Evolutionary dynamics of bipartite begomoviruses revealed by complete genome analysis

César A. D. Xavier | Márcio T. Godinho | Talita B. Mar | Camila G. Ferro | Osvaldo F. L. Sande | José C. Silva | Roberto Ramos-Sobrinho | Renato N. Nascimento | Iraíldes Assunção | Gaus S. A. Lima | Alison T. M. Lima | F. Murilo Zerbini

1Dep. de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil
2Dep. de Bioquímica e Biologia Molecular/ BIOAGRO, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil
3Centro de Ciências Agrárias/ Fitossanidade, Universidade Federal de Alagoas, Rio Largo, Alagoas, Brazil
4Instituto de Ciências Agrárias, Universidade Federal de Uberlândia, Uberlândia, Minas Gerais, Brazil

Correspondence
F. Murilo Zerbini, Dep. de Fitopatologia/ BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil. Email: zerbini@ufv.br

Present address
César A. D. Xavier, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, North Carolina, USA
Talita B. Mar, Embrapa Trigo, Passo Fundo, Rio Grande do Norte, Brazil
Camila G. Ferro, Dep. de Fitopatologia e Nematologia, Universidade de São Paulo, Piracicaba, São Paulo, Brazil
Roberto Ramos-Sobrinho, School of Plant Sciences, University of Arizona, Tucson, Arizona, USA
Renato N. Nascimento, Dep. de Ciências Biológicas, Universidade Federal da Paraíba, Areia, Paraíba, Brazil

Funding information
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: Finance code 01; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 409599/2016-6; Fundação de Amparo à Pesquisa do Estado de Minas Gerais, Grant/Award Number: APQ-03444-16

Abstract
Several key evolutionary events marked the evolution of geminiviruses, culminating with the emergence of divided (bipartite) genomes represented by viruses classified in the genus Begomovirus. This genus represents the most abundant group of multipartite viruses, contributing significantly to the observed abundance of multipartite species in the virosphere. Although aspects related to virus-host interactions and evolutionary dynamics have been extensively studied, the bipartite nature of these viruses has been little explored in evolutionary studies. Here, we performed a parallel evolutionary analysis of the DNA-A and DNA-B segments of New World begomoviruses. A total of 239 full-length DNA-B sequences obtained in this study, combined with 292 DNA-A and 76 DNA-B sequences retrieved from GenBank, were analysed. The results indicate that the DNA-A and DNA-B respond differentially to evolutionary processes, with the DNA-B being more permissive to variation and more prone to recombination than the DNA-A. Although a clear geographic segregation was observed for both segments, differences in the genetic structure between DNA-A and DNA-B were also observed, with cognate segments belonging to distinct genetic clusters. DNA-B coding regions evolve under the same selection pressures than DNA-A coding regions. Together, our results indicate an interplay between reassortment and recombination acting at different levels across distinct subpopulations and segments.

KEYWORDS
divided genome, geminivirus, recombination, virus evolution
1 | INTRODUCTION

Although structurally simple, viruses possess complex evolutionary histories (Dolja & Koonin, 2018; Koonin et al., 2020; Solé, 2016). Throughout their evolutionary process, viruses incorporated several unique features, exhibiting a wide diversity of genome organizations, mechanisms of replication and gene expression strategies (Baltimore, 1971; Elena, 2016; Gale et al., 2000). This diversity allows viral populations to be dynamic, with a high adaptive capacity, infecting hosts in the three domains of life (Prangishvili, 2013; Weinbauer, 2004).

An intriguing aspect is the emergence of viruses with segmented genomes (Lister, 1966; Nee, 1987; Szathmary, 1992). Segmented genomes can be found in families of DNA and RNA viruses infecting animals, plants and fungi (Shi et al., 2016, 2018; Sicard et al., 2016). The evolution of segmented genomes involves not only the gain and loss of genes, but also the break-up and reunification of structural and nonstructural genes (Shi et al., 2016). Nevertheless, the biological and evolutionary significance of genome segmentation remains unclear (McDonald et al., 2016; Pressing & Reanney, 1984).

A special case of genome segmentation are viruses with multipartite genomes, which have their genome segments packed into separate particles. While each segment reaches complete physical independence, they are all needed to complete the infectious cycle (Lucia-Sanz & Manrubia, 2017). Therefore, multipartite genomes can suffer a high cost because in order for all segments to be present in the next cell or transmitted to a new host all segments must be in a high multiplicity of infection (MO); Garcia-Arriaza et al., 2004; Gutierrez & Zwart, 2018; Lucia-Sanz & Manrubia, 2017; Sanchez-Navarro et al., 2013; Sicard et al., 2013). It has been suggested that the lower number of errors introduced per replication round in small genome segments may counterbalance the disadvantageous effects of multipartite genomes (Nee, 1987). In addition, Ojosnegros et al. (2011) proposed that increased capsid stability due to the encapsidation of smaller genomic segments is a factor that can at least in part contribute to the maintenance of multipartite genomes. Evolutionarily, it has been proposed that reassortment (the exchange of genomic segments between strains or related species) in segmented viruses may represent a form of “sexual reproduction”, allowing the maintenance of genomic integrity through the elimination of deleterious mutations (Chao, 1991) and increasing population-level genotypic diversity (Turner, 2003).

Given the trade-offs between functional complementation and independence among the distinct segments in multipartite viruses, it would be expected that the different segments would be in an intimate process of co-evolution, with similar evolutionary histories. In support of this hypothesis, some studies indicated that reassortants between isolates of multipartite plant viruses did not become established in the population (Fraile et al., 1997; Grigoras et al., 2014). However, other studies suggested that the different segments of multipartite genomes may experience different evolutionary histories (Briddon et al., 2010; Hu et al., 2007; Lozano et al., 2009; Roossinck, 2002). These conflicting results could be related to varying degrees of interdependency among the proteins encoded by genomic segments of distinct viruses.

The genus Begomovirus (family Geminiviridae) is comprised of viral species with one or two genomic segments of circular, single-strand DNA (ssDNA) of about 2,700 nucleotides (Zerbini et al., 2017). Beyond their economic importance (Rojas et al., 2018), two main characteristics make begomoviruses an attractive model to study the evolutionary dynamics of viral populations. First, begomoviruses are fast-evolving viruses, with substitution rates as high as those of ssRNA viruses (Duffy & Holmes, 2008, 2009; Ge et al., 2007). Second, begomoviruses include viruses with nonsegmented (monopartite) as well as segmented (bipartite) genomes. The bipartite nature of these viruses has been little explored in evolutionary studies, and little is known about the evolutionary dynamics of the different segments.

The two genomic segments of bipartite begomoviruses are referred to as DNA-A and DNA-B. The DNA-A contains five genes involved in replication (Rep, ReN), transcriptional control and suppression of host defences (TrAP, AC4) and encapsidation of the viral progeny (CP). The DNA-B contains two genes (MP, NSP) required for movement in the plant (Figure 1). The two segments share an intergenic region (IR) of ~200 nucleotides which includes the viral replication origin (v-ori) and conserved repeat sequences (iterons) that are specifically recognized by the viral replication-associated protein, Rep (Arguello-Astorga et al., 1994; Fontes et al., 1994). The IR is important for maintaining the integrity of the bipartite genome, allowing both segments to be replicated since Rep proteins show high specificity for their cognate ori (Rojas et al., 2005).

A number of studies have been conducted to determine and quantify the mechanisms dictating the evolutionary dynamics of begomoviruses (Duffy & Holmes, 2008, 2009; García-Andrés et al., 2007; Ge et al., 2007; Lefeuvre et al., 2010; Lima et al., 2013, 2017; Mar et al., 2017; Rocha et al., 2013; Sanz et al., 1999). These studies were based on analysis of nonsegmented viruses or of the DNA-A of bipartite viruses. Considering the important role played by the DNA-B in the infection process, evolutionary studies that consider this segment are needed. The analysis of the complete genome can provide a more accurate picture of the evolution of begomovirus and a better understanding of the evolution of multipartite genomes in general.

Here, using a population genetics approach, we performed a parallel evolutionary analysis of the DNA-A and DNA-B segments of NW begomoviruses from cultivated and noncultivated hosts. Our results indicate an interplay between reassortment and recombination acting at different levels across distinct subpopulations and genomic segments.

2 | MATERIALS AND METHODS

2.1 | Cloning and sequencing of DNA-B segments

Begomovirus DNA-B clones were obtained from total DNA extracted from the samples collected for the studies of Lima et al. (2013), Rocha et al. (2013) and Ramos-Sobrinho et al. (2014), which
have been stored in our laboratory at –80°C. Total DNA was used as a template for rolling-circle amplification (RCA) of viral genomes as described by Inoue-Nagata et al. (2004). To facilitate the cloning of DNA-B segments, restriction analysis of the previously cloned DNA-As was performed using Ape 2.0 Plasmid Editor, and only restriction enzymes that do not cleave the cognate DNA-A were chosen. Unit genome-length fragments (~2,600 nucleotides) were excised and ligated into the pBLUESCRIPT-KS+ (pKS+) plasmid vector (Stratagene), previously cleaved with the same enzyme. Viral inserts were sequenced commercially (Macrogen Inc.). Full-length begomovirus genomes were assembled using Geneious v. 8.1 (Kearse et al., 2012). Sequences were analysed with the BLAST algorithm (Altschul et al., 1990) to confirm that they corresponded to a DNA-B.

2.2 Multiple sequence alignments and phylogenetic analysis

All genome sequences were organized to begin at the nicking site at the origin of replication (5’-TAATATT//AC-3’). Multiple sequence alignments were prepared for the full-length nucleotide sequences of DNA-A and DNA-B, of the CP, Rep, REn, TrAP, AC4, MP and NSP genes, and of the DNA-A intergenic region (IR-A), DNA-B large intergenic region (LIR-B) and DNA-B small intergenic region (SIR-B), using the MUSCLE option in MEGA6 (Tamura et al., 2013). The same alignments were used for all subsequently performed analyses.

Phylogenetic trees were constructed based on the full-length DNA-A and DNA-B nucleotide sequence using Bayesian inference performed with MrBayes v. 3.0b4 (Ronquist & Huelsenbeck, 2003). MrModeltest v. 2.2 (Nylander, 2004) was used to select the nucleotide substitution model with the best fit for each data set in the Akaike Information Criterion (GTR+I+G for all data sets except BGMV DNA-B, GTR+G). The analyses were carried out running 10,000,000 generations and excluding the first 2,500,000 generations. Trees were visualized and edited using FigTree (tree.bio.ed.ac.uk/software/figtree/). For data sets that showed evidence of recombination, trees with and without recombinant blocks were generated. The regions corresponding to the recombinant blocks were replaced by missing data (?), using a script (available from the authors upon request) written in R software (R Core Team, 2017).

2.3 Variability indices and genetic structure

The average pairwise number of nucleotide differences per site (nucleotide diversity, π) was calculated using a script written in R software (available from the authors upon request). The statistical significance of the differences amongst the mean π obtained from different data sets was calculated by estimating their 95% bootstrap confidence intervals from 1,000 nonparametric simulations using the simpleboot statistical package in R software (Peng, 2008) as previously described (Lima et al., 2017). This method was
chosen because it does not make assumptions about the distribution of data and allows comparing data sets with different sample sizes (Lima et al., 2017). Nucleotide diversity was also calculated on a 100-nucleotide sliding window with a step size of 10 nucleotides across the full-length DNA-A and DNA-B nucleotide sequence for each data set. The number of segregating sites (S), number of haplotypes (H) and haplotype diversity (Hd) were estimated for each data set using dnasp v. 5.10 (Rozas et al., 2003).

Inferences about population genetic structure were performed using discriminant analysis of principal components (DAPC) (Jombart et al., 2010) implemented in the Adegenet package (Jombart & Ahmed, 2011) in R software. DAPC is a model-free method, with no assumptions about Hardy-Weinberg equilibrium or linkage disequilibrium, thus being ideal for viruses. For the full-length DNA-A and DNA-B, the number of genetic clusters (k) were predefined using a k-means algorithm. To find the best number of genetic clusters to describe the data, the k-means was run sequentially with values of k varying from 2 to 10, retaining all principal components (PC). The choice of the best clustering model was based primarily on the minimum number of k after which the Bayesian information criterion (BIC) decreases by an insignificant amount, as proposed by Jombart et al. (2010). In addition, to support and verify the degree of differentiation between genetic clusters inferred by DAPC, the coefficient of nucleotide differentiation (Nst) was estimated using dnasp v. 5 (Rozas et al., 2003).

2.4 Recombination and reassortment analysis

The occurrence of recombination was analysed for the DNA-A and DNA-B separately using intraspecific data sets. Recombination analysis was performed using the Recombination Detection Program v. 4 (Martin et al., 2015). Alignments were scanned with default settings for the different methods. Statistical significance was inferred by p-values lower than a Bonferroni-corrected cutoff of 0.05. Only recombination events detected by at least four different methods were considered reliable.

To assess the occurrence of reassortment and compare global congruence across different data sets, the topological congruence between the inferred Bayesian phylogenetic trees based on the DNA-A and DNA-B were analysed with Procrustean Approach to Cophylogeny (PACo). While both genomic segments are necessary for bipartite begomoviruses to systemically infect the host and be transmitted, which segment is driving the evolution of the other is unclear. Thus, in our cophylogenetic analyses, we tested the dependence of both phylogenies on each other by setting the symmetric argument of the paco function as true (Hutchinson et al., 2017). Matrices of patristic distances from DNA-A and DNA-B phylogenies were calculated using the ape package (Paradis et al., 2004) and transformed into principle coordinates (PCo) using the paco package (Hutchinson et al., 2017), both in R software. The PCo and DNA-A-DNA-B association matrices were used for procrustean superimposition analysis to produce the residual square sum (m^2_{XY}), which is inversely proportional to the level of global congruence between the phylogenies. The contribution of each individual DNA-A and DNA-B link for the global congruence and 95% confidence intervals were estimated using a jackknife method. The statistical significance of global congruence was assessed with 999 permutations in r software using the paco package (Hutchinson et al., 2017). The assumption behind this approach is that incongruence between DNA-A and DNA-B phylogenies is due to reassortment events. In addition, to access the statistical significance of the differences in global congruence across species data sets, 95% confidence intervals were estimated for the mean normalized residuals links, from 1,000 nonparametric simulations in R software using the Simpleboot statistical package. Individual contributions of DNA-A-DNA-B links for the global congruence statistic were normalized by dividing them for the global m^2_{XY} (Pérez-Escobar et al., 2015). Since recombination can also lead to incongruence between phylogenies, analyses were performed in data sets with and without recombinant blocks. Reassortment was also analysed by concatenating the sequences of the DNA-A and DNA-B and scanning sequences for recombination breakpoints near the artificial joint between the two genomic segments using RDP as described previously.

2.5 Selection analysis

Potentials sites under positive and negative selection in coding regions of the CP, Rep, TrAP, REn, AC4, MP and NSP genes were identified using four distinct methods: single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), random effects likelihood (REL) and partitioning for robust inference of selection (PARRIS), implemented in the DataMonkey webservice (Kosakovsky-Pond & Frost, 2005). The mean ratios of nonsynonymous to synonymous substitutions (dN/dS) were estimated for all genes using the SLAC method. All methods were applied using nucleotide substitution models with the best fit for each data set determined in the DataMonkey webservice. To avoid misleading selection results, we searched for recombination breakpoints in each data set using genetic algorithm recombination detection (GARD) (Pond et al., 2006). All analyses were based on the inferred GARD-corrected phylogenetic trees. To access the statistical significance of the differences in selection pressure, 95% confidence intervals were estimated for the mean dN/dS ratio and for the average number of negatively and positively selected sites, from 1,000 nonparametric simulations in r software using the Simpleboot statistical package.

3 RESULTS

A total of 239 full-length DNA-B sequences were obtained in this study, all from samples of cultivated and noncultivated plants from which a DNA-A had been previously cloned and analysed (Lima et al., 2013; Rocha et al., 2013; Ramos-Sobrinho et al., 2014). These sequences were combined with DNA-A and DNA-B sequences
retrieved from GenBank. Detailed information on the samples and the corresponding DNA-A and DNA-B sequences are presented in Figure 1 and Tables S1 and S2. The global data set included sequences of bean golden mosaic virus (BGMV), Blainvillea yellow spot virus (BlYSV), Euphorbia yellow mosaic virus (EuYMV), Macroptilium yellow spot virus (MaYSV) and tomato severe rugose virus (ToSRV), with a total of 292 DNA-A and 315 DNA-B sequences (Figure 1). For each species data set, the number of DNA-A and DNA-B sequences was similar ($\chi^2 = 2.2432, P = 0.6911$).

### 3.1 Phylogenetic analysis

Bayesian phylogenetic trees based on full-length DNA-A and DNA-B nucleotide sequences were constructed for each species data set. Although the PACo analysis provided significant evidence for global congruence between DNA-A and DNA-B trees ($p \leq 0.001$; Figure 2a), the residual square sum ($m^2_{XY}$) values were significantly variable across the different data sets (Figure 2b), indicating that congruence levels are variable among them.

DNA-A and DNA-B phylogenetic trees for BGMV were perfectly symmetric, yielding "mirror image" trees (Figure 3a). In both trees it is possible to visualize two major clades supported by high posterior probability values and separated by long branches (Figure 3a). The first clade (DF/GO/MG) is comprised of isolates collected in Distrito Federal (DF), Góias (GO) and Minas Gerais (MG) states, and the second clade (AL) is comprised of isolates sampled in Alagoas (AL) state (Figure 3a).

Although the BlYSV data set contains isolates sampled in distant geographical regions (~1800 km apart), phylogenetic trees do not display a clear and consistent pattern of geography-based clustering as noted for other species data sets (Figure 3b). There was a tendency in isolates collected in the northeastern states (AL and Bahia, BA) to form separate clades from isolates collected in the southeast (MG), however short internal branches can be observed, suggesting a low level of differentiation. In addition, recombination events were detected in DNA-A segments among isolates of the northeast with putative major parents in the southeast (Table S3). For the DNA-B, isolates sampled in the southeast showed recombination events with putative minor parents in the northeast (Table S3). These results suggest that, despite the greater distance between sampled sites, some connection between the two regions exists, resulting in an incipient pattern of differentiation. Sampling of intermediate regions between AL and MG may provide a better picture of the genetic make-up of the BlYSV population.

In agreement with Mar et al. (2017), a near perfect segregation based on geography was observed in the EuYMV DNA-A and DNA-B trees (Figure 3c). For the DNA-A, two major clades were well supported, with isolates sampled in Rio Grande do Sul (RS) and Paraná states (PR) clustered separately from those sampled in Góias (GO) and Mato Grosso do Sul (MS). The same clusters could be observed in the DNA-B tree, except that isolates sampled in MS clustered with those from RS and PR, suggesting a reassortment event.

Due to the complex recombination profile of the MaYSV data set (see below), DNA-A and DNA-B phylogenetic trees showed a high degree of incongruence and displayed no obvious pattern of...
segregation (Figure 3d). For the DNA-A tree two well-supported major clades can be observed, while the DNA-B tree shows several poorly supported clades (Figure 3d). To minimize the effect of recombination, phylogenetic trees were reconstructed based on recombination-free DNA-A and DNA-B data sets (Figure 3e). Two well-supported clades based on geographical location can be observed in the DNA-A tree, with isolates sampled in Pernambuco (PE) state clustered separately from those collected in neighbouring AL (Figure 3e). Although two well-supported clades were also observed in the DNA-B tree, these were not perfectly congruent with the DNA-A tree (Figure 3e). Only two DNA-B haplotypes were sampled in PE, with the first (isolates BR-PE-Cau-22-3, BR-PE-Cau-22-4 and BR-PE-Cau-22-5) clustering together with BR-Sti-34-11 sampled in AL, and the second (BR-PE-Cau-23-1) clustering with the remaining sequences also sampled in AL. Therefore, a larger number of haplotypes sampled in PE will be necessary to obtain a better resolution of the pattern of DNA-B segregation.

Phylogenetic trees for ToSRV DNA-A and DNA-B were not completely congruent (Figure 3f). A clear geographical clustering was observed for the DNA-B tree, with three clades comprised almost exclusively of isolates from Coimbra, Carandai and Florestal/2014. A fourth cluster contained isolates from Florestal/2008, indicating a remarkable temporal structure considering the short period of time between samplings (only six years). The DNA-A tree displayed the same topology as the DNA-B tree for the isolates from Florestal, but isolates from Coimbra and Carandai were not segregated (Figure 3f). Interestingly, the DNA-B of isolate BR-Flo01-14 from Florestal clustered together with isolates sampled in Coimbra, due to a recombination event involving parental sequences from these two locations (Table S3). This indicates that the exchange of segments between isolates from Coimbra and Florestal has occurred.

### 3.2 | Genetic structure of populations

To compare the genetic structure based on DNA-A and DNA-B segments, a nonparametric multivariate statistical analysis was performed using DAPC (Figure 4; Table S4). For most data sets it was possible to define clear genetic clusters. However, for BiYSV DNA-A and DNA-B and MaYSV DNA-B data sets, in agreement with phylogenetic analysis, it was not possible to define consistent genetic clusters (Figure S1). Therefore, for these three segment data sets all individuals were considered to constitute a single population in later analyses.

Although a clear geographical segregation can be observed for the two segments, especially for BGMV, EuYMV and ToSRV (consistent with phylogenetic analysis), slight differences between the genetic structure based on DNA-A and DNA-B could be observed, with cognate segments belonging to distinct genetic clusters.

For BGMV, DAPC analysis discriminated five and four genetics clusters for DNA-A and DNA-B, respectively, primarily based on geography (Figure 4). The clusters named AL-1, AL-2 and MG-1 were observed for both segments, displaying a perfect pattern based on sampling location. For isolates sampled in the midwestern region, two genetic clusters (MW-1 and MW-2) and a single cluster (MW) were detected for DNA-A and DNA-B, respectively. Interestingly, no clear pattern was observed to explain the formation of MW-1 and MW-2 clusters, with all four subpopulations from the midwest including DNA-A sequences from both genetic clusters (Figure 4). To try to understand what could be influencing this process, we checked which sites in the DNA-A were contributing the most to the observed structure. We found 22 contributing sites located between the C-terminal region of Rep and the C-terminal region of CP, mostly displaying synonymous substitutions (Table S5). This suggests the action of genetic drift, rather than positive selection, leading to a change in the frequency of the two genotypes in the different subpopulations.

Similar to the patterns observed for BGMV, a clear geographical segregation was also observed for both EuYMV segments, with only minor differences in the assignment of some DNA-A and DNA-B segments to the three genetic clusters (South-1, South-2 and GO) (Figure 4).

Although it was not possible to define clear genetic clusters for MaYSV DNA-B segments based on DAPC analysis, the DNA-A segments were split into four groups (Figure 4). Isolates from PE clustered separately from isolates sampled in AL. Moreover, isolates from AL grouped into three separate genetic clusters (AL-1, AL-2, AL-3) with no association with sampling location or host. When the data were analysed without recombinant blocks, only two clusters were formed, with isolates sampled in AL forming a single cluster separately from those sampled in PE (Figure S2).

For ToSRV it was possible to discriminate four clusters for both segments. However, there was no perfect correspondence between DNA-A and DNA-B clusters in the same locations. The MG-1 and MG-2 clusters showed the same composition for both segments, comprised of isolates sampled in Florestal in 2008 and 2014, respectively. The DNA-A segments sampled in Coimbra were split into two genetic clusters (MG-3 and MG-4), while DNA-A segments from Carandai were assigned only to the MG-3 cluster (Figure 4). For the DNA-B, in agreement with the phylogenetic analysis, isolates
sampled in Carandai and Coimbra were assigned to MG-3 and MG-4 clusters, respectively, except for BR-Coi179-14 from Coimbra which was assigned to the MG-3 cluster (Figure 4). These results suggest that asymmetric flow of DNA-A segments can occur from Carandai to Coimbra, but the opposite is not necessarily true, since the Carandai subpopulation was entirely comprised of individuals belonging to same genetic cluster. In addition, there is not a large mixture of DNA-B segments, suggesting that the exchange may be restricted to the DNA-A.

To assess the consistence of the genetic clusters inferred by DAPC analysis, the differentiation degree among the clusters was estimated by means of the coefficient of nucleotide differentiation (Nst) (Table 1). In general, high values of pairwise Nst were observed among different clusters, indicating great differentiation, and fully supporting the subdivision proposed by DAPC analysis.

3.3 | Genetic variability

Although there is evidence that DNA-B segments are more variable than DNA-A segments (Briddon et al., 2010), a quantitative and comparative analysis using large intraspecific data sets had not yet been performed. To address this issue, we calculated the average pairwise number of nucleotide differences (nucleotide diversity index, \( \pi \)) for the DNA-A and DNA-B segments from the five species data sets. The \( \pi \) values for the DNA-B were statistically higher than those for the DNA-A for almost all species data sets (Figure 5a; Tables S6 and S7). The exception was MaYSV, for which the value for the DNA-A (\( \pi = 0.07170 \)) was statistically higher than for the DNA-B (\( \pi = 0.05567 \)) (Figure 5a; Tables S6 and S7). These results confirm the highest variability of the DNA-B in relation to the DNA-A at the populational level.
The nucleotide diversity index was also calculated on a sliding window across the DNA-A and DNA-B for each species data set (Figure 5). The results show an uneven distribution of variation along the segments for almost all data sets. The exception was the ToSRV DNA-A (and to a lesser extent also its DNA-B), which in addition to presenting a very low degree of variation, this variation is evenly distributed along the genome (Figure 5). To verify the statistical significance of the uneven distribution of variation along the genomic segments, the genome was partitioned and nucleotide sequences of coding and non-coding regions were compared for each species data set (Figure S3). Unsurprisingly, non-coding regions were more variable compared to coding regions (Figure S3). The SIR-B was more variable than the LIR-B, which in turn was more variable than the IR-A.

### Table 1

Results of subdivision test performed for bean golden mosaic virus (BGMV), Euphorbia yellow mosaic virus (EuYMV), Macroptilium yellow spot virus (MaYSV) and tomato severe rugose virus (ToSRV) data sets

| Virus   | Genomic segment | DNA-A | DNA-B |
|---------|-----------------|-------|-------|
|         | Cluster 1 | Cluster 2 | Nst   | Cluster 1 | Cluster 2 | Nst   |
| BGMV    | MW-2     | MW-5     | 0.704 | MW-2     | MW-5     | 0.704 |
|         | MW-2     | MG-1     | 0.922 | MW-2     | MG-1     | 0.922 |
|         | MW-2     | AL-2     | 0.966 | MW-2     | AL-2     | 0.966 |
|         | MW-2     | AL1-1    | 0.975 | -        | -        | -     |
|         | MW-5     | MG-1     | 0.908 | -        | -        | -     |
|         | MW-5     | AL-2     | 0.964 | -        | -        | -     |
|         | MW-5     | AL-1     | 0.973 | -        | -        | -     |
|         | MG-1     | AL-2     | 0.971 | MG-1     | AL-2     | 0.971 |
|         | MG-1     | AL-1     | 0.981 | MG-1     | AL-1     | 0.981 |
|         | AL-2     | AL-1     | 0.695 | AL-2     | AL-1     | 0.695 |
| EuYMV   | GO       | SO-1     | 0.487 | GO       | SO-1     | 0.487 |
|         | GO       | SO-2     | 0.640 | GO       | SO-1     | 0.640 |
|         | SO-1     | SO-2     | 0.442 | SO-1     | SO-2     | 0.442 |
| MaYSV   | AL-2     | AL-1     | 0.834 | -        | -        | -     |
|         | AL-2     | PE       | 0.827 | -        | -        | -     |
|         | AL-2     | AL-3     | 0.801 | -        | -        | -     |
|         | AL-1     | PE       | 0.929 | -        | -        | -     |
|         | AL-1     | AL-3     | 0.726 | -        | -        | -     |
|         | PE       | AL-3     | 0.895 | -        | -        | -     |
| ToSRV   | MG-3     | MG-4     | 0.420 | MG-3     | MG-4     | 0.420 |
|         | MG-3     | MG-1     | 0.718 | MG-3     | MG-1     | 0.718 |
|         | MG-3     | MG-2     | 0.623 | MG-3     | MG-2     | 0.623 |
|         | MG-4     | MG-1     | 0.752 | MG-4     | MG-1     | 0.752 |
|         | MG-4     | MG-2     | 0.703 | MG-4     | MG-2     | 0.703 |
|         | MG-1     | MG-2     | 0.776 | MG-1     | MG-2     | 0.776 |

*Although there are subpopulations with the same names in the DNA-A and DNA-B data sets of some viruses, the composition of isolates in these clusters is different. AL, Alagoas; GO, Goiás; MG, Minas Gerais; MW, Midwest; PE, Pernambuco; SO, South

*Values from 0 to 0.05 indicate little genetic differentiation; from 0.05 to 0.15, moderate differentiation; from 0.15 to 0.25, great differentiation; >0.25, high differentiation.

3.4 Recombination analysis

It is well established that recombination plays an important role in the diversification and evolution of begomoviruses (Lefeuvre et al., 2007; Lima et al., 2017; Monci et al., 2002). However, most studies were based on analysis of nonsegmented viruses or of the DNA-A segment of bipartite viruses, and little is known about the effect of recombination on the DNA-B segment (Lefeuvre & Moriones, 2015).

There is a clear correlation between the number of intraspecific recombination events and genetic variability. BlYSV and MaYSV, which showed the higher number of events for both genomic segments, are also the viruses with the higher genetic variability. Likewise, viruses with low genetic variability (BGMV, EuYMV and especially ToSRV) had few or no recombination events detected.
For all species data sets, the DNA-B was more prone to recombination than the DNA-A, with a higher number of unique, well-supported recombination events (detected by at least four methods and a \( p < 0.001 \)) (Figure 6a,b; Table S3). Considering all unique events, 32 recombination events were detected in the DNA-B data sets and only 10 in the DNA-A data sets. No recombination events were detected for the DNA-A data sets from BGMV, EuYMV and ToSRV, whereas three and seven unique events were detected for the DNA-A data sets from BlYSV and MaYSV, respectively (Figure 6a). The MaYSV DNA-B was the most prone to recombination, with 13...
unique recombination events, followed by BIYSV, EuYMV, BGMV and ToSRV, with nine, five, four and one unique events, respectively (Figure 6b). These results demonstrate an asymmetric distribution of recombination events between genomic segments, with a higher propensity of the DNA-B to recombine compared to the DNA-A, and also suggest that populations of begomoviruses from noncultivated hosts are more prone to recombination, as shown for MaYSV and BIYSV and to a lesser extent for EuYMV DNA-B.

**FIGURE 6** Recombination events detected in the (a) DNA-A and (b) DNA-B segments of begomoviruses using RDP4 (Martin et al., 2015). The genome map at the top of the figure corresponds to the schematic representations of the sequences below. Regions highlighted in black correspond to the donated (minor parent) portion, while the remaining portion corresponds to the receiving (major parent) sequence (see Table S4 for more details on recombination events). IR, intergenic region in the DNA-A; LIR, large intergenic region and SIR, small intergenic region, both in the DNA-B. N, Number of isolates containing each event. (c) Truncated Rep (tRep) ORFs present in the DNA-B of Macroptilium yellow spot virus (MaYSV) isolates. For each type of tRep insertion (type I, type II and type III), the DNA-B is represented with genes indicated by arrows (MP in the viral sense, NSP in the complementary sense), and the portion of the tRep ORF filled in black represents homologous sequences to the N-terminal portion of the Rep protein in the cognate DNA-A. Each DNA-B map indicates the amino acid sequence alignments of the tRep ORF and the Rep protein N-terminal region of the cognate DNA-A (identical residues in black, similar residues in grey). (d) Nucleotide sequence alignment of the DNA-B LIR containing cis-acting elements. The nonanucleotide at the origin of replication is highlighted in bold and underlined. Putative Rep-binding elements (iterons) are shaded in grey and their orientation is indicated by arrows. Putative cis-acting elements found in eukaryotic promoters located in the LIR are marked by dotted boxes located at the right side of the tRep start codon (indicated in bold with asterisks).
Recombination breakpoints were mapped across the DNA-A and DNA-B. Overall, a nonrandom distribution of recombination breakpoints was observed along both segments. In agreement with previous studies (Lefevre et al., 2007), most recombination events detected in the DNA-A involved breakpoints located in the Rep gene (10 out of 20), followed by intergenic region (IR-A; 6/20). Conversely, few recombination breakpoints were located in the CP (2/20) and at the interface between the TrAP and REn genes (2/20). Most recombination events in the DNA-B have breakpoints in the LIR-B (37 out of 64), especially upstream of the initiation codon of the MP gene. In contrast, the SIR-B showed only 2/64 recombination breakpoints. Despite approximately one third of the DNA-B length being comprised of intergenic regions, the number of breakpoints maped to intergenic and coding regions was similar ($x^2=3.0625, P = 0.08012$).

### 3.5 Fragments of Rep ORFs in the large intergenic regions of the DNA-B (LIR-B) of MaYSV

Interestingly, small ORFs in the complementary-sense strand of the LIR-B of several MaYSV isolates are homologous to the N-terminal region of the Rep gene located in its DNA-A (Figure 6c; Table 2). These small ORFs were named truncated Rep (tRep) and, based on their common features, were grouped into three types named tRep I, II and III (Table 2). tRep ORFs were detected in 10 MaYSV isolates (42%) collected in three different regions (Table 2), ruling out the possibility of them being cloning artefacts. Despite their size ranging from 35 to 95 amino acids, the conserved Motif I (FLTYP) and iteron related domains were detected in all three types (Figure 6c). Interestingly, all tRep ORFs were located upstream of the start codon and downstream of the promoter of the MP gene (Figure 6d), suggesting that they may be transcribed and translated, and possibly interfering with expression of the MP gene. Although a recombination analysis was not performed (it was not possible to obtain a good alignment between nonhomologous segments), the presence of these tRep ORFs in the LIR-B is a strong indication of intersegment recombination.

### 3.6 Reassortment analysis

As mentioned above for phylogenetic reconstruction, PACo analysis provided significant evidence for the global congruence between DNA-A and DNA-B trees (Figure 2). However, it does not rule out the possibility of reassortment, since distance-based methods do not take into account the degree of congruence between links when accepting the hypothesis of global congruence between two trees (de Vienne et al., 2013). Indeed, the residual square sum ($m^2_{XY}$) values were significantly variable across the different species data sets (Figure 2b), ranging from 0.124 for BGMV to 0.522 for MaYSV, indicating that congruence levels are variable among different species data sets and suggesting that reassortment can occur at different frequencies across data sets.

To verify the occurrence of potential reassortment events, the individual contribution of each DNA-A and DNA-B link for the global square sum was assessed (Figure 7). Analysis of the estimated jackknifed squared residuals for each link clearly indicated the occurrence of reassortment in all species data sets (Figure 7). Jackknifed squared residuals values were grouped according to sampling location and compared by estimating their 95% bootstrap confidence intervals from 1,000 nonparametric simulations. Some groups showed significantly higher residual values, indicative of a higher contribution to incongruence and a higher frequency of reassortant genomes. For example, for the EuYMV data set, significantly higher residual values were observed for sequences sampled in RS and MS.

**Table 2** Features of truncated Rep (tRep) ORFs located in the complementary-sense strand of the DNA-B large intergenic region (LIR-B) of Macroptilium yellow spot virus (MaYSV) isolates

| Isolate  | Type | Length of ORF (nt/aa) | Location in the genome | Coverage (% nt/aa) | Identity (% nt/aa) | E-value (nt/aa) |
|----------|------|----------------------|------------------------|--------------------|--------------------|-----------------|
| BR:Oaf8:11 | I    | 105/35               | 2471-2366              | 97/100             | 91/88              | $4e^{-31}/6e^{-12}$ |
| BR:Crb10:11 | I    | 108/36               | 2472-2364              | 98/100             | 97/94              | $2e^{-41}/1e^{-13}$ |
| BR:Sti34:11 | I    | 105/35               | 2468-2363              | 97/100             | 98/94              | $6e^{-41}/1e^{-12}$ |
| BR:Sti2:11  | II   | 162/54               | 2473-2312              | 82/96              | 92/80              | $2e^{-44}/4e^{-18}$ |
| BR:Sti8:11  | II   | 162/54               | 2470-2308              | 82/81              | 97/93              | $3e^{-54}/9e^{-12}$ |
| BR:Sti4:11  | II   | 162/54               | 2473-2312              | 71/79              | 90/79              | $3e^{-34}/7e^{-17}$ |
| BR:Sti26:11 | II   | 174/58               | 2486-2312              | 62/87              | 90/68              | $6e^{-32}/3e^{-11}$ |
| BR:Sti35:11 | II   | 162/54               | 2470-2308              | 82/81              | 99/98              | $6e^{-57}/5e^{-20}$ |
| BR:Crb1:11  | III  | 285/95               | 2439-2154              | 48/47              | 99/100             | $1e^{-61}/1e^{-23}$ |
| BR:Crb2:11  | III  | 285/95               | 2439-2154              | 48/47              | 99/100             | $1e^{-61}/1e^{-23}$ |

*aNumbering starts at the first nucleotide after the cleavage site at the origin of replication and increases clockwise.

*bPercentage of tRep ORF (nt, nucleotides; aa, amino acids) which is homologous to the Rep protein N-terminal region of the cognate DNA-A.

Analyses performed using the BLASTn and BLASTp algorithms for nt and aa sequences, respectively.
compared to GO and PR (Figure 7). Conversely, for the BIYSV data set there was no difference in residual values between sequences sampled at different locations, indicating no differences in reassortment frequency (Figure 7). Furthermore, most reassortment events occurred between isolates from the same or close regions.

Analysis of concatenated sequences confirmed the results obtained by the topological congruence test (Table 3; Table S8). For almost all data sets, most of the reassortment events detected by RDP were supported by PACo analysis, displaying high squared residuals (Figure 7). The frequency of reassortment sequences detected by RDP, in agreement with PACo global test, was higher for MaYSV (19/25; 76%), followed by EuYMV (29/55; 53%), BGMV (35/117; 29%) and BIYSV (5/24; 21%) (χ² = 25.881, P = 1.01 × 10⁻⁵; Table 3). For ToSRV only two sequences were detected as putatively reassortant by RDP. However, PACo analysis suggests a greater number of reassortment events, with 16/67 links (24%) showing squared residual values greater than two times the median value (indicative of strong incongruence). This discrepancy between RDP and PACo analyses may be due to the high sequence identity among isolates of the ToSRV data set (97%–100% and 95%–100% for DNA-A and DNA-B, respectively), impairing the capacity of RDP to detect reliable reassortment events.

Interestingly and surprisingly, when data sets including recombinant blocks were analysed, the global square sum was slightly lower compared to the same data sets without recombinant blocks (Figure 2a). Although significant differences in the global congruence were not observed (Figure 2b), this result suggests that recombination may be restoring the congruence of specific links between the DNA-A and DNA-B phylogenies. For example, recombination after a migration event followed by reassortment could restore the congruence, but this needs to be further investigated (Geoghegan et al., 2017).

### 3.7 Selection analysis

To understand the possible effect of selection pressure on the uneven distribution of variation observed between DNA-A and DNA-B, the ratio of nonsynonymous to synonymous substitutions (dN/dS) was compared for each gene/genomic segment. The dN/dS ratios were <1 for almost all genes, except AC4 (Table 4), indicating the predominance of negative selection acting on both segments. In agreement with these results, a higher number of individual sites under negative selection was detected in all species data sets (Table 4).
Although negative selection is predominant, dN/dS ratios were highly variable among genes/genomic segments (0.028–2.530 for the DNA-A, 0.0667–0.2311 for the DNA-B) (Table 4). These results indicate that different genes/genomic segments can evolve under distinct selection pressures and, unexpectedly, that DNA-B coding regions evolve under equal or more strict selection compared to DNA-A coding regions. When the dN/dS ratio was compared between DNA-A and DNA-B data set excluding the AC4 and TrAP genes, there was no significant difference between dN/dS ratio (data not shown). Thus, it is suggested that DNA-B coding regions evolve at least under equivalent selection pressure as those of the DNA-A.

In agreement with the high dN/dS ratio of the AC4 gene of BGMV and EuYMV (Table 4), we found evidence of positive selection by PARRIS (p < 0.04), and one positively selected site for BGMV (codon 54) and two sites for EuYMV (codons 93 and 110) were detected by FEL. For AC4 of MaYSV, which also exhibited dN/dS > 1, only one positively selected site (codon 39) was detected by FEL, without evidence by PARRIS (Table 4). In addition, for BIYSV TrAP we found evidence of positive selection by PARRIS (p < 0.043), and one (codon 90) and 10 (codons 34, 79, 85, 88, 89, 90, 93, 95, 97, 128) positively selected sites were detected by FEL and REL, respectively. Other genes showed sites with weaker evidence of positive selection, detected by only one or two methods (Table 4).

### DISCUSSION

Several key events marked the evolution of geminiviruses, culminating with the emergence of segmented (bipartite) genomes in viruses classified in the genus Begomovirus (Rojas et al., 2005). Expansion of the genome through the capture of a new segment by an ancestral nonsegmented virus, together with the loss of the V2 gene, are the main events related to the emergence of NW bipartite begomoviruses. Although NW begomoviruses are predominantly bipartite, very few studies have addressed evolutionary aspects of the complete (DNA-A and DNA-B) genome (Briddon et al., 2010; Rodelo-Urrego et al., 2015).

Our results, based on analysis of five well-sampled NW bipartite begomoviruses, indicate that DNA-A and DNA-B segments respond differently to evolutionary processes, with the DNA-B being more permissive to variation and more prone to recombination than the DNA-A. Although the DNA-B is usually more variable than DNA-A.
at the population level, this is not an absolute rule, as MaYSV DNA-A is more variable than its DNA-B (basically because it has a recombinant fragment encompassing the Rep gene which contains a large number of synonymous mutations (Lima et al., 2013). All viruses have similar (rather narrow) host ranges. BGMV and MaYSV were sampled from both cultivated and noncultivated hosts. A previous study (based on the DNA-A only) showed that they have contrasting genetic structures regardless of host (Ramos-Sobrinho et al., 2014),
and the DNA-B analysis confirms that result. Overall, our results do not indicate a correlation between host range and population structure or ecology.

It has been proposed that the DNA-B can tolerate greater variation since it does not contain overlapping genes, while the DNA-A encodes four overlapping genes (Rep/AC4 and TrAP/REn) (Briddon et al., 2010). With overlapping genes, a higher proportion of mutations will cause amino acid changes, which could lead to fitness trade-offs and negative selection. However, the fact that the AC4 gene (which is entirely within the Rep gene) is often under positive selection seems to undermine this hypothesis. Another explanation would be that the proteins encoded by the DNA-A are involved in several cis and trans interactions which could be negatively affected by changes in their amino acid sequences, placing this segment under stricter evolutionary constraints. However, using an interspecific data set, Ho et al. (2014) demonstrated that while DNA-B coding regions of OW begomoviruses are indeed more variable than DNA-A coding regions, this pattern was not consistently observed for NW begomoviruses. In agreement with this observation, our results indicate that patterns of variation for NW begomovirus DNA segments coding regions are species-specific. For example, there is no difference in variation among ToSRV CP, Rep, MP and NSP genes, although the DNA-B as a whole is more variable than the DNA-A. For MaYSV not only is the DNA-A more variable than the DNA-B, but Rep is the most variable gene followed by AC4, with CP more variable than MP and no difference between CP and NSP. In contrast, the BiYSV MP and NSP genes are the most variable genes. We also show that coding regions of the DNA-B evolve under at least the same selection pressure than DNA-A coding regions. This pattern might reflect the specialized movement function, encoded exclusively by the DNA-B in NW begomoviruses, impairing its ability to tolerate nonsynonymous substitutions (Gardiner et al., 1988; Jeffrey et al., 1996). Conversely, bipartite OW begomoviruses, whose movement functions may partially be provided by DNA-A, can allow the DNA-B to evolve in a more relaxed fashion, becoming more permissive to variation (Padidam et al., 1996; Roshan et al., 2018; Rothenstein et al., 2007). Moreover, NW DNA-B segments accommodate additional functions besides movement, such as suppression of plant defence mechanisms (Zorzatto et al., 2015), which may impose additional constraints.

In contrast to the ssDNA Faba bean necrotic stunt virus (FBNSV, family Nanoviridae), where there is no difference in variability between coding and non-coding regions (Grigoras et al., 2010), a clear difference was observed here where intergenic regions (IRs) are more relaxed to variation than coding regions. Thus, the DNA-B may support a greater accumulation of variation compared to the DNA-A, since it has two IRs comprising approximately one third of its length, in contrast to the DNA-A IR encompassing only one eighth of its length.

Asymmetric accumulation among the different genomic segments in multipartite viruses seems to be a common trait shared by RNA and DNA viruses infecting animals and plants (Hu et al., 2016; Lozano et al., 2009; Sicard et al., 2013; Wu et al., 2017; Yeh et al., 2000). During the infection cycle, each segment reaches a stable relative frequency, referred to as the "setpoint genome formula" (Sicard et al., 2013; Wu et al., 2017). Although very little is known about the biological and evolutionary meaning of this mechanism, it was proposed that it may be related to regulation of gene expression by controlling the copy number of each segment, and more interestingly, that it can directly modulate the evolutionary rates of the different segments (Gallet et al., 2018; Sicard et al., 2013). Segments with higher accumulation would have a greater mutation load, and consequently would evolve faster than other segments present at a lower frequency (Gutierrez & Zwart, 2018; Lozano et al., 2009). Although no study has previously addressed whether the genome formula applies to bipartite begomoviruses, indirect evidence obtained using high-throughput sequencing from unamplified libraries suggest that the DNA-B can accumulate at twice the level of the DNA-A for both NW and OW begomoviruses (Chen et al., 2019; Pinto et al., 2021). Considering that the two segments are exposed to the same basal rate of mutation, a higher accumulation of the DNA-B could contribute to its higher variation at the populational level.

Viral populations are subject to severe bottlenecks, especially during systemic infection of the host and horizontal transmission by vectors (Betancourt et al., 2008; Gallet et al., 2018; Monsion et al., 2008; Moury et al., 2007). After severe bottlenecks, a large reduction in the effective population size (Ne) can favour the action of genetic drift, leading to a dramatic loss of genetic diversity, as demonstrated for monopartite (Monsion et al., 2008; Moury et al., 2007) and tripartite (Ali et al., 2006) ssRNA viruses. The aphid-transmitted FBNSV (an octopartite ssDNA virus) also undergoes severe bottlenecks during vector transmission (Gallet et al., 2018). Interestingly, the size of the bottleneck differs for the two FBNSV segments studied, and a direct correlation between Ne and the relative frequency of each segment was observed. These results suggest that genetic drift may differentially affect each genomic segment according to their relative frequencies, driving each one to distinct patterns of variation and evolution. It is reasonable to assume that differences in variability observed between begomovirus DNA-A and DNA-B segments could at least in part be influenced by genetic drift, due to differences in the strength of the bottlenecks experienced by each segment especially during transmission by B. tabaci.

Recombination is considered one of the main forces driving begomovirus evolution (Lefevre et al., 2007, 2009; Lima et al., 2013, 2017; Rocha et al., 2013; Saleem et al., 2016). Our results showed a much greater propensity of the DNA-B to recombine compared to the DNA-A, with a higher number of unique events for DNA-B segments compared to DNA-A segments in all data sets. Martin et al. (2005) demonstrated experimentally that tolerance to recombination of a given region of the genome is correlated to the degree of divergence between the segments exchanged and the number of inter- and intragenomic interactions established. Thus, the greater propensity of the DNA-B to recombine may be explained by its lowest organizational complexity and the lowest number of inter- and intragenomic interactions of its encoded proteins (Hanley-Bowdoin et al., 2013). Interestingly, most recombination events involving the
DNA-B had breakpoints in the LIR, but none mapped to the v-ori, a known hot spot for recombination in OW begomoviruses (Lefeuvre et al., 2009). Although breakpoint mapping has a margin of error, this suggests that NW and OW begomoviruses may have distinct recombination profiles, which deserves further study.

Rodelo-Urrego et al. (2015) observed no difference in the frequency of recombinant sequences between DNA-A and DNA-B segments of Pepper golden mosaic virus (PepGMV) and Pepper huasteco yellow vein virus (PHYVV), two bipartite begomoviruses infecting chiltepin (Capsicum annuum var. glabriusculum) plants in Mexico. We found equivalent recombination frequencies for the DNA-A and DNA-B of MaYSV; however, the number of unique events was much higher for the DNA-B compared to the DNA-A. For BIYSV, both the number of unique events and the frequency of recombinant sequences were higher for the DNA-B, with 41% of the isolates carrying the same recombination event. Thus, the frequency of recombinant sequences, more than the propensity to recombination, may reflect the amplification of a few successful events in the population which are positively selected. Considering the five viruses, our results indicate a higher propensity for recombination in the DNA-B compared to the DNA-A, but also that selective advantage of specific recombination events may vary across DNA segments and viruses. Although less frequent, intersegment homologous recombination has been reported in multipartite viruses (Grigoras et al., 2014; Hu et al., 2007; Hughes, 2004; Saunders et al., 2002). Interestingly, Gregorio-Jorge et al. (2010) reported the presence of small fragments of Rep gene-derived sequences in the DNA-B intergenic region of several NW begomoviruses. This suggests that nonhomologous recombination among heterologous segments may also occur. We detected small truncated Rep (tRep) ORFs in the complementary-sense strand of the LIR-B of several MaYSV isolates. The presence of these tRep ORFs is a strong indication of intersegment recombination. Although nonhomologous recombination may be deleterious due to the breakage of important coding or regulatory regions, these results suggest that it can occur in a less congested segment such as the DNA-B.

While the cophylogenetic analysis suggests a global congruence between DNA-A and DNA-B trees across all data sets, congruence levels were variable among them, consistent with the role of reassortment in the evolution of multipartite viruses (Ohshima et al., 2016; Varsani et al., 2018). We were not able to estimate and compare evolutionary rates between DNA-A and DNA-B data sets, since the sequences were sampled on too short a time span. However, the phylogenetic concordance observed, which would indicate a co-evolution scenario, may be caused by biogeographic processes and limited gene flow among subpopulations, rather than similar evolutionary rates driven by mutual selection between the segments. Indeed, DNA-A and DNA-B phylogenies depict the formation of large clades according to geographical origin, providing evidence of population subdivision based on geography, as confirmed by DAPC analysis. Structural differences and the specialized functions of each segment would argue against the assumption of similar evolutionary rates, but nevertheless it needs to be tested. In spite of the selection analysis, which suggests that coding regions from each segment may evolve under the same selection pressure (which could lead to similar evolutionary rates), different segments in multipartite viruses may be differentially influenced by recombination and genetic drift (Gallet et al., 2018). That could overcome, to some extent, the homogenizing effect of selection and contribute to distinct rates. In addition, differences in the propensity for recombination between DNA-A and DNA-B may drastically affect their evolution, guiding each segment to distinct patterns of variation.

Previous studies, based only on the DNA-A analysis, have shown that begomoviruses populations segregate based on geographical location on a global as well as on a local scale (Mar et al., 2017; Prasanna et al., 2010; Rocha et al., 2013; Ramos-Sobrinho et al., 2014). Our results showed this to occur for the DNA-B as well. Although a direct comparison among different data sets should be treated with caution, due to differences in geographic scale and number of sequences sampled, our results demonstrate that geography-based subdivision is clear for both segments of BGMV, EuYMV and ToSRV, and to a lesser extent for the DNA-A of MaYSV. For BIYSV, a consistent clustering pattern was not verified by either phylogenetic or DAPC analyses, even analyzing isolates sampled in distant geographical regions. This result suggests an incipient pattern of differentiation, probably due a connection between the two sampled regions, allowing gene flow between them or a recent founder effect.

We were able to detect reassortment events across all data sets. However, all events detected were intraspecific and most exchanged segments were restricted among geographically and genetically close subpopulations, probably contributing little for the genetic makeup of the subpopulation.

Our results demonstrate that the DNA-A and DNA-B segments of NW begomoviruses, as well as different regions in a gene and genomic segment, display different evolutionary patterns, with significant differences in variation and recombination levels. Thus, while regulatory regions responsible for maintaining the integrity and functionality of the multipartite genome might be under strong selection pressure in different segments, most of the genome may evolve in a more relaxed fashion, driving distinct patterns of variation. This more relaxed evolution may be an advantage of multipartition, as each segment may behave as a single independent entity, exploring a larger portion of sequence space and allowing for faster adaptation to a constantly changing environment.

ACKNOWLEDGEMENTS
This work was funded by CAPES (Finance Code 001), CNPq (grant 409599/2016-6 to FMZ) and FAPEMIG (grant APQ-03444-16 to FMZ). The authors wish to thank Eduardo S.G. Mizubuti and Siobain Duffy for helpful discussions, and Fernando García-Arenal for critical review of the manuscript.

AUTHOR CONTRIBUTIONS
César A.D. Xavier, Márcio T. Godinho and F. Murilo Zerbini designed the research. César A.D. Xavier, Márcio T. Godinho, Talita B. Mar, Camila G. Ferro, Osvaldo F.L. Sande, Roberto Ramos-Sobrinho and...
Renato N. Nascimento performed the research. Márcio T. Godinho, José C. Silva, Iraldes Assunção, Gaus S.A. Lima and Alison T.M. Lima contributed new reagents or analytical tools. César A.D. Xavier, Márcio T. Godinho, Talita B. Mar, José C. Silva, Renato N. Nascimento, Alison T.M. Lima and F. Murilo Zerbini analysed the data. César A.D. Xavier and F. Murilo Zerbini wrote the manuscript.

DATA AVAILABILITY STATEMENT
All sequences obtained in this study have been deposited in GenBank. Accession numbers are listed in Table S1.

ORCID
F. Murilo Zerbini https://orcid.org/0000-0001-8617-0200

REFERENCES
Ali, A., Li, H., Schneider, W. L., Sherman, D. J., Gray, S., Smith, D., & F. Murilo Zerbini. ORCID
DATA AVAILABILITY STATEMENT
All sequences obtained in this study have been deposited in GenBank. Accession numbers are listed in Table S1.

ORCID
F. Murilo Zerbini https://orcid.org/0000-0001-8617-0200

REFERENCES
Ali, A., Li, H., Schneider, W. L., Sherman, D. J., Gray, S., Smith, D., & Roossinck, M. J. (2006). Analysis of genetic bottlenecks during horizontal transmission of Cucumber mosaic virus. *Journal of Virology*, 80, 8345–8350. https://doi.org/10.1128/JVI.00568-06.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic alignment search tool. *Journal of Molecular Biology*, 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Arguell-Astorga, G. R., Guevara-González, R. G., Herrera-Estrella, L. R., & Rivera-Bustamante, R. F. (1994). Geminivirus replication origins have a group-specific organization of interactive elements: a model for replication. *Virology*, 203, 90–100.

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews*, 35, 235–241. https://doi.org/10.1128/BR.35.3.235-241.1971.

Betancourt, M., Fereres, A., Fraile, A., & García-Arenal, F. (2008). Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. *Journal of Virology*, 82, 12416–12421. https://doi.org/10.1128/jvi.01542-08.

Briddon, R. W., Patil, B. L., Bagewadi, B., Nawaz-ul-Rehman, M. S., & Fauquet, C. M. (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evolutionary Biology*, 10, 97. https://doi.org/10.1186/1471-2148-10-97.

Chao, L. (1991). Levels of selection, evolution of sex in RNA viruses, and the origin of life. *Journal of Theoretical Biology*, 135, 229–246. https://doi.org/10.1016/S0022-5193(05)80424-2.

Chen, K., Khatabi, B., & Fondong, V. N. (2019). The AC4 protein of *African cassava mosaic viruses*. *Journal of General Virology*, 90, 1539–1547. https://doi.org/10.1099/jgv.0.009266-0.

Elena, S. F. (2016). Evolutionary transitions during RNA virus experimental evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20150441. https://doi.org/10.1098/rstb.2015.0441.

Fontes, E. P. B., Eagle, P. A., Sipe, P. S., Luckow, V. A., & Hanley-Bowdoin, L. (1994). Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *Journal of Biological Chemistry*, 269, 8459–8465. https://doi.org/10.1016/S0021-9258(17)37216-2.

Frailé, A., Alonso-Prados, J. L., Aranda, M. A., Bernal, J. J., Malpica, J. M., & García-Arenal, F. (1997). Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. *Journal of Virology*, 71, 934–940. https://doi.org/10.1128/JVI.71.2.934-940.1997.

Gale, M. J., Tan, S. L., & Katze, M. G. (2000). Translational control of viral gene expression in eukaryotes. *Microbiology and Molecular Biology Reviews*, 64, 239–280. https://doi.org/10.1128/MMBR.64.2.239-280.2000.

Gallet, R., Fabre, F., Thebaut, G., Sofonea, M. T., Sicard, A., Blanc, S., & Michalakis, Y. (2018). Small bottleneck size in a highly multipartite virus during a complete infection cycle. *Journal of Virology*, 92, e00139-00118. https://doi.org/10.1128/JVI.00139-18.

García-Andrés, S., Accotto, G. P., Navas-Castillo, J., & Moriones, E. (2007). Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. *Virology*, 359, 302–312. https://doi.org/10.1016/j.virology.2006.09.030.

García-Ariáza, J., Manrubia, S. C., Toja, M., Domingo, E., & Escarmís, C. (2004). Evolutionary transition toward defective RNAs that are infectious by complementation. *Journal of Virology*, 78, 11678–11685. https://doi.org/10.1128/JVI.78.21.11678-11685.2004.

Gardiner, W., Sunter, G., Brand, L., Elmer, J. S., Rogers, S. G., & Bisaro, D. M. (1988). Genetic analysis of tomato golden mosaic virus: The coat protein is not required for systemic spread or symptom development. *EMBO Journal*, 7, 899–904. https://doi.org/10.1002/j.1460-2075.1988.tb02894.x.

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. *Journal of Virology*, 81, 5902–5907. https://doi.org/10.1128/jvi.02431-06.

Geoghegan, J. L., Duchene, S., & Holmes, E. C. (2017). Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Path*, 13, e1006215. https://doi.org/10.1371/journal.ppat.1006215.

Gregorio-Jorge, J., Bernal-Alcocer, A., Bañuelos-Hernández, B., Alpuche-Solis, Á. G., Hernández-Zepeda, C., Moreno-Valenzuela, O., Frías-Treviño, G., & Argüello-Astorga, G. R. (2010). Analysis of a new strain of *Euphorbia mosaic virus* with distinct replication specificity unveils a lineage of begomoviruses with short Rep sequences in the DNA-B intergenic region. *Virology Journal*, 7, 275. https://doi.org/10.1186/1743-422x-7-275.

Grigoras, I., Grinz, A. D. C., Martin, D. P., Varsani, A., Romero, J., Mammadov, A. C., Huseynova, I. M., Aliyev, J. A., Kheyr-Pour, A., Huss, H., Ziebell, H., Timchenko, T., Vetten, H.-J., & Gronenborn, B. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. *Journal of General Virology*, 95, 1178–1191. https://doi.org/10.1099/vir.0.063115-0.

Grigoras, I., Timchenko, T., Grande-Perez, A., Katul, L., Vetten, H.-J., & Gronenborn, B. (2010). High variability and rapid evolution of a nanovirus. *Journal of Virology*, 84, 9105–9117. https://doi.org/10.1128/jvi.00607-10.

Gutiérrez, S., & Zwart, M. P. (2018). Population bottlenecks in multipartite viruses: first forays into the uncharted territory of genome-formula drift. *Current Opinion in Virology*, 33, 184–190. https://doi.org/10.1016/j.coviro.2018.09.001.

Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D., & Mansoor, S. (2013). Geminiviruses: Masters at redirecting and reprogramming plant
processes. Nature Reviews Microbiology, 11, 777–788. https://doi.org/10.1038/nrmicro3117.

Ho, E. S., Kuchie, J., & Duffy, S. (2014). Bioinformatic analysis reveals genome size reduction and the emergence of tyrosine phosphorylation site in the movement protein of New World bipartite begomoviruses. PLoS One, 9, e111957. https://doi.org/10.1371/journal.pone.0111957.

Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J., & Yeh, H. H. (2007). Reassortment and concerted evolution in Banana bunchy top virus genomes. Journal of Virology, 81, 1746–1761. https://doi.org/10.1128/JVI.01390-06.

Hu, Z., Zhang, X., Liu, W., Zhou, Q., Zhang, Q., Li, G., & Yao, Q. (2016). Genome segments accumulate with different frequencies in Bombyx mori bidensivirus. Journal of Basic Microbiology, 56, 1338–1343. https://doi.org/10.1002/jobm.201600120.

Hughes, A. L. (2004). Birth-and-death evolution of protein-coding regions and concerted evolution of non-coding regions in the multi-component genomes of nanoviruses. Molecular Phylogenetics and Evolution, 30, 287–294. https://doi.org/10.1016/S1055-7903(03)00189-1.

Hutchinson, M. C., Cagua, E. F., Balbuena, J. A., Stouffer, D. B., & Poisot, T. (2017). PACC: Implementing Procrustean approach to cophylogeny in R. Methods in Ecology and Evolution, 8, 932–940. https://doi.org/10.1111/2041-210X.12736.

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 phi29 DNA polymerase. Journal of Virological Methods, 116, 209–211. https://doi.org/10.1016/j.jviromet.2003.11.015.

Jeffrey, J. L., Pooma, W., & Petty, I. T. (1996). Genetic requirements for local and systemic movement of tomato golden mosaic virus in infected plants. Virology, 223, 208–218. https://doi.org/10.1006/viro.1996.0469.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics, 27, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics, 11, 94. https://doi.org/10.1186/1471-2156-11-94.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Koonin, E. V., Dolja, V. V., Krupovic, M., Varani, A., Wolf, Y. I., Yutin, N., Zerbini, F. M., & Koonin, E. V. (2002). Global organization and analysis of sequence data. Journal of General Virology, 83, 1343–1353. https://doi.org/10.1099/jgv.0.000784.

Koonin, E. V., Dolja, V. V., Krupovic, M., Varani, A., Wolf, Y. I., Yutin, N., Zerbini, F. M., & Koonin, E. V. (2002). Global organization and analysis of sequence data. Journal of General Virology, 83, 1343–1353. https://doi.org/10.1099/jgv.0.000784.

Lauber, C. L., Hamady, M., Knight, R., & Gordon, L. J. (2009). Diversity and rarity assessment for microorganisms using 16S rRNA gene pyrosequencing. Nature Methods, 6, 363–370. https://doi.org/10.1038/nmeth.1304.

Lefebvre, P., Leffler, V., & Moriones, E. (2015). Recombination as a motor of host switches and virus emergence: Geminiviruses as case studies. Current Opinion in Virology, 10, 14–19. https://doi.org/10.1016/j.coviro.2014.12.005.

Lima, A. T. M., Silva, J. C. F., Silva, F. N., Castillo-Urquiza, G. P., Silva, F. F., Seah, Y. M., Mizubuti, E. S. G., Duffy, S., & Zerbini, F. M. (2017). The diversification of begomovirus populations is predominantly driven by mutational dynamics. Virus Evolution, 3, vex005. https://doi.org/10.1093/ve/vex005.

Lima, A. T. M., Sobrinho, R. R., Gonzalez-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. Journal of General Virology, 94, 418–431. https://doi.org/10.1099/vir.0.047241-0.

Lister, R. M. (1966). Possible relationships of virus-specific products of tobacco rattle virus infections. Virology, 28, 350–353. https://doi.org/10.1016/S0042-6822(66)90161-9.

Lozano, G., Grande-Perez, A., & Navas-Castillo, J. (2009). Populations of genomic RNAs devoted to the replication or spread of a bipartite plant virus differ in genetic structure. Journal of Virology, 83, 12973–12983. https://doi.org/10.1128/JVI.00950-09.

Lucia-Sanz, A., & Manrubia, S. (2017). Multiparticle viruses: Adaptive trick or evolutionary treat? NPJ Systems Biology and Applications, 3, 34. https://doi.org/10.1038/s41540-017-0035-y.

Mar, T. B., Xavier, C. A. D., Lima, A. T. M., Nogueira, A. M., Silva, J. C. F., Ramos-Sobrinho, R., Lau, D., & Zerbini, F. M. (2017). Genetic variability and population structure of the New World begomovirus Euphorbia yellow mosaic virus. Journal of General Virology, 98, 1537–1551. https://doi.org/10.1099/jgv.0.000784.

Martin, D., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution, 1, vev003. https://doi.org/10.1093/ve/vev003.

Martin, D. P., van der Walt, E., Posada, D., & Rybicki, E. P. (2005). The evolutionary value of recombination is constrained by genome modularity. PLoS Genetics, 1, e51. https://doi.org/10.1371/journal.pgen.0010051.

McDonald, S. M., Nelson, M. I., Turner, P. E., & Patton, J. T. (2016). Reassortment in segmented RNA viruses: Mechanisms and outcomes. Nature Reviews Microbiology, 14, 448–460. https://doi.org/10.1038/nrmicro.2016.46.

Monci, F., Sanchez-Campoz, 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. https://doi.org/10.1006/viro.2002.1633.

Monsion, B., Fraissart, R., Michalakis, Y., & Blanc, S. (2008). Large bottleneck size in cauliflower mosaic virus populations during host plant colonization. PLoS Path, 4, e1000174. https://doi.org/10.1371/journal.ppat.1000174.

Moury, B., Fabre, F., & Senoussi, R. (2007). Estimation of the number of virus particles transmitted by an insect vector. Proceedings of the National Academy of Sciences, USA, 104, 17891–17896. https://doi.org/10.1073/pnas.0702739104.

Nee, S. (1987). The evolution of multicompartmental genomes in viruses. Journal of Molecular Evolution, 25, 277–281. https://doi.org/10.1007/BF02603110.

Nylander, J. A. A. (2004). MrModeltest v2. PLoS Path, 1, e7. https://doi.org/10.1371/journal.ppat.1000005.
Turner, P. E. (2003). Searching for the advantages of virus sex. * Origins of Life and Evolution of the Biosphere*, 33, 95–108.

Varsani, A., Lefeuvre, P., Roumagnac, P., & Martin, D. (2018). Notes on recombination and reassortment in multipartite/segmented viruses. *Current Opinion in Virology*, 33, 156–166. https://doi.org/10.1016/j.coviro.2018.08.013.

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, 28, 127–181. https://doi.org/10.1016/j.femsre.2003.08.001.

Wu, B., Zwart, M. P., Sánchez-Navarro, J. A., & Elena, S. F. (2017). Within-host evolution of segments ratio for the tripartite genome of alfalfa mosaic virus. *Scientific Reports*, 7, 5004. https://doi.org/10.1038/s41598-017-05335-8.

Yeh, H. H., Tian, T., Rubio, L., Crawford, B., & Falk, B. W. (2000). Asynchronous accumulation of lettuce infectious yellows virus RNAs 1 and 2 and identification of an RNA 1 trans enhancer of RNA 2 accumulation. *Journal of Virology*, 74, 5762–5768. https://doi.org/10.1128/JVI.74.13.5762-5768.2000.

Zerbini, F. M., Briddon, R. W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., Rivera-Bustamante, R., Roumagnac, P., Varsani, A., & ICTV Consortium (2017). ICTV Virus Taxonomy Profile: Geminiviridae. *Journal of General Virology*, 98, 131–133. https://doi.org/10.1099/jgv.0.000738.

Zorzatto, C., Machado, J. P. B., Lopes, K. V. G., Nascimento, K. J. T., Pereira, W. A., Brustoloni, O. J. B., Reis, P. A. B., Calil, I. P., Deguchi, M., Sachetto-Martins, G., Gouveia, B. C., Loriato, V. A. P., Silva, M. A. C., Silva, F. F., Santos, A. A., Chory, J., & Fontes, E. P. B. (2015). NIK1-mediated translation suppression functions as a plant antiviral immunity mechanism. *Nature*, 520, 679–682. https://doi.org/10.1038/nature14171.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

---

**How to cite this article:** Xavier, C. A. D., Godinho, M. T., Mar, T. B., Ferro, C. G., Sande, O. F. L., Silva, J. C., Ramos-Sobrinho, R., Nascimento, R. N., Assunção, I., Lima, G. S. A., Lima, A. T. M., & Murilo Zerbini, F. (2021). Evolutionary dynamics of bipartite begomoviruses revealed by complete genome analysis. *Molecular Ecology*, 30, 3747–3767. https://doi.org/10.1111/mec.15997