algorithm. The robustness of each individual branch was estimated from 2000 nearest neighbour interchange bootstrap replicates. Trees were visualized and edited using FigTree (tree.bio.ed.ac.uk/software/figtree).

Variability indices. The haplotype diversity (H_d), the mean pairwise number of nucleotide differences per site (nucleotide diversity, π) and Wright's F fixation index were calculated using DnaSP v. 5.10 (Rozas *et al.*, 2003). The π statistic was also calculated on a sliding window of 100 bases, with a step size of 10 bases in order to estimate the nucleotide diversity across the length of the CP and Rep datasets.

Recombination analysis. Recombination analysis was performed using the RDP, Geneconv, Bootscan, Maximum Chi Square, Chimaera, SisterScan and 3Seq methods as implemented in Recombination Detection Program (RDP) version 3.0 (Martin *et al.*, 2010) (RDP project files available from the authors upon request). Alignments were scanned with default settings for the different methods. Statistical significance was inferred by *P*-values lower than a Bonferroni-corrected cut-off. Only recombination events detected by at least three of the analysis methods available in the program were considered reliable.

Detection of positive and negative selection at amino acid sites. To detect sites under positive and negative selection, we analysed the CP and Rep datasets using three ML-based methods implemented in DataMonkey (www.datamonkey.org): SLAC, REL and PARRIS (Kosakovsky Pond & Frost, 2005; Scheffler *et al.*, 2006). As recombinant sequences cause spurious selection results, we searched for breakpoints in sequences implicated as recombinant by GARD prior to running these analyses. Additionally, PARRIS allowed synonymous rate variation, topology and branch lengths to vary across recombination breakpoints. The SLAC method was also used to estimate the mean ratios of non-synonymous to synonymous substitutions (d_N/d_S) for the CP and Rep datasets from both begomovirus populations based on the inferred GARD phylogenetic trees. These methods were applied under the nucleotide substitution models determined in MODELTEST as described in Table S2. Bayes factors larger than 50 and *P*-values smaller than 0.1 were used as thresholds for the REL method.

Relative contribution of recombination and mutation in the genetic variability of ToSRV and MaYSV populations. A phylogeny based approach was developed to determine the relative contribution of recombination and mutation to the total variability observed in the ToSRV and MaYSV populations. We identified groups of sequences descended from a shared recombination event (based on RDP analysis), and frequently these formed clades on midpoint-rooted CP and Rep ML trees. The sequence of the ancestral node of each of these clades reflected the recombination event (ancestral state reconstruction by marginal ML method in PAUP*), but its parental node did not. Wherever possible, we assigned each recombination event to the branch between these nodes. We did not consider the minor/major parent information from RDP analyses to be definitive, and instead considered the portion of the sequence that differed the most from the parental nodes on the ML tree to be the recombined, introduced block. In the case of recombination events associated with only one sequence, the event was assigned to the branch leading to the corresponding tip. Using PAUP*, we identified all substitutions that occurred over each ML tree, and calculated η , the total number of substitutions over the phylogeny (Fu & Li, 1993). We then looked at the location within the gene of substitutions that occurred on branches associated with recombination. Substitutions that occurred in the region likely introduced by recombination were added to η_r , the total number of substitutions on the phylogeny likely due to recombination. All other substitutions, including those on branches associated with recombination but outside of the region implicated in the recombination event, were added to η_{mutation} (η_m). Thus, $\eta = \eta_r + \eta_m$.

ACKNOWLEDGEMENTS

The authors wish to thank Eduardo Mizubuti and Elizabeth Fontes for their continuous and invaluable support. A. T. M. L. was the

recipient of a CNPq doctoral fellowship. S.D. was supported by the NSF (DEB 1026095). F.M.Z. is supported by CNPq (484447/2011-4) and Fapemig (APQ-949-09).

REFERENCES

- Ala-Poikela, M., Svensson, E., Rojas, A., Horko, T., Paulin, L., Valkonen, J. P. T. & Kvarnheden, A. (2005). Genetic diversity and mixed infections of begomoviruses infecting tomato, pepper and cucurbit crops in Nicaragua. *Plant Pathol* **54**, 448–459.
- Alabi, O. J., Ogbé, F. O., Bandyopadhyay, R., Dixon, A. G., Hughes, J. & Naidu, R. A. (2007). The occurrence of African cassava mosaic virus and East African cassava mosaic Cameroon virus in natural hosts other than cassava in Nigeria. *Phytopathology* **97**, S3.
- Alabi, O. J., Ogbé, F. O., Bandyopadhyay, R., Lava Kumar, P., Dixon, A. G. O., Hughes, J. D. & Naidu, R. A. (2008). Alternate hosts of African cassava mosaic virus and East African cassava mosaic Cameroon virus in Nigeria. *Arch Virol* **153**, 1743–1747.
- Barbosa, J. C., Barreto, S. S., Inoue-Nagata, A. K., Reis, M. S., Firmino, A. C., Bergamin Filho, A. & Rezende, J. A. M. (2009). Natural infection of *Nicandra physaloides* by *Tomato severe rugose virus* in Brazil. *J Gen Plant Pathol* **75**, 440–443.
- Bedford, I. D., Kelly, A., Banks, G. K., Briddon, R. W., Cenis, J. L. & Markham, P. G. (1998). *Solanum nigrum*: an indigenous weed reservoir for a tomato yellow leaf curl geminivirus in southern Spain. *Eur J Plant Pathol* **104**, 221–222.
- Berrie, L. C., Rybicki, E. P. & Rey, M. E. C. (2001). Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst begomoviruses. *J Gen Virol* **82**, 53–58.
- Bezerra-Agassie, I. C., Ferreira, G. B., Ávila, A. C. & Inoue-Nagata, A. K. (2006). First report of *Tomato severe rugose virus* in chili pepper in Brazil. *Plant Dis* **90**, 114.
- Blair, M. W., Basset, M. J., Abouzid, A. M., Hiebert, E., Polston, J. E., McMillan, R. T., Graves, W. & Lamberts, M. (1995). Occurrence of bean golden mosaic virus in Florida. *Plant Dis* **79**, 529–533.
- Briddon, R. W., Bedford, I. D., Tsai, J. H. & Markham, P. G. (1996). Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. *Virology* **219**, 387–394.
- Brown, J. K. & Bird, J. (1992). Whitefly-transmitted geminiviruses and associated disorders in the Americas and the Caribbean basin. *Plant Dis* **76**, 220–225.
- Brown, J. K., Fauquet, C. M., Briddon, R. W., Zerbini, F. M., Moriones, E. & Navas-Castillo, J. (2012). Family *Geminiviridae*. In *Virus Taxonomy 9th Report of the International Committee on Taxonomy of Viruses*, pp. 351–373. Edited by A. M. Q. King, M. J. Adams, E. B. Carstens & E. J. Lefkowitz. London, UK: Elsevier Academic Press.
- Bull, S. E., Briddon, R. W., Sserubombwe, W. S., Ngugi, K., Markham, P. G. & Stanley, J. (2006). Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. *J Gen Virol* **87**, 3053–3065.
- Chare, E. R. & Holmes, E. C. (2004). Selection pressures in the capsid genes of plant RNA viruses reflect mode of transmission. *J Gen Virol* **85**, 3149–3157.
- Davino, S., Napoli, C., Dellacroce, C., Miozzi, L., Noris, E., Davino, M. & Accotto, G. P. (2009). Two new natural begomovirus recombinants associated with the tomato yellow leaf curl disease co-exist with parental viruses in tomato epidemics in Italy. *Virus Res* **143**, 15–23.
- Doyle, J. J. & Doyle, J. L. (1987). A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochem Bull* **19**, 11–15.
- Duffy, S. & Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus *Tomato yellow leaf curl virus*. *J Virol* **82**, 957–965.
- Duffy, S. & Holmes, E. C. (2009). Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *J Gen Virol* **90**, 1539–1547.
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113.
- Fernandes, F. R., de Albuquerque, L. C., de Britto Giordano, L., Boiteux, L. S., de Avila, A. C. & Inoue-Nagata, A. K. (2008). Diversity and prevalence of Brazilian bipartite begomovirus species associated to tomatoes. *Virus Genes* **36**, 251–258.
- Fiallo-Olivé, E., Navas-Castillo, J., Moriones, E. & Martínez-Zubiaur, Y. (2012). Begomoviruses infecting weeds in Cuba: increased host range and a novel virus infecting *Sida rhombifolia*. *Arch Virol* **157**, 141–146.
- Fondong, V. N., Pita, J. S., Rey, M. E. C., de Kochko, A., Beachy, R. N. & 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.
- Fu, Y. X. & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.
- García-Andrés, S., Monci, F., Navas-Castillo, J. & Moriones, E. (2006). Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: evidence for the presence of a new virus species of recombinant nature. *Virology* **350**, 433–442.
- García-Andrés, S., Accotto, G. P., Navas-Castillo, J. & Moriones, E. (2007a). 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.
- García-Andrés, S., Tomás, D. M., Sánchez-Campos, S., Navas-Castillo, J. & Moriones, E. (2007b). Frequent occurrence of recombinants in mixed infections of tomato yellow leaf curl disease-associated begomoviruses. *Virology* **365**, 210–219.
- García-Arenal, F., Fraile, A. & Malpica, J. M. (2003). Variation and evolution of plant virus populations. *Int Microbiol* **6**, 225–232.
- Gilbertson, R. L., Faria, J. C., Ahlquist, P. & Maxwell, D. P. (1993). Genetic diversity in geminiviruses causing bean golden mosaic disease: the nucleotide sequence of the infectious cloned DNA components of a Brazilian isolate of bean golden mosaic geminivirus. *Phytopathology* **83**, 709–715.
- González-Aguilera, A., Tavares, S. S., Ramos-Sobrinho, R., Xavier, C. A. D., Dueñas-Hurtado, F., Lara-Rodrigues, R. M., Silva, D. J. H. & Zerbini, F. M. (2012). Genetic structure of a Brazilian population of the begomovirus *Tomato severe rugose virus* (ToSRV). *Trop Plant Pathol* **37**, 346–353.
- Graham, A. P., Martin, D. P. & Roye, M. E. (2010). Molecular characterization and phylogeny of two begomoviruses infecting *Malvastrum americanum* in Jamaica: evidence of the contribution of inter-species recombination to the evolution of malvaceous weed-associated begomoviruses from the northern Caribbean. *Virus Genes* **40**, 256–266.
- Hanley-Bowdoin, L., Settlege, S. B. & Robertson, D. (2004). Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. *Mol Plant Pathol* **5**, 149–156.
- Harrison, B. D. & Robinson, D. J. (1999). Natural genomic and antigenic variation in white-fly transmitted geminiviruses (begomoviruses). *Annu Rev Phytopathol* **37**, 369–398.
- Harrison, B. D., Zhou, X., Otim Nape, G. W., Liu, Y. & Robinson, D. J. (1997). Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. *Ann Appl Biol* **131**, 437–448.

- Ilyina, T. V. & Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria. *Nucleic Acids Res* **20**, 3279–3285.
- Inoue-Nagata, A. K., Albuquerque, L. C., Rocha, W. B. & Nagata, T. (2004). A simple method for cloning the complete begomovirus genome using the bacteriophage phi 29 DNA polymerase. *J Virol Met* **116**, 209–211.
- Jeske, H., Lütgemeier, M. & Preiss, W. (2001). DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO J* **20**, 6158–6167.
- Kosakovsky Pond, S. L. & Frost, S. D. W. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208–1222.
- Lefevre, P., Lett, J. M., Reynaud, B. & Martin, D. P. (2007a). Avoidance of protein fold disruption in natural virus recombinants. *PLoS Pathog* **3**, e181.
- Lefevre, P., Martin, D. P., Hoareau, M., Naze, F., Delatte, H., Thierry, M., Varsani, A., Becker, N., Reynaud, B. & Lett, J. M. (2007b). Begomovirus ‘melting pot’ in the south-west Indian Ocean islands: molecular diversity and evolution through recombination. *J Gen Virol* **88**, 3458–3468.
- Lefevre, P., Lett, J. M., Varsani, A. & Martin, D. P. (2009). Widely conserved recombination patterns among single-stranded DNA viruses. *J Virol* **83**, 2697–2707.
- Legg, J. P. & Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. *Plant Mol Biol* **56**, 585–599.
- Legg, J. P. & Thresh, J. M. (2000). Cassava mosaic virus disease in East Africa: a dynamic disease in a changing environment. *Virus Res* **71**, 135–149.
- Londoño, A., Riego-Ruiz, L. & Argüello-Astorga, G. R. (2010). DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. *Arch Virol* **155**, 1033–1046.
- Lozano, G., Trenado, H. P., Valverde, R. A. & Navas-Castillo, J. (2009). Novel begomovirus species of recombinant nature in sweet potato (*Ipomoea batatas*) and *Ipomoea indica*: taxonomic and phylogenetic implications. *J Gen Virol* **90**, 2550–2562.
- Martin, D. P., van der Walt, E., Posada, D. & Rybicki, E. P. (2005). The evolutionary value of recombination is constrained by genome modularity. *PLoS Genet* **1**, e51.
- Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefevre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462–2463.
- Martin, D. P., Lefevre, P., Varsani, A., Hoareau, M., Semegni, J. Y., Dijoux, B., Vincent, C., Reynaud, B. & Lett, J. M. (2011). Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. *PLoS Pathog* **7**, e1002203.
- Monci, F., Sánchez-Campos, S., Navas-Castillo, J. & Moriones, E. (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.
- Monde, G., Walangululu, J., Winter, S. & Bragard, C. (2010). Dual infection by cassava begomoviruses in two leguminous species (Fabaceae) in Yangambi, Northeastern Democratic Republic of Congo. *Arch Virol* **155**, 1865–1869.
- Morales, F. J. & Anderson, P. K. (2001). The emergence and dissemination of whitefly-transmitted geminiviruses in Latin America. *Arch Virol* **146**, 415–441.
- Morales, F. J. & Jones, P. G. (2004). The ecology and epidemiology of whitefly-transmitted viruses in Latin America. *Virus Res* **100**, 57–65.
- Morilla, G., Krenz, B., Jeske, H., Bejarano, E. R. & Wege, C. (2004). Tête à tête of tomato yellow leaf curl virus and tomato yellow leaf curl sardinia virus in single nuclei. *J Virol* **78**, 10715–10723.
- Moriones, E. & Navas-Castillo, J. (2000). *Tomato yellow leaf curl virus*, an emerging virus complex causing epidemics worldwide. *Virus Res* **71**, 123–134.
- Navas-Castillo, J., Sánchez-Campos, S., Noris, E., Louro, D., Accotto, G. P. & Moriones, E. (2000). Natural recombination between *Tomato yellow leaf curl virus-is* and *Tomato leaf curl virus*. *J Gen Virol* **81**, 2797–2801.
- Ndunguru, J., Legg, J. P., Aveling, T. A., Thompson, G. & Fauquet, C. M. (2005). Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. *Virol J* **2**, 21.
- Orozco, B. M., Kong, L. J., Batts, L. A., Elledge, S. & Hanley-Bowdoin, L. (2000). The multifunctional character of a geminivirus replication protein is reflected by its complex oligomerization properties. *J Biol Chem* **275**, 6114–6122.
- Padidam, M., Sawyer, S. & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218–225.
- Pita, J. S., Fondong, V. N., Sangaré, A., Otim-Nape, G. W., Ogwal, S. & Fauquet, C. M. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J Gen Virol* **82**, 655–665.
- Polston, J. E. & Anderson, P. K. (1997). The emergence of whitefly-transmitted geminiviruses in tomato in the western hemisphere. *Plant Dis* **81**, 1358–1369.
- Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.
- Posada, D. & Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* **54**, 396–402.
- Power, A. G. (2000). Insect transmission of plant viruses: a constraint on virus variability. *Curr Opin Plant Biol* **3**, 336–340.
- Prasanna, H. C. & Rai, M. (2007). Detection and frequency of recombination in tomato-infecting begomoviruses of South and Southeast Asia. *Virol J* **4**, 111.
- Reddy, R. V. C., Colvin, J., Muniyappa, V. & Seal, S. (2005). Diversity and distribution of begomoviruses infecting tomato in India. *Arch Virol* **150**, 845–867.
- Ribeiro, S. G., Ambrozevícius, L. P., Ávila, A. C., Bezerra, I. C., Calegario, R. F., Fernandes, J. J., Lima, M. F., de Mello, R. N., Rocha, H. & Zerbini, F. M. (2003). Distribution and genetic diversity of tomato-infecting begomoviruses in Brazil. *Arch Virol* **148**, 281–295.
- Rocha, C. S. (2011). Variability and genetic structure of begomovirus populations infecting tomato and non-cultivated hosts in south-western Brazil. DS Thesis, *Dep de Fitopatologia*, 129 p. Viçosa, MG: Universidade Federal de Viçosa.
- Rojas, M. R., Hagen, C., Lucas, W. J. & Gilbertson, R. L. (2005). Exploiting chinks in the plant’s armor: evolution and emergence of geminiviruses. *Annu Rev Phytopathol* **43**, 361–394.
- Roosinck, M. J. (2003). Plant RNA virus evolution. *Curr Opin Microbiol* **6**, 406–409.
- Rothenstein, D., Haible, D., Dasgupta, I., Dutt, N., Patil, B. L. & Jeske, H. (2006). Biodiversity and recombination of cassava-infecting begomoviruses from southern India. *Arch Virol* **151**, 55–69.

- Rozas, J., Sánchez-DelBarrio, J. C., Messegueur, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.
- Sanz, A. I., Fraile, A., Gallego, J. M., Malpica, J. M. & García-Arenal, F. (1999). Genetic variability of natural populations of cotton leaf curl geminivirus, a single-stranded DNA virus. *J Mol Evol* **49**, 672–681.
- Sanz, A. I., Fraile, A., García-Arenal, F., Zhou, X., Robinson, D. J., Khalid, S., Butt, T. & Harrison, B. D. (2000). Multiple infection, recombination and genome relationships among begomovirus isolates found in cotton and other plants in Pakistan. *J Gen Virol* **81**, 1839–1849.
- Saunders, K., Bedford, I. D. & Stanley, J. (2001). Pathogenicity of a natural recombinant associated with ageratum yellow vein disease: implications for geminivirus evolution and disease aetiology. *Virology* **282**, 38–47.
- Scheffler, K., Martin, D. P. & Seoighe, C. (2006). Robust inference of positive selection from recombining coding sequences. *Bioinformatics* **22**, 2493–2499.
- Schnippenkoetter, W. H., Martin, D. P., Willment, J. A. & Rybicki, E. P. (2001). Forced recombination between distinct strains of *Maize streak virus*. *J Gen Virol* **82**, 3081–3090.
- Seal, S. E., Jeger, M. J. & Van den Bosch, F. (2006). Begomovirus evolution and disease management. *Adv Virus Res* **67**, 297–316.
- Silva, S. J., Castillo-Urquiza, G. P., Hora Júnior, B. T., Assunção, I. P., Lima, G. S. A., Pio-Ribeiro, G., Mizubuti, E. S. G. & Zerbini, F. M. (2011). High genetic variability and recombination in a begomovirus population infecting the ubiquitous weed *Cleome affinis* in north-eastern Brazil. *Arch Virol* **156**, 2205–2213.
- Silva, S. J. C., Castillo-Urquiza, G. P., Hora-Junior, B. T., Assunção, I. P., Lima, G. S. A., Pio-Ribeiro, G., Mizubuti, E. S. G. & Zerbini, F. M. (2012). Species diversity, phylogeny and genetic variability of begomovirus populations infecting leguminous weeds in northeastern Brazil. *Plant Pathol* **61**, 457–467.
- Souza-Dias, J. A. C., Sawazaki, H. E., Pernambuco-Fo, P. C. A., Elias, L. M. & Maluf, H. (2008). Tomato severe rugose virus: another begomovirus causing leaf deformation and mosaic symptoms on potato in Brazil. *Plant Dis* **92**, 487–488.
- Sserubombwe, W. S., Briddon, R. W., Baguma, Y. K., Ssemakula, G. N., Bull, S. E., Bua, A., Alicai, T., Omongo, C., Otim-Nape, G. W. & Stanley, J. (2008). Diversity of begomoviruses associated with mosaic disease of cultivated cassava (*Manihot esculenta* Crantz) and its wild relative (*Manihot glaziovii* Mull. Arg.) in Uganda. *J Gen Virol* **89**, 1759–1769.
- Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland, Massachusetts: Sinauer Associates.
- Tiendrébéogo, F., Lefeuvre, P., Hoareau, M., Harimalala, M. A., De Bruyn, A., Villemot, J., Traoré, V. S., Konaté, G., Traoré, A. S. & other authors (2012). Evolution of *African cassava mosaic virus* by recombination between bipartite and monopartite begomoviruses. *Viol J* **9**, 67.
- Torres-Pacheco, I., Garzón-Tiznado, J. A., Brown, J. K., Becerra-Flora, A. & Rivera-Bustamante, R. (1996). Detection and distribution of geminiviruses in Mexico and the Southern United States. *Phytopathology* **86**, 1186–1192.
- Were, H. K., Winter, S. & Maiss, E. (2004). Viruses infecting cassava in Kenya. *Plant Dis* **88**, 17–22.
- Wyant, P. S., Gotthardt, D., Schäfer, B., Krenz, B. & Jeske, H. (2011). The genomes of four novel begomoviruses and a new *Sida micrantha* mosaic virus strain from Bolivian weeds. *Arch Virol* **156**, 347–352.
- Zerbini, F. M., Andrade, E. C., Barros, D. R., Ferreira, S. S., Lima, A. T. M., Alfenas, P. F. & Mello, R. N. (2005). Traditional and novel strategies for geminivirus management in Brazil. *Australas Plant Pathol* **34**, 475–480.
- Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G. W., Robinson, D. J. & 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.