Journal of General Virology (2014), **95**, 2540–2552 **DOI** 10.1099/vir.0.067009-0

Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts

Roberto Ramos Sobrinho,[1] César Augusto Diniz Xavier,[1] Hermano Monteiro de Barros Pereira,[1] Gaus Silvestre de Andrade Lima,[2] Iraíldes Pereira Assunção,[2] Eduardo Seiti Gomide Mizubuti,[1] Siobain Duffy[3] and Francisco Murilo Zerbini[1]

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

[1]Departamento de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil

[2]Departamento de Fitossanidade/CECA, Universidade Federal de Alagoas, Rio Largo, AL 57100-000, Brazil

[3]Department of Ecology, Evolution and Natural Resources, Rutgers, The State University Of New Jersey, New Brunswick, NJ 08901, USA

Begomoviruses are whitefly-transmitted, ssDNA plant viruses and are among the most damaging pathogens causing epidemics in economically important crops worldwide. Wild/non-cultivated plants play a crucial epidemiological role, acting as begomovirus reservoirs and as 'mixing vessels' where recombination can occur. Previous work suggests a higher degree of genetic variability in begomovirus populations from non-cultivated hosts compared with cultivated hosts. To assess this supposed host effect on the genetic variability of begomovirus populations, cultivated (common bean, *Phaseolus vulgaris*, and lima bean, *Phaseolus lunatus*) and non-cultivated (*Macroptilium lathyroides*) legume hosts were sampled from two regions of Brazil. A total of 212 full-length DNA-A genome segments were sequenced from samples collected between 2005 and 2012, and populations of the begomoviruses *Bean golden mosaic virus* (BGMV) and *Macroptilium yellow spot virus* (MaYSV) were obtained. We found, for each begomovirus species, similar genetic variation between populations infecting cultivated and non-cultivated hosts, indicating that the presumed genetic variability of the host did not a priori affect viral variability. We observed a higher degree of genetic variation in isolates from MaYSV populations than BGMV populations, which was explained by numerous recombination events in MaYSV. MaYSV and BGMV showed distinct distributions of genetic variation, with the BGMV population (but not MaYSV) being structured by both host and geography.

Received 15 April 2014
Accepted 11 July 2014

INTRODUCTION

Viruses belonging to the family *Geminiviridae* have circular ssDNA genomes and are widely distributed in tropical and subtropical regions, infecting several economically important crop species (Legg & Fauquet, 2004; Morales, 2006). The family is divided into seven genera (*Becurtovirus*, *Begomovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus* and *Turncurtovirus*) according to the type of insect vector, host range, genome organization and phylogeny (Varsani *et al.*, 2014). Viruses classified within the genus *Begomovirus* are transmitted by the whitefly *Bemisia tabaci* (Brown *et al.*, 2012). Begomoviruses usually found in the New World have two genomic components known as DNA-A and DNA-B, but only the DNA-A is used in the taxonomic classification of bipartite begomoviruses (Brown *et al.*, 2012). The DNA-A contains genes involved in replication and encapsidation, while the DNA-B contains genes responsible for intra- and intercellular movement (Hanley-Bowdoin *et al.*, 2013). Begomoviruses are among the most damaging pathogens infecting cultivated plants worldwide (Legg & Fauquet, 2004; Morales & Anderson, 2001; Navas-Castillo *et al.*, 2011; Varma & Malathi, 2003), and specifically limit production of tomatoes, peppers and legumes in the New World (Morales, 2006; Navas-Castillo *et al.*, 2011). In Brazil, *Bean golden mosaic virus* (BGMV) has been an important pathogen infecting beans

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are KJ939707–KJ939918.

Four supplementary tables are available with the online version of this paper.

(*Phaseolus* spp.) since the 1970s (Costa, 1976; Faria & Maxwell, 1999; Gilbertson *et al.*, 1991b), causing yield losses between 40 and 100 % (Morales, 2006). Despite its economic and social importance, few studies have focused on BGMV and other begomovirus populations infecting these crops (Faria & Maxwell, 1999; Lima *et al.*, 2013). Other begomovirus species, such as *Macroptilium yellow spot virus* (MaYSV) and *Sida micrantha mosaic virus* (SimMV), have been reported naturally infecting common bean and non-cultivated legume hosts, but there is no information about their epidemiological importance for bean crops.

Wild/non-cultivated plants from different botanical families can sustain a high species diversity of begomoviruses (Castillo-Urquiza *et al.*, 2008; Fiallo-Olive *et al.*, 2012; García-Andrés *et al.*, 2006; Silva *et al.*, 2012; Tavares *et al.*, 2012; Wyant *et al.*, 2011), and can play an important epidemiological role serving as alternate/reservoir hosts, preventing local extinctions of the virus when the cultivated host is absent (Alabi *et al.*, 2008; Barbosa *et al.*, 2009; Rocha *et al.*, 2013). In these cases, whiteflies transmitting begomoviruses between cultivated and non-cultivated hosts contribute to virus evolution and disease epidemics (Alabi *et al.*, 2008; Power, 2000). Additionally, mixed infections by different begomoviruses are common in wild hosts (Alabi *et al.*, 2008; García-Andrés *et al.*, 2006; Monde *et al.*, 2010), facilitating recombination among distantly related begomoviruses. Recombination is an important evolutionary mechanism in begomoviruses (Lefeuvre *et al.*, 2007b, 2009; Lima *et al.*, 2013; Martin *et al.*, 2005, 2011), and resulted in higher genetic variability found in begomoviruses populations infecting primarily non-cultivated hosts (Lima *et al.*, 2013).

We have carried out a large-scale study to obtain more information about the genetic structure and factors shaping genetic variability in begomovirus populations infecting legume hosts in Brazil. Foliar samples of common bean (*Phaseolus vulgaris*), lima bean (*Phaseolus lunatus*) and of the weed *Macroptilium lathyroides* (located near common bean or lima bean fields) were collected over an 8-year period from different regions where the begomoviruses MaYSV and BGMV were previously reported infecting cultivated and/or non-cultivated hosts (Gilbertson *et al.*, 1991a; Lima *et al.*, 2013; Silva *et al.*, 2012; Wyant *et al.*, 2012). We hypothesized that non-cultivated hosts, which are more genetically variable, would harbour more genetically variable virus sequences, whereas cultivated hosts would select for fewer haplotypes. Instead, we found that the presumed genetic variability of the host did not affect viral sequence variability. Our results corroborate previous studies indicating high genetic variability in MaYSV populations due to interspecific recombination (Lima *et al.*, 2013; Silva *et al.*, 2012) and low genetic variability in BGMV populations (Faria & Maxwell, 1999). Nevertheless, MaYSV and BGMV showed different biogeographical patterns, and the genetic structure of BGMV populations was strongly shaped by geography and host, while that of MaYSV populations was not.

RESULTS

The prevalence of begomoviruses in leguminous hosts in Brazil shifts temporally and spatially

A total of 515 plant samples (300 common bean, 115 lima bean and 100 *M. lathyroides*) were collected. From these samples, 212 full-length DNA-A components were cloned and sequenced. Each component was sequenced from individual plants (see Table S1, available in the online Supplementary Material). Only 11 full-length DNA-B components of common bean-infecting BGMV were obtained and thus no additional analyses were performed based on this component.

Using pairwise comparisons of the DNA-A sequences and the recently established $\geqslant$91 % nucleotide identity criterion established by the *Geminiviridae* Study Group of the International Committee on Taxonomy of Viruses (ICTV) (http://talk.ictvonline.org/files/proposals/taxonomy_proposals_plant1/m/plant04/4399.aspx), the 212 isolates were assigned to four begomovirus species: MaYSV, BGMV, *Soybean chlorotic spot virus* (SoCSV) and *Macroptilium yellow vein virus* (MaYVV). While we have sequenced isolates of these viruses, we assume they are representative of the populations from which they were isolated, and so refer to them collectively as populations.

MaYSV was cloned from the three host species, but was found only in the north-eastern region [states of Alagoas (AL) and Sergipe (SE)]. MaYSV is known to naturally infect non-cultivated plants (*M. lathyroides*, *Calopogonium mucunoides* and *Canavalia* sp.) as well as the common bean (Lima *et al.*, 2013; Silva *et al.*, 2012). Our results extend its natural host range to include the cultivated *P. lunatus*. BGMV was cloned from lima bean in AL, common bean in the central region [states of Minas Gerais (MG), Goiás (GO) and the Federal District (DF)], and *M. lathyroides* in MG. SoCSV was found in *M. lathyroides* samples from MG and MaYVV was only isolated from *M. lathyroides* plants adjacent to a lima bean field in AL (Table S1).

The diversity of begomoviruses infecting *P. lunatus* in the north-east was different between sampling times. In 2005, only BGMV isolates were obtained from this host in AL and Pernambuco (PE), but in 2011, nearly a third of the clones obtained from *P. lunatus* were classified as MaYSV. In two surveys in 2011 (this study and Lima *et al.*, 2013), MaYSV was the begomovirus most frequently cloned from legume hosts in AL (87 of 146 clones obtained), while BGMV and MaYVV were less prevalent (41 and 18 clones, respectively). These results highlight the rapid emergence of MaYSV in legume crops in AL, where it is now the prevalent begomovirus.

MaYSV populations are more variable than BGMV populations

A total of 99 MaYSV and 152 BGMV full-length DNA-A sequences were analysed. The MaYSV dataset contained 55

sequences previously published (Lima *et al.*, 2013; Silva *et al.*, 2012) plus 44 new sequences described here, and the BGMV dataset contained five sequences previously published (Fernandes *et al.*, 2009; Gilbertson *et al.*, 1991a; Silva *et al.*, 2012) plus 147 new sequences described here (Tables S1 and S2).

The genetic variability of MaYSV was evenly distributed among isolates infecting the different hosts, and was similar to the variability calculated for all isolates (Table 1). Among the different BGMV subpopulations from each of the three hosts, isolates infecting lima bean were slightly more variable than those from other hosts, despite a smaller population size than isolates from common bean (Table 1). The genetic variability calculated for all BGMV isolates was higher than that calculated for isolates infecting each host, suggesting that isolates infecting each host comprise distinct subpopulations and that the variability of the whole population reflects the variability between (rather than within) each subpopulation (Table 1). The MaYSV Rep gene was four times more variable than its CP gene (Table 1), and the MaYSV population showed greater genetic variability than the BGMV populations, especially in the Rep gene (Table 1).

Recombination is very common in MaYSV but not in BGMV

To investigate intraspecific recombination events, full-length DNA-A genomes of MaYSV and BGMV were analysed using the Recombination Detection Program (RDP)3 package (Table S3). For MaYSV, a complex pattern of recombination was found, with seven unique events. Six events had putative breakpoints in the Rep gene and the common region, and only one event was in the CP gene (Table S3). In contrast, only two putative recombination events were detected in BGMV (Table S3). These events were restricted to isolates infecting lima bean from AL, where almost all isolates differed from other BGMV isolates by a recombinant region with breakpoints in the Rep and CP genes, and where one isolate (BR : Ata2 : 05) has a recombinant region in its CP gene and the common region (Table S3).

The effect of recombination in the CP and Rep genes was visualized using neighborNet analysis. The networks included MaYSV, BGMV and other begomoviruses that showed greatest nucleotide identity to putative recombinant regions detected in MaYSV and BGMV by RDP3 analysis (Table S2). In the MaYSV CP dataset, isolates BR : Bas1 : 09 and BR : Mac1 : 10 may also be recombinant, with another begomovirus donating sequence to their common ancestor (Fig. 1a). The network of MaYSV Rep sequences confirmed the complex pattern of recombination detected by RDP3, with both intra- and inter-species events (Fig. 1b). Isolates in Cluster I were closely related to *Sida yellow blotch virus* (SiYBV) and *Sida mosaic Alagoas virus* (SiMAlV, obtained from *Sida* spp.), isolates in Cluster II to SoCSV, and those in Cluster III to *Blainvillea yellow spot virus* (BlYSV, from *Blainvillea rhomboidea*) (Fig. 1b). Recombination was also suggested in the BGMV dataset, albeit at a lower degree (Fig. 2). The *M. lathyroides*-infecting BGMV isolates BR : Car3 : 10 and BR : Car4 : 10 from AL were more closely placed to other begomovirus species in the CP network compared to the Rep network (Fig. 2). Similarly, the lima bean-infecting isolates BR : Ata2 : 05, BR : Rec1 : 05 and BR : Rec2 : 05 were located away from the main BGMV clusters in both the CP and Rep networks (Fig. 2).

Full-length DNA-A datasets including the begomovirus species used in the neighborNet analysis were then assessed for interspecific recombination using RDP3. For MaYSV, the event involving isolates in Cluster I had as putative minor parents SiYBV and SiMAlV, and the single event involving the CP gene had *Macroptilium yellow net virus* (MaYNV) as the putative minor parent (data not shown). For BGMV, both events were confirmed as intra-species recombination, with different BGMV isolates identified as the putative parents (data not shown).

BGMV but not MaYSV populations show genetic structuring by geography and host

The MaYSV CP and Rep maximum likelihood (ML) phylogenetic trees are not congruent (Fig. 3). The Rep tree has four well-supported clades that are not resolved in the CP tree. In both trees, there is no evidence for structuring

Table 1. Genetic variability of the begomoviruses BGMV and MaYSV infecting three distinct leguminous hosts

Results are given as pairwise, per-site nucleotide diversity ± SD.

Population	No. of sequences	DNA-A π	CP π	Rep π
BGMV (total)	147	0.0489 ± 0.0015	0.0442 ± 0.0012	0.0371 ± 0.0013
BGMV (common bean)	75	0.0067 ± 0.0002	0.0058 ± 0.0002	0.0043 ± 0.0002
BGMV (lima bean)	59	0.0153 ± 0.0039	0.0117 ± 0.0031	0.0144 ± 0.0036
BGMV (*Macroptilium*)	13	0.0018 ± 0.0005	0.0020 ± 0.0006	0.0009 ± 0.0002
MaYSV (total)	99	0.0627 ± 0.0021	0.0262 ± 0.0020	0.1055 ± 0.0040
MaYSV (common bean)	50	0.0622 ± 0.0032	0.0224 ± 0.0011	0.1068 ± 0.0065
MaYSV (lima bean)	21	0.0590 ± 0.0062	0.0212 ± 0.0019	0.0976 ± 0.0119
MaYSV (*Macroptilium*)	28	0.0662 ± 0.0037	0.0352 ± 0.0056	0.1037 ± 0.0078

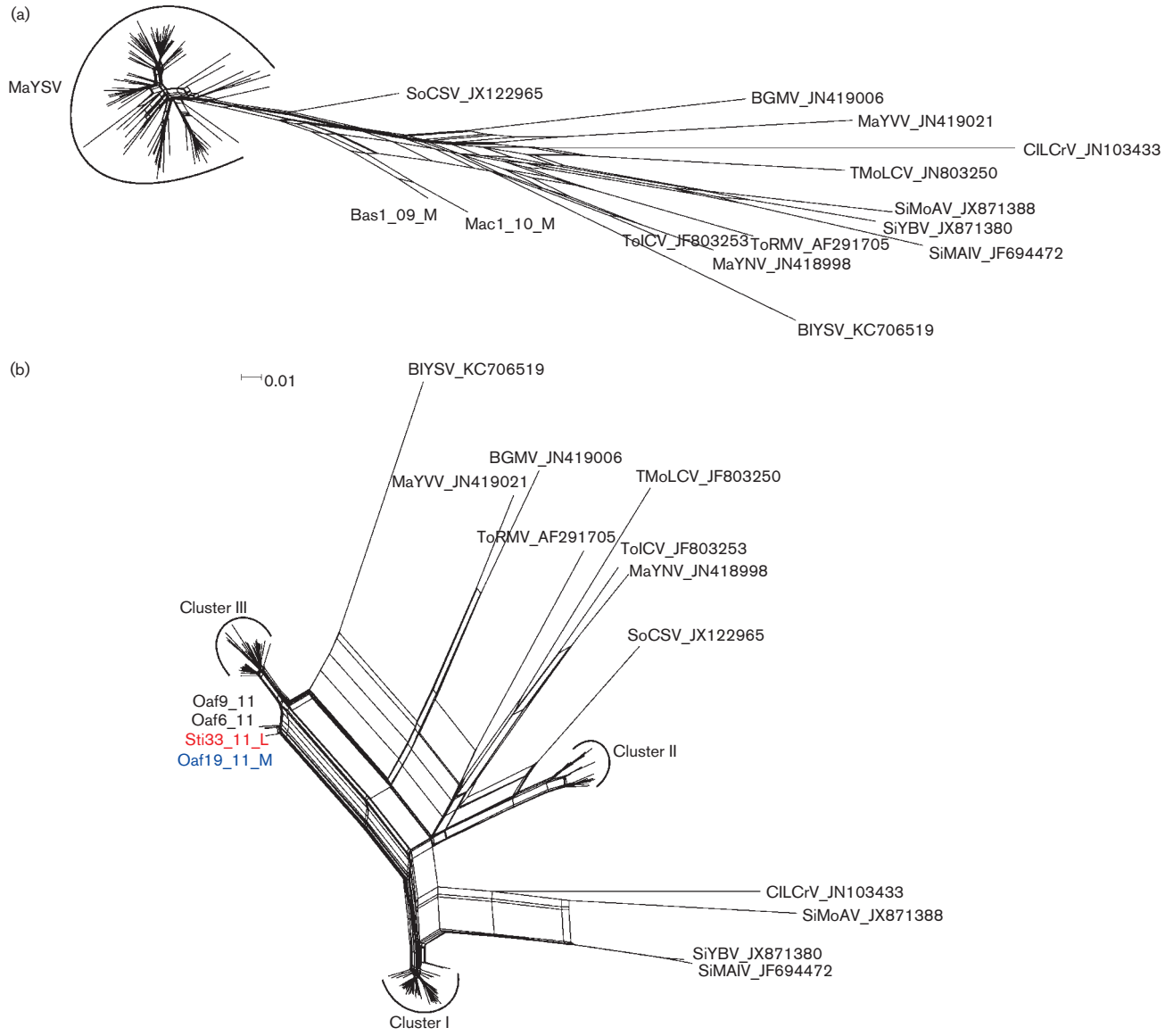

Fig. 1. NeighborNet networks based on (a) CP and (b) Rep nucleotide sequences of MaYSV and selected begomoviruses from Brazil. In the Rep network, isolates from *P. vulgaris* are indicated in black, from *P. lunatus* in red, and from *M. lathyroides* in blue. The clusters labelled MaYSV in the CP network and I, II and III in the Rep network contain isolates from all three hosts.

by location or host species, as each clade includes isolates from different hosts and/or locations (Fig. 3). These results were confirmed by subdivision analysis based on full-length DNA-A sequences ($F_{ST/host}$=0.0223; $F_{ST/location}$=0.0794), which indicates that putative subpopulations have little genetic differentiation.

The BGMV CP and Rep ML phylogenetic trees are similar (Fig. 4), with strong evidence for population structuring both by location and host plant. Both trees showed three main clades, with isolates from AL (primarily from lima bean) being a separate population from MG, GO and DF (common bean) and MG (non-cultivated plants). Isolates from Florestal (MG) formed a sister subpopulation to isolates from Unaí (MG), GO and DF (Fig. 4). As Unaí is closer to fields located in the neighbouring state of GO [i.e. 85 km from Cristalina (GO) but 470 km from Florestal (MG)], geographical proximity can explain the clustering observed in the phylogenetic trees. Two isolates (BR : Rec1 : 05 and BR : Rec2 : 05) from Recife (PE) clustered with MG isolates, suggesting that there has been migration between these states (Fig. 4). Based on F_{ST} analysis for full-length DNA-A sequences, BGMV showed strong genetic differentiation between subpopulations isolated from different hosts/locations ($F_{ST/host/location}$=0.8667).

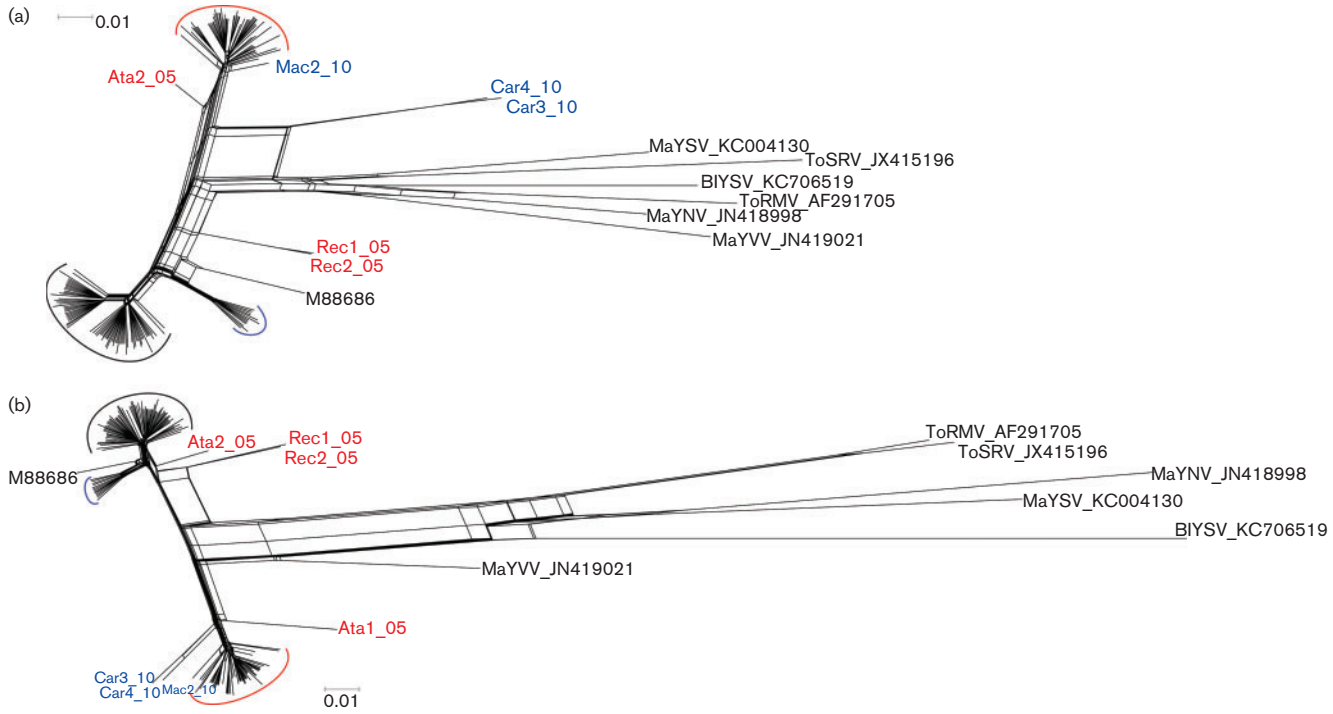

Fig. 2. NeighborNet networks based on (a) CP and (b) Rep nucleotide sequences of BGMV and selected begomoviruses from Brazil. Isolates from *P. vulgaris* are indicated in black, from *P. lunatus* in red, and from *M. lathyroides* in blue.

Purifying selection predominates in BGMV and MaYSV populations

To investigate the extent in which selection pressures have shaped the standing genetic variability in MaYSV and BGMV populations, CP and Rep datasets were analysed using different ML-based methods. We analysed subpopulations separately to minimize the impact of recombination on our results. Most datasets showed mean non-synonymous to synonymous substitution ratios (d_N/d_S) <1 (Table 2), indicating negative selection. Although neutral evolution may be occurring in the Rep encoded by BGMV infecting *Macroptilium* ($d_N/d_S=1.0$), this analysis contained only four distinct sequences, and a larger dataset needs to be analysed before definitive conclusions can be drawn.

A higher number of sites were under statistically detectable negative selection than positive selection in both MaYSV and BGMV populations. For example, using the fixed-effect likelihood (FEL) method on the MaYSV CP dataset, only three sites (142, 163 and 250) were identified as under positive selection, with site 250 also detected by Single-Likelihood Ancestor Counting (SLAC), while 72 sites were under negative selection (Table S4). A few sites under positive selection were detected in MaYSV Rep datasets. In Clusters I, II and III, one, six and two sites were detected by the random-effect likelihood (REL), REL and FEL, respectively (Table S4). In the BGMV CP gene, seven sites under positive selection were identified in the population infecting *P. vulgaris* (Table S4). A larger number of sites under positive selection were detected in the small (only four haplotypes) *Macroptilium*-infecting BGMV dataset: 28 sites were under positive selection, with only two sites under negative selection (Table S4). Partitioning for Robust Inference of Selection (PARRIS) did not identify any sites under positive selection in any of the datasets analysed, emphasizing that negative selection is the most important selective force acting upon BGMV and MaYSV populations.

DISCUSSION

The species diversity of begomoviruses has been surveyed extensively (Ala-Poikela *et al.*, 2005; Albuquerque *et al.*, 2012; Bull *et al.*, 2006; Castillo-Urquiza *et al.*, 2008; Fernandes *et al.*, 2009; García-Andrés *et al.*, 2006; Lozano *et al.*, 2009; Ndunguru *et al.*, 2005; Reddy *et al.*, 2005; Ribeiro *et al.*, 2003; Rothenstein *et al.*, 2006; Sserubombwe *et al.*, 2008; Tavares *et al.*, 2012), and some studies have investigated the genetic structure of begomovirus populations in cultivated and non-cultivated hosts in different geographical regions (Ge *et al.*, 2007; González-Aguilera *et al.*, 2012; Lima *et al.*, 2013; Rocha *et al.*, 2013; Silva *et al.*, 2012, 2011). Strikingly, in spite of its great economic importance as the most damaging viral pathogen of bean

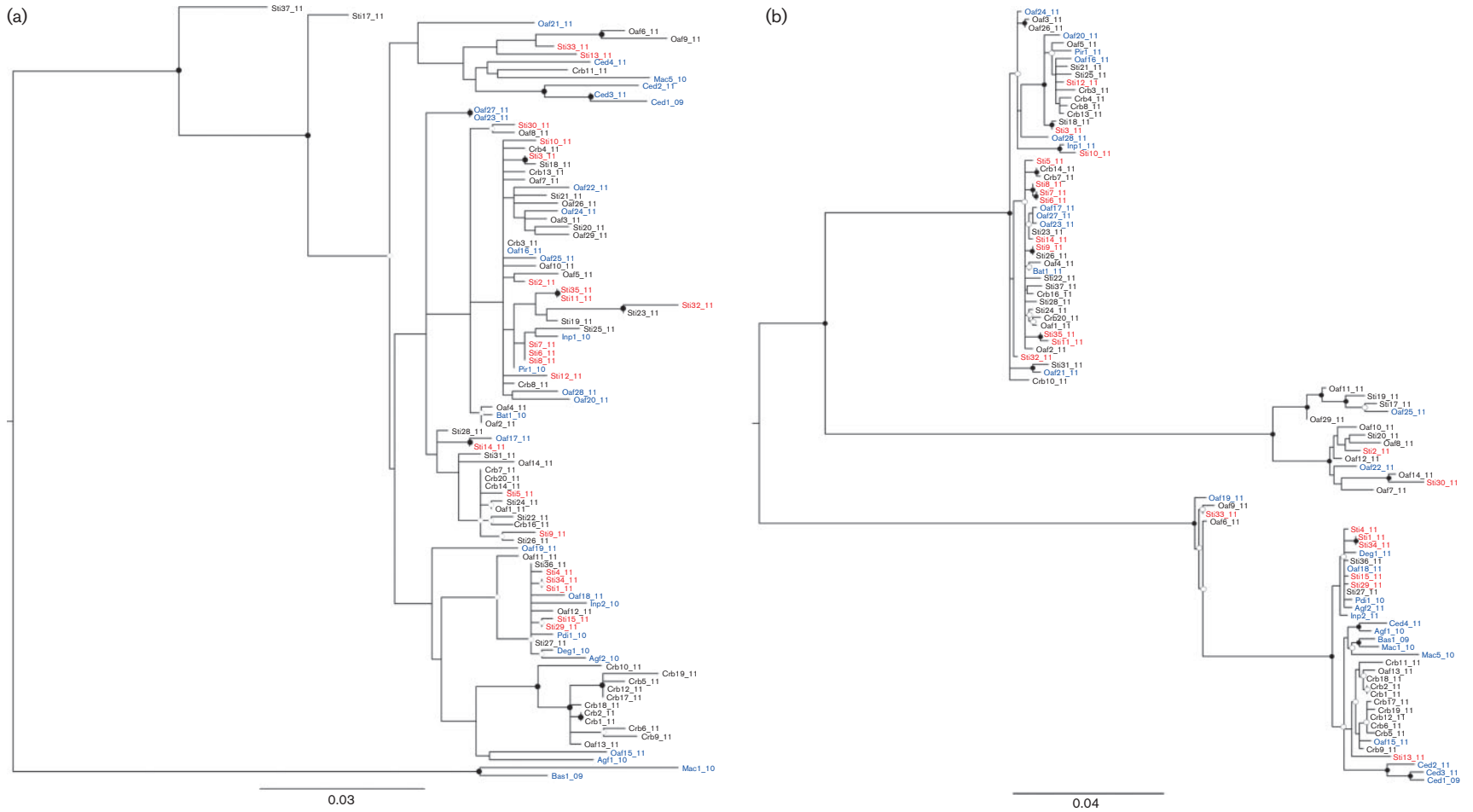

Fig. 3. Midpoint-rooted ML trees based on CP (a) and Rep (b) nucleotide sequences of MaYSV isolates from *P. vulgaris* (indicated in black), *P. lunatus* (red) and *M. lathyroides* (blue). Nodes with bootstrap values >50 % and <80 % are indicated by open circles, and those with values ⩾80 % by filled circles. Isolates with identical names (first three letters) were collected at the same location. Bar, number of nucleotide substitutions per site.

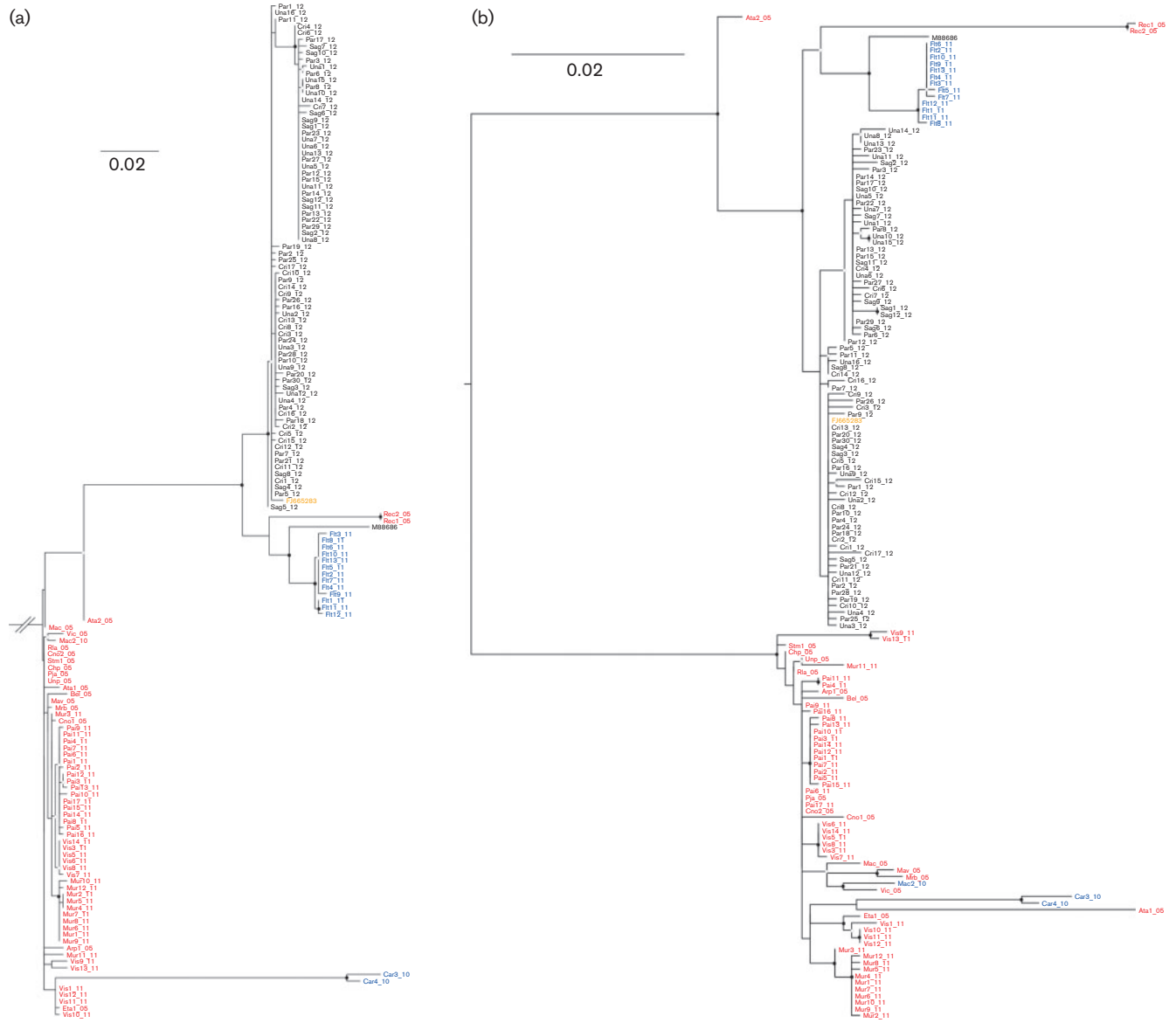

Fig. 4. Midpoint-rooted ML trees based on CP (a) and Rep (b) nucleotide sequences of BGMV isolates from *P. vulgaris* (indicated in black), *P. lunatus* (red) and *M. lathyroides* (blue). An isolate infecting *Glycine max* is indicated in orange. Nodes with bootstrap values >50 % and <80 % are indicated by open circles, and those with values ⩾80 % by filled circles. Isolates with identical names (first three letters) were collected at the same location. Bar, number of nucleotide substitutions per site.

crops in South America (Morales, 2006), no studies on the genetic variability of BGMV have been conducted since Faria & Maxwell (1999) analysed 20 partial DNA-A sequences. Our study of 272 full-length DNA-A begomovirus sequences across Brazil found no effect of genetic background of the host species on molecular variability of begomovirus populations, confirmed the high genetic variability in MaYSV populations, and detected large differences between BGMV populations infecting each of three leguminous hosts.

BGMV was the predominant begomovirus species found infecting lima bean in north-eastern Brazil in 2005, but was a minority of isolates sampled in this region in 2011. Instead, MaYSV was the main begomovirus infecting legumes in 2011 in AL, but was not present in the central region, where BGMV was still prevalent. MaYSV was reported for the first time infecting non-cultivated legumes in the north-east in 2010 (Silva *et al.*, 2012), common bean in AL in 2011 (Lima *et al.*, 2013) and now naturally infecting another cultivated host, *P. lunatus* (this study). The absence of MaYSV in samples from 2005 could be related to the cloning method used at that time, which was biased for previously known sequences. However, a study using cloning-independent (i.e. unbiased) pyro-sequencing

Table 2. Mean values of non-synonymous to synonymous substitution ratios (d_N/d_S) for the subpopulation-specific CP and Rep genes of BGMV and MaYSV

Datasets	d_N/d_S
BGMV CP (common bean)	0.0936
BGMV CP (lima bean)	0.1207
BGMV CP (*Macroptilium*)	0.4844
BGMV Rep (common bean)	0.3541
BGMV Rep (lima bean)	0.2073
BGMV Rep (*Macroptilium*)	1.0000
MaYSV CP	0.0919
MaYSV Rep CI	0.1757
MaYSV Rep CII	0.1725
MaYSV Rep CIII	0.2427

of cultivated and non-cultivated legume samples collected in 2003 and 2004 in three north-eastern states (AL, PE and Bahia) also did not detect the presence of MaYSV (Wyant *et al.*, 2012). Our combined results provide strong evidence for the recent emergence of MaYSV as an important agricultural pathogen, for a rapid expansion of its host range and for its swift spread throughout AL. Indeed, MaYSV seems to be replacing BGMV as the dominant begomovirus infecting common bean and lima bean crops in AL. If the trends in AL hold for other regions in Brazil, monitoring the spread of MaYSV would be critical to guarantee the success of BGMV-resistant transgenic common bean.

High genetic variability in begomovirus populations infecting different non-cultivated hosts has been reported (Lima *et al.*, 2013; Rocha *et al.*, 2013; Silva *et al.*, 2012, 2011). Conversely, begomovirus populations infecting cultivated hosts seem to have lower variability (Faria & Maxwell, 1999; González-Aguilera *et al.*, 2012; Lima *et al.*, 2013; Rocha *et al.*, 2013). Here, we found that viral sequence variability was similar in one non-cultivated and two cultivated hosts for both BGMV and MaYSV. It will be interesting to determine whether this result for legume-infecting begomoviruses can be reproduced for begomoviruses infecting solanaceous or malvaceous hosts.

Our results on BGMV, based on 152 full-length DNA-A sequences, are in agreement with Faria & Maxwell (1999). We found little genetic variability in common bean-infecting BGMV isolates, while the lima bean-infecting population, which showed evidence of recombination, was slightly more variable. The variability in MaYSV populations observed here was similar to previous reports (Lima *et al.*, 2013; Silva *et al.*, 2012), and when compared to other begomovirus species, MaYSV was 2.5 times more variable than the next most variable begomovirus reported in Brazil, BlYSV (González-Aguilera *et al.*, 2012; Rocha *et al.*, 2013; Silva *et al.*, 2011). We found additional begomovirus species infecting *M. lathyroides*, consistent with previously published reports where this host sustains a high species diversity of begomoviruses (Silva *et al.*, 2012).

Recombination is very common and important for the emergence of different geminiviruses, and occurs within species, between species and across genera in the family (García-Andrés *et al.*, 2007a, b; Hou & Gilbertson, 1996; Lefeuvre *et al.*, 2009, 2007b; Monci *et al.*, 2002; Padidam *et al.*, 1999; Pita *et al.*, 2001; Schnippenkoetter *et al.*, 2001; Tiendrébéogo *et al.*, 2012; Umaharan *et al.*, 1998). Different studies have shown that it is also an important evolutionary mechanism acting within begomovirus populations in Brazil (Lima *et al.*, 2013; Rocha *et al.*, 2013; Silva *et al.*, 2012, 2011), with detectable events of recombination occurring in regular, non-random regions in the genome (Lima *et al.*, 2013; Rocha *et al.*, 2013).

A complex recombination pattern was found in MaYSV. The more variable populations all showed evidence of recombination (e.g. the frequently recombinant Rep gene in MaYSV), suggesting that recombination could, in part, explain the higher genetic variability found in MaYSV, consistent with previously published reports (Lima *et al.*, 2013). Although we are unable to provide definitive evidence that the parental sequences identified here are the actual parents, these results provide a strong indication that MaYSV evolved through recombination between begomoviruses infecting non-cultivated hosts belonging to different botanical families. Similar results were observed for tomato-infecting begomoviruses in Brazil, which are thought to have arisen from inter-species recombination between begomoviruses infecting non-cultivated hosts (Ribeiro *et al.*, 2007; Rocha *et al.*, 2013). Although some studies indicate that a large number of recombinants arising from events between distantly related genomes are defective and probably would be removed from the population by selection (Liu *et al.*, 1999; Martin *et al.*, 2005), our results continue to emphasize the importance of recombination for microevolution and macroevolution of agronomically important begomoviruses.

MaYSV isolates with detectable recombination in the Rep gene were found in all three hosts studied, but isolates with recombination in the CP region were only found twice, both in *M. lathyroide*s in 2009 and 2010. The lack of spread of CP recombinants could be evidence that they suffer a greater fitness cost than Rep recombinants. Several studies have shown that selection pressure is an important factor determining which recombinants survive in viral populations (Lefeuvre *et al.*, 2007a; Martin *et al.*, 2011; Rokyta & Wichman, 2009; Simon-Loriere *et al.*, 2009), and purifying selection is strongest in the CP gene of begomoviruses (Lefeuvre *et al.*, 2009). Interestingly, recombination seems infrequent or selected against in BGMV populations, and the few putative events detected here were restricted to isolates located in AL.

The genetic structure of begomovirus populations is determined by mutation, recombination and the interplay between adaptation to host species and vector biotypes, and is influenced by the geographical distribution of the hosts, vectors and other begomoviruses (Lima *et al.*, 2013;

Navas-Castillo *et al.*, 2011; Rocha *et al.*, 2013). Recently, phylogeographic segregation of begomoviruses infecting the same crop was observed in Brazil, where different species were prevalent in different tomato-growing areas (Rocha *et al.*, 2013). Here, similar results were observed for BGMV infecting cultivated and non-cultivated legume hosts.

MaYSV populations were not neatly subdivided according to host or sampled area, providing strong evidence of migration among fields. These phylogenetic results were confirmed by extremely low F_{ST} values. In contrast, evidence of structuring by both host and geography was found for BGMV populations. However, related isolates were found infecting both lima bean and *M. lathyroides* in AL (Silva *et al.*, 2012), and infecting common bean and *M. lathyroides* in MG, which indicate that this segregation is not absolute. Furthermore, while there appears to have been migration between fields located in PE and MG, BGMV isolates in north-eastern and central Brazil cannot be considered to comprise one unified population.

The contrasting population structure between MaYSV and BGMV could be due to the distances among sampling sites. The MaYSV isolates are separated by a maximum of 240 km [Santana do Ipanema (AL) and Barra de Santana (PB)], but all but one isolate (BR : Bas1 : 10) were sampled no more than 105 km apart. BGMV was isolated within MG from fields 472 km apart (Unaí and Florestal), and the maximum distance between sampling sites was 1425 km [Cristalina (GO) and Arapiraca (AL)]. Over this larger area, we observed clustering between isolates collected at closer locations, such as isolates from Unaí (MG) forming one population with isolates from GO and DF, whose sampling sites were no more than 85 km away. The observed lack of population structure for MaYSV may be due to its restricted geographical range compared to BGMV, and that substructure could appear as MaYSV spreads over larger distances. Furthermore, it is an unknown question whether we are observing viral populations at 'equilibrium variability', but it seems doubtful that the structures observed in this study will persist over the next decades. For instance, we would hesitate to consider BGMV stable in light of the fact that it was much more frequent 20 years ago, when it was considered the only major begomovirus threat to Brazilian agriculture.

Recently, a genetically modified BGMV-resistant common bean was approved by the Brazilian National Biosafety Commission (http://www.ctnbio.gov.br/index.php/content/view/16526.html) in which resistance is due to RNA silencing (Bonfim *et al.*, 2007), a process which is highly sequence-specific (Raja *et al.*, 2010). Incidentally, our results could impact the implementation of this BGMV-resistant transgenic common bean in Brazil. As this resistance is based on the highly sequence-specific mechanism of RNA silencing, the effectiveness of transgenic cultivars must be assessed against the genetic variation in BGMV and other common bean-infecting begomoviruses present in Brazil. In particular, in AL where the highly variable MaYSV is widespread and efficiently infects common bean, the use of this cultivar may fail to control the losses due to begomoviruses. Fortunately, MaYSV seems to be currently restricted to a small part of north-eastern Brazil, and efforts could be made to contain its spread into other regions. The far less variable BGMV found in the main areas growing common bean in central Brazil provides hope that the RNAi (RNA interference) resistance may be durable under field conditions.

Purifying selection was the dominant force acting on BGMV and MaYSV CP and Rep genes, in agreement with most begomovirus studies (García-Andrés *et al.*, 2007a; González-Aguilera *et al.*, 2012; Lima *et al.*, 2013; Rocha *et al.*, 2013; Sanz *et al.*, 1999; Silva *et al.*, 2012, 2011). Positive selection was infrequently observed, and in both species some positively selected sites in the CP can be associated with insect transmission (i.e. sites 123 for BGMV and 142 for MaYSV). Sites located in the region between amino acids 123 and 152 of the CP gene have been shown to be involved in insect transmissibility (Caciagli *et al.*, 2009; Noris *et al.*, 1998; Hohnle *et al.*, 2001; Kheyr-Pour *et al.*, 2000). Despite the long length of this transmission-determining region (compared to the concise motif conferring aphid transmission in potyviruses; James & Bryce, 2006), changes in only one (Kheyr-Pour *et al.*, 2000) or two (Hohnle *et al.*, 2001; Noris *et al.*, 1998) amino acids in this region may prevent whitefly transmission. Therefore, the positively selected sites should be improving whitefly transmission in these begomovirus populations. Additional studies are needed to determine if the positively selected sites identified here are involved in whitefly transmission.

Although the N-terminal region in MaYSV Rep is known to be highly recombinant (and highly variable), the few sites we identified to be under positive selection were mostly located in the C-terminal region. Previously identified sites under positive selection in MaYSV Rep were hypothesized to be spurious results due to recombination events (Lima *et al.*, 2013), which is corroborated by our subpopulation-specific results. Interestingly, neutral evolution was observed in BGMV Rep ($d_N/d_S=1.0$) in the subpopulation infecting *M. lathyroides*. These results provide additional evidence that high nucleotide variability in the N-terminal portion of the Rep gene (e.g. the MaYSV population) is accompanied by strong purifying selection that preserves the amino acid sequence. The Rep N-terminal region in geminiviruses includes conserved motifs essential for rolling-circle replication (Ilyina & Koonin, 1992; Koonin & Ilyina, 1992; Nash *et al.*, 2011). Conservation of the integrity of these elements is critical for successful infection cycles, despite the variation introduced by frequent recombination.

We have determined the genetic structure of two legume-infecting begomovirus populations. BGMV populations are less variable than MaYSV ones, mostly due to recombination acting upon MaYSV. BGMV populations are strongly structured by geography/host, while MaYSV populations are not. These results suggest that, at least for begomoviruses,

genetic variability is an intrinsic viral property, rather than a malleable feature that could be affected by the host (also seen in Ge *et al.*, 2007). The emergence of the highly variable MaYSV in *Phaseolus* spp. could seriously complicate disease management in this important crop in Brazil and other countries in South America.

METHODS

Sample collection and storage. Foliar samples with virus-like symptoms such as yellow mosaic, leaf curl and stunting were collected in common bean and lima bean fields in different states of Brazil. Lima bean (*P. lunatus*) landrace samples were collected in fields in PE in 2005 and AL in 2005 and 2011 (detailed isolation information given in Table S1). Common bean (*P. vulgaris*, 'carioca' type) samples were collected in fields in AL in 2011 and in GO and MG states and in the DF in 2012 (Table S1). Most common bean samples belonged to cultivar Pérola, except samples from AL, for which the cultivar was not determined. Samples of non-cultivated *M. lathyroides* near common bean or lima bean fields were collected in AL and SE states in 2011 and MG in 2012 (Table S1). For each sample collected in 2011 or 2012, the following information was recorded: plant species (and cultivar for the common bean samples), date of collection, GPS coordinates of the sampling location and a digital image of the sample at the time of collection. Incomplete information was available for samples collected in 2005. Samples were analysed while fresh or were press dried at room temperature for storage as dried leaf samples until analysed.

DNA amplification and cloning. Total DNA was extracted from fresh tissue or dried leaf samples as described by Doyle & Doyle (1987) and used as a template for rolling-circle amplification (RCA) of begomovirus genomes (Inoue-Nagata *et al.*, 2004). RCA products were individually cleaved with *Bam*HI, *Cla*I, *Hin*dIII or *Pst*I restriction endonucleases and ligated to the pBluescript KS+ plasmid vector (Stratagene), previously cleaved with the same enzyme. Different enzymes were used to ensure that no virus was left uncloned because it lacked one or more of the sites (only ten samples had undigested RCA products after incubation with all four restriction enzymes). Viral inserts were sequenced commercially by primer walking at Macrogen. All genome sequences were organized to begin at the nicking site in the invariant nonanucleotide at the origin of replication (5′-TAATATT//AC-3′). Isolates were named to include information on the region of sampling (country and municipality) and year of isolation (last two numbers; Table S1).

Sequence comparisons. Full-length begomovirus genomes were assembled using CodonCode Aligner v. 4.1.1 (www.codoncode.com). Sequences were initially analysed with the BLAST*n* algorithm (Altschul *et al.*, 1990) and the GenBank non-redundant nucleotide database to determine the viral species with which they shared greatest identity. These similar sequences from GenBank were used to classify the novel isolates using the Species Demarcation Tool v. 1.0 (Muhire *et al.*, 2013).

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 using the MUSCLE algorithm (Edgar, 2004) and manually adjusted using Se-Al v. 2.0a11 (tree.bio.ed.ac.uk/software/seal/). ML trees were inferred for CP and Rep sequences using RAxML v. 7.0.3 (Stamatakis, 2006), assuming a general time reversible nucleotide substitution model with a gamma model of rate heterogeneity. The CP and Rep genes were chosen for analysis because of their essential role for insect transmission and viral replication, therefore being subject to stricter variability constraints (Rojas *et al.*, 2005). Additionally, they encompass ~70 % of the full-length DNA-A genome. The robustness of each individual branch was estimated from 2000 bootstrap replicates. Trees were visualized and edited using FigTree (tree.bio.ed.ac.uk/software/figtree) and Adobe Illustrator.

Genetic structure and variability indices. Partitioning of genetic variability and inferences about population structure were based on Wright's F fixation index (Wright, 1951) calculated using DnaSP v. 5.10 (Rozas *et al.*, 2003). Subpopulations were tested for structure according to host species and geographical location. The mean pairwise number of nucleotide differences per site (nucleotide diversity, π) was estimated for each population/subpopulation using DnaSP v. 5.10.

Recombination analysis. Possible parental sequences and recombination breakpoints were determined using the RDP, Geneconv, Bootscan, Maximum Chi Square, Chimaera, SisterScan and 3Seq methods implemented in RDP v. 3.44 (Martin *et al.*, 2010). Alignments were scanned with default settings for the different methods and statistical significance was inferred by a *P* value lower than a Bonferroni-corrected cut-off of 0.05. Only recombination events detected by at least three different methods were considered to be reliable. Putatively recombinant portions of genomes were BLASTed against the GenBank non-redundant nucleotide database to identify additional species that may have served as parental sequences. These were added to the datasets for additional analysis (Table S2). Evidence of non-tree-like evolution was assessed for CP and Rep datasets using the neighborNet method (Bryant & Moulton, 2004) implemented in SplitsTree v. 4.10 (Huson & Bryant, 2006). Images of networks were edited using Adobe Illustrator.

Detection of positive and negative selection at amino acid sites. To detect sites under positive and negative selection the CP and Rep datasets were analysed using four different ML-based methods available in the DataMonkey server (www.datamonkey.org): SLAC, FEL, REL, PARRIS (Kosakovsky-Pond & Frost, 2005; Scheffler *et al.*, 2006). To avoid spurious selection results caused by recombination, different predominantly non-recombinant clusters of sequences were defined based on phylogenetic and recombination analyses. The SLAC method was also used to estimate the mean d_N/d_S based on the inferred Genetic Algorithm Recombination Detection (GARD) (Kosakovsky-Pond & Frost, 2005) phylogenetic trees. A general time reversible nucleotide substitution model was assumed, and Bayes factors larger than 50 and a *P* value <0.1 were used as significance thresholds for the REL and FEL methods.

ACKNOWLEDGEMENTS

The authors wish to thank Drs Raquel Mello and Flávia Rabelo (Embrapa) for their support during sample collection. R. R. S. was the recipient of a CNPq doctoral fellowship and a CAPES sandwich PhD fellowship. S. D. was supported by the NSF of the USA (1026095) and funds from the NJ Agricultural Experiment Station. F. M. Z. was supported by Fapemig (APQ-00949-09) and CNPq (484447/2011-4).

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., Ogbe, F. O., Bandyopadhyay, R., Lava Kumar, P., Dixon, A. G. O., Hughes, Jd. & 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.

Albuquerque, L. C., Inoue-Nagata, A. K., Pinheiro, B., Resende, R. O., Moriones, E. & Navas-Castillo, J. (2012). Genetic diversity and recombination analysis of sweepoviruses from Brazil. *Virol J* **9**, 241.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

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.

Bonfim, K., Faria, J. C., Nogueira, E. O., Mendes, E. A. & Aragão, F. J. (2007). RNAi-mediated resistance to *Bean golden mosaic virus* in genetically engineered common bean (*Phaseolus vulgaris*). *Mol Plant Microbe Interact* **20**, 717–726.

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.

Bryant, D. & Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**, 255–265.

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.

Caciagli, P., Medina Piles, V., Marian, D., Vecchiati, M., Masenga, V., Mason, G., Falcioni, T. & Noris, E. (2009). Virion stability is important for the circulative transmission of *tomato yellow leaf curl Sardinia virus* by *Bemisia tabaci*, but virion access to salivary glands does not guarantee transmissibility. *J Virol* **83**, 5784–5795.

Castillo-Urquiza, G. P., Beserra, J. E. A., Jr, Bruckner, F. P., Lima, A. T. M., Varsani, A., Alfenas-Zerbini, P. & Murilo Zerbini, F. (2008). Six novel begomoviruses infecting tomato and associated weeds in southeastern Brazil. *Arch Virol* **153**, 1985–1989.

Costa, A. S. (1976). Whitefly-transmitted plant diseases. *Annu Rev Phytopathol* **14**, 429–449.

da 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.

Doyle, J. J. & Doyle, J. L. (1987). A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochem Bull* **19**, 11–15.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 1–19.

Faria, J. C. & Maxwell, D. P. (1999). Variability in geminivirus isolates associated with *Phaseolus* spp. in Brazil. *Phytopathology* **89**, 262–268.

Fernandes, F. R., Cruz, A. R. R., Faria, J. C., Zerbini, F. M. & Aragão, F. J. L. (2009). Three distinct begomoviruses associated with soybean in central Brazil. *Arch Virol* **154**, 1567–1570.

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.

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.

Ge, L. M., Zhang, J. T., Zhou, X. P. & Li, H. Y. (2007). Genetic structure and population variability of tomato yellow leaf curl China virus. *J Virol* **81**, 5902–5907.

Gilbertson, R. L., Faria, J. C., Hanson, S. F., Morales, F. J., Ahlquist, P. G., Maxwell, D. P. & Russell, D. R. (1991a). Cloning of the complete DNA genomes of four bean-infecting geminiviruses and determining their infectivity by electric discharge particle acceleration. *Phytopathology* **81**, 980–985.

Gilbertson, R. L., Hidayat, S. H., Martinez, R. T., Leong, S. A., Faria, J. C., Morales, F. J. & Maxwell, D. P. (1991b). Differentiation of bean-infecting geminiviruses by nucleic acid hybridization probes and aspects of bean golden mosaic in Brazil. *Plant Dis* **75**, 336–342.

González-Aguilera, J., Tavares, S. S., Sobrinho, R. 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.

Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D. & Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat Rev Microbiol* **11**, 777–788.

Höhnle, M., Höfer, P., Bedford, I. D., Briddon, R. W., Markham, P. G. & Frischmuth, T. (2001). Exchange of three amino acids in the coat protein results in efficient whitefly transmission of a nontransmissible *Abutilon mosaic virus* isolate. *Virology* **290**, 164–171.

Hou, Y. M. & Gilbertson, R. L. (1996). Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. *J Virol* **70**, 5430–5436.

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254–267.

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 archaebacteria. *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 φ29 DNA polymerase. *J Virol Met* **116**, 209–211.

Kheyr-Pour, A., Bananej, K., Dafalla, G. A., Caciagli, P., Noris, E., Ahoonmanesh, A., Lecoq, H. & Gronenborn, B. (2000). *Watermelon chlorotic stunt virus* from the Sudan and Iran: sequence comparisons and identification of a whitefly-transmission determinant. *Phytopathology* **90**, 629–635.

Koonin, E. V. & Ilyina, T. V. (1992). Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. *J Gen Virol* **73**, 2763–2766.

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.

Lefeuvre, P., Lett, J. M., Reynaud, B. & Martin, D. P. (2007a). Avoidance of protein fold disruption in natural virus recombinants. *PLoS Pathog* **3**, e181.

Lefeuvre, 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.

Lefeuvre, 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.

Lima, A. T. M., Sobrinho, R. R., González-Aguilera, J., Rocha, C. S., Silva, S. J. C., Xavier, C. A. D., Silva, F. N., Duffy, S. & Zerbini, F. M. (2013). Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. *J Gen Virol* **94**, 418–431.

Liu, L., Pinner, M. S., Davies, J. W. & Stanley, J. (1999). Adaptation of the geminivirus bean yellow dwarf virus to dicotyledonous hosts involves both virion-sense and complementary-sense genes. *J Gen Virol* **80**, 501–506.

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. & Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462–2463.

Martin, D. P., Lefeuvre, 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. (2006). History and current distribution of begomoviruses in Latin America. *Adv Virus Res* **67**, 127–162.

Morales, F. J. & Anderson, P. K. (2001). The emergence and dissemination of whitefly-transmitted geminiviruses in Latin America. *Arch Virol* **146**, 415–441.

Muhire, B., Martin, D. P., Brown, J. K., Navas-Castillo, J., Moriones, E., Zerbini, F. M., Rivera-Bustamante, R., Malathi, V. G., Briddon, R. W. & Varsani, A. (2013). A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus *Mastrevirus* (family *Geminiviridae*). *Arch Virol* **158**, 1411–1424.

Nash, T. E., Dallas, M. B., Reyes, M. I., Buhrman, G. K., Ascencio-Ibañez, J. T. & Hanley-Bowdoin, L. (2011). Functional analysis of a novel motif conserved across geminivirus Rep proteins. *J Virol* **85**, 1182–1192.

Navas-Castillo, J., Fiallo-Olivé, E. & Sánchez-Campos, S. (2011). Emerging virus diseases transmitted by whiteflies. *Annu Rev Phytopathol* **49**, 219–248.

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.

Ng, J. C. & Falk, B. W. (2006). Virus–vector interactions mediating nonpersistent and semipersistent transmission of plant viruses. *Annu Rev Phytopathol* **44**, 183–212.

Noris, E., Vaira, A. M., Caciagli, P., Masenga, V., Gronenborn, B. & Accotto, G. P. (1998). Amino acids in the capsid protein of tomato yellow leaf curl virus that are crucial for systemic infection, particle formation, and insect transmission. *J Virol* **72**, 10050–10057.

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.

Power, A. G. (2000). Insect transmission of plant viruses: a constraint on virus variability. *Curr Opin Plant Biol* **3**, 336–340.

Raja, P., Wolf, J. N. & Bisaro, D. M. (2010). RNA silencing directed against geminiviruses: post-transcriptional and epigenetic components. *Biochim Biophys Acta* **1799**, 337–351.

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.

Ribeiro, S. G., Martin, D. P., Lacorte, C., Simões, I. C., Orlandini, D. R. S. & Inoue-Nagata, A. K. (2007). Molecular and biological characterization of *Tomato chlorotic mottle virus* suggests that recombination underlies the evolution and diversity of Brazilian tomato begomoviruses. *Phytopathology* **97**, 702–711.

Rocha, C. S., Castillo-Urquiza, G. P., Lima, A. T. M., Silva, F. N., Xavier, C. A. D., Hora-Júnior, B. T., Beserra-Júnior, J. E. A., Malta, A. W. O., Martin, D. P. & other authors (2013). Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. *J Virol* **87**, 5784–5799.

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.

Rokyta, D. R. & Wichman, H. A. (2009). Genic incompatibilities in two hybrid bacteriophages. *Mol Biol Evol* **26**, 2831–2839.

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., Messeguer, 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.

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.

Silva, S. J. C., 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. (2012). Species diversity, phylogeny and genetic variability of begomovirus populations infecting leguminous weeds in northeastern Brazil. *Plant Pathol* **61**, 457–467.

Simon-Loriere, E., Galetto, R., Hamoudi, M., Archer, J., Lefeuvre, P., Martin, D. P., Robertson, D. L. & Negroni, M. (2009). Molecular mechanisms of recombination restriction in the envelope gene of the *Human immunodeficiency virus*. *PLoS Pathog* **5**, e1000418.

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.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690.

Tavares, S. S., Ramos-Sobrinho, R., Gonzalez-Aguilera, J., Lima, G. S. A., Assunção, I. P. & Zerbini, F. M. (2012). Further molecular characterization of weed-associated begomoviruses in Brazil with an emphasis on *Sida* spp. *Planta Daninha* **30**, 305–315.

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. *Virol J* **9**, 67.

Umaharan, P., Padidam, M., Phelps, R. H., Beachy, R. N. & Fauquet, C. M. (1998). Distribution and diversity of geminiviruses in Trinidad and Tobago. *Phytopathology* **88**, 1262–1268.

Varma, A. & Malathi, V. G. (2003). Emerging geminivirus problems: a serious threat to crop production. *Ann Appl Biol* **142**, 145–164.

Varsani, A., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Brown, J. K., Zerbini, F. M. & Martin, D. P. (2014). Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus, Eragrovirus* and *Turncurtovirus*. *Arch Virol* **159** [Epub ahead of print].

Wright, S. (1949). The genetical structure of populations. *Ann Eugen* **15**, 323–354.

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.

Wyant, P. S., Strohmeier, S., Schäfer, B., Krenz, B., Assunção, I. P., Lima, G. S. D. & Jeske, H. (2012). Circular DNA genomics (circomics) exemplified for geminiviruses in bean crops and weeds of north-eastern Brazil. *Virology* **427**, 151–157.