CÉSAR AUGUSTO DINIZ XAVIER

BEGOMOVIRUSES: VARIABILITY, EVOLUTION, AND INTERACTIONS WITH HOSTS AND VECTORS

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitopatologia, para obtenção do título de Doctor Scientiae.

Orientador: Francisco Murilo Zerbini Júnior

VIÇOSA - MINAS GERAIS
2019

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
X3b
2019

Xavier, César Augusto Diniz, 1990-
Begomoviruses : variability, evolution, and interactions with hosts and vectors / César Augusto Diniz Xavier. – Viçosa, MG, 2019.
206 f. : il. ; 29 cm.

Orientador: Francisco Murilo Zerbini Junior.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Begomovírus. 2. DNA de forma B. 3. Deaminação. 4. Diversidade genética. 5. Mosca-branca. I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-Graduação em Fitopatologia. II. Título.

CDD 22. ed. 579.28

CÉSAR AUGUSTO DINIZ XAVIER

BEGOMOVIRUSES: VARIABILITY, EVOLUTION, AND INTERACTIONS WITH HOSTS AND VECTORS

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitopatologia, para obtenção do título de Doctor Scientiae.

APROVADA: 13 de setembro de 2019.

_____

César Augusto Diniz Xavier
Autor

_____

Francisco Murilo Zerbini Júnior
Orientador

Em nome da minha mãe Maria de Lourdes e meu pai Heleno, dedico.

AGRADECIMENTOS

A Deus por ter-me concedido o dom da vida, por sempre guiar meus passos e iluminar meu caminho, especialmente em momentos quando tudo parecia perdido;

Aos meus pais, Heleno e Maria de Lourdes, pela confiança, pelo apoio incondicional nos momentos de alegria e tristezas, pelo incentivo, amor e dedicação em todos os momentos da minha vida;

A minha amada Kellyne por ter abraçado este sonho comigo, pelo amor incondicional, incentivo, paciência, carinho e confiança;

Aos meus irmãos e sobrinhos, por sempre acreditarem em mim e estarem presentes ao longo desta caminhada;

A Universidade Federal de Viçosa (UFV), pela oportunidade de realização deste curso;

Ao Professor Francisco Murilo Zerbini, pelo aprendizado transmitido, orientação e dedicação durante a realização deste trabalho;

A todos os amigos do Laboratório de Virologia Vegetal Molecular, pela amizade e pelo agradável ambiente de trabalho ao longo destes 10 anos de laboratório. Em especial aos amigos André, Angélica e Larissa.

Ao ex-funcionário Joaquim e a funcionária Patrícia, pelo apoio para a realização deste trabalho e à amizade;

Aos amigos do Laboratório de Microbiologia Industrial pela amizade;

Agradeço a professora Linda Hanlew-Bowdoin e ao professor Trino pelo apoio, ensinamentos transmitidos e pela amizade. Agradeço a todos os membros do laboratório da Prof. Linda pela amizade, pela paciência e por todas as discussões.

Agradeço a todos os professores e amigos do Departamento de Fitopatologia;

A todos que contribuíram direta ou indiretamente para a realização deste trabalho, muito obrigado.

ABSTRACT

XAVIER, César Augusto Diniz, D.Sc., Universidade Federal de Viçosa, August, 2019. **Begomoviruses: variability, evolution, and interactions with hosts and vectors.** Adviser: Francisco Murilo Zerbini Júnior.

Begomoviruses constitute an important group of plant pathogens being widespread in the world. Although New World begomoviruses are predominantly bipartite, the bipartite nature of these viruses has been little explored in terms of their evolutionary significance. We performed a parallel evolutionary analysis of the DNA-A and DNA-B components of New World begomoviruses from cultivated and non-cultivated hosts. The two components respond differentially to evolutionary process, with the DNA-B being more variable and more prone to recombination than the DNA-A. Our results indicate an interplay between reassortment and recombination acting at different levels across distinct subpopulations and genomic components. This more relaxed evolution may be an advantage of multipartition, as each component may behave as a single almost independent evolutionary entity, allowing a faster adaptation to an ever-changing environment. Furthermore, begomovirus population present a high level of genetic variability. Interestingly, a mutational bias has been detected in begomovirus genomes, favoring C to T transitions, suggesting that cytidine deamination may act on these viruses. We investigated a possible role of A. thaliana cytidine deaminase (AtCDA and AtCDAL4) as components of plant antiviral immunity. We demonstrate distinct functional roles for AtCDAs upon begomovirus infection. While AtCDA has a proviral effect during Cabbage leaf curl virus infection, AtCDAL4 has antiviral activity, constituting a putative branch of antiviral immune response in plants. However, unlike in animals, our results suggest that AtCDAs affect viral infection independently of deaminase activity. We also studied factors underlying leading to emergence and the current patterns of prevalence of begomoviruses in tomato by quantifying the replicative and transmission fitness of two begomoviruses infecting tomato in Brazil, Tomato severe rugose virus (ToSRV) and Tomato yellow spot virus (ToYSV). Interestingly, despite ToYSV is not widespread in the field, it presented similar replicative fitness compared to ToSRV, a widespread begomovirus, in tomato. These results suggest that adaptation to the host does not explain the prevalence of ToSRV over ToYSV, as both viruses presented similar adaptation levels as evaluated here. However, while ToSRV was transmitted by B. tabaci MEAM1 and MED, ToYSV was not transmitted by them, reason by which this virus probably is not spread in the field. Finally, it has been hypothesized that the absence of a competent vector that efficiently colonizes cassava is the reason why begomoviruses have not emerged in South America. To test this hypothesis, we performed a country-wide whitefly

diversity study in cassava in Brazil. A high degree of species richness was observed, with five previously described species and two putative new ones. The most prevalent species were Tetraleurodes acaciae and Bemisia tuberculata, representing over 75% of the analyzed individuals. Although we detected, for the first time, the presence of whitefly species of the Bemisia tabaci complex colonizing cassava in Brazil, they were not prevalent. Our results indicate an ongoing adaptation process of B. tabaci Middle East-Asia Minor 1 to cassava, increasing the likelihood of begomovirus emergence in this crop in the future. The results obtained here, expand our knowledge about the wide spectrum of interactions among begomoviruses-hosts-vectors.

Keywords: DNA-B. Base deamination. Whitefly diversity.

RESUMO

XAVIER, César Augusto Diniz, Universidade Federal de Viçosa, agosto de 2019. **Begomovírus: variabilidade, evolução e interações com hospedeiros e vetores**. Orientador: Francisco Murilo Zerbini Júnior.

Begomovírus é um grupo importante de patógenos de plantas e apresentam uma ampla distribuição. Embora os begomovírus na América são predominantemente bipartidos, a natureza bipartida desses vírus tem sido pouco explorada em termos do significado evolutivo. Neste trabalho, nós realizamos uma análise evolutiva dos componentes DNA-A e DNA-B de begomovírus que infectam hospedeiros cultivados e não cultivados. Os resultados obtidos aqui indicam que os dois componentes respondem diferentemente aos processos evolutivos, com o DNA-B sendo mais variável e mais propenso à recombinação comparado ao DNA-A. Nossos resultados indicam que a atuação de recombinação e pseudo-recombição em diferentes níveis entre subpopulações distintas e componentes genômicos. Possivelmente, essa evolução mais relaxada pode ser uma vantagem da multipartição, pois cada componente pode comportar como uma entidade evolutiva quase independente, permitindo uma adaptação mais rápida a um ambiente em constante mudança. Além disso, populações de begomovírus apresentam um alto nível de variabilidade genética. Curiosamente, um viés mutacional foi detectado nos genomas de begomovírus, favorecendo as transições de C para T, sugerindo que a deaminação de citidina pode atuar sobre esses vírus. Nós investigamos o possível papel de citidinas desaminases de A. thaliana (AtCDA e AtCDAL4) como componentes da imunidade antiviral de plantas. Nós demonstramos papéis funcionais distintos para as AtCDAs durante a infecção por begomovírus. Enquanto AtCDA teve um efeito positivo durante a infecção de Cabbage leaf curl virus, AtCDAL4 teve uma atividade antiviral, sugerindo ser parte da resposta imune antiviral em plantas. No entanto, diferentemente dos animais, nossos resultados sugerem que AtCDAs afetam a infecção viral independentemente da atividade de deaminase. Nós também estudamos os fatores subjacentes à emergência e aos padrões atuais de prevalência de begomovírus em tomateiro, quantificando a adaptabilidade replicativa e de transmissão de dois begomovírus que infectam tomateiro no Brasil, Tomato severe rugose virus (ToSRV) e o Tomato yellow spot virus (ToYSV). Curiosamente, apesar do ToYSV não estar amplamente disseminado no campo, ele apresentou adaptabilidade replicativa semelhante ao ToSRV, um begomovírus amplamente disperso. Esses resultados sugerem que a adaptação ao hospedeiro não explica a prevalência de ToSRV sobre ToYSV, pois ambos os vírus apresentaram níveis de adaptação semelhantes. No entanto, enquanto o ToSRV foi transmitido por B. tabaci MEAM1 e MED, o ToYSV não foi transmitido por ambas as espécies, razão pela qual esse vírus provavelmente não está

disseminado no campo. Finalmente, nós testamos a hipótese de que a ausência de um vetor competente que colonize mandioca eficientemente é a razão pela qual os begomovírus não emergiram na América do Sul. Para testar esta hipótese, nós realizamos um estudo de diversidade de mosca branca em todo o país em mandioca. Interessantemente, uma alta riqueza de espécies foi detectada, com cinco espécies descritas anteriormente e duas supostas novas espécies. As espécies mais prevalentes foram Tetraleurodes acaciae e Bemisia tuberculata, representando mais de 75% dos indivíduos analisados. Embora tenhamos detectado, pela primeira vez, a presença de espécies de mosca branca do complexo Bemisia tabaci colonizando mandioca no Brasil, elas não foram prevalentes. Nossos resultados indicam um processo de adaptação incipiente de B. tabaci Middle East-Asia Minor 1 à mandioca, aumentando a probabilidade de surgimento de begomovírus nesta cultura no futuro. Os resultados aqui obtidos, expandem nosso conhecimento sobre o amplo espectro de interações entre begomovírus, hospedeiros e vetores.

Palavras-chave: DNA-B. Deaminação de base. Diversidade de mosca-branca.

SUMÁRIO

GENERAL INTRODUCTION

The family Geminiviridae is composed of viruses with one or two circular, single-stranded DNA (ssDNA) components encapsidated in twinned icosahedral particles. The family consists of the genera Becurtovirus, Begomovirus, Capulavirus, Curtovirus, Eragrovirus, Grablovirus, Mastrevirus, Topocuvirus, and Turncurtovirus, defined based on the vector, host range, genomic organization, and phylogenetic relationships.

The genus Begomovirus (family Geminiviridae) includes viral species with one or two genomic components which are transmitted by whiteflies of the Bemisia tabaci cryptic species complex. Begomoviruses constitute an important group of plant pathogens responsible for severe diseases in several crops of economic importance, especially in tropical and subtropical regions. The main examples include cassava mosaic disease in Africa, caused by a begomovirus complex composed of nine species, and tomato yellow leaf curl disease caused by Tomato yellow leaf curl virus (TYLCV) and Tomato yellow leaf curl Sardinia virus (TYLCSV), a major constraint on tomato production in the Middle East, the Mediterranean and North and Central America. The main factors that have contributed to the emergence and spread of diseases caused by begomoviruses are the increased spread of polyphagous B. tabaci populations and the emergence of more aggressive viral variants, the latter attributed to the ability of these viruses to evolve rapidly.

According to phylogenetic relationships and genome architecture, begomoviruses are divided into two major groups, named Old World (OW - Europe, Asia and Africa) and New World (NW - the Americas). OW begomoviruses can be non-segmented or bipartite and are normally associated with satellite DNA molecules. Begomoviruses in the NW are mostly bipartite, with a few exceptions. The two genomic components of bipartite begomoviruses are referred to as DNA-A and DNA-B. The DNA-A of NW begomoviruses encodes five genes: CP (capsid protein) in the virion-sense strand, Rep (replication-associated protein), TrAP (trans-activating protein), REn (replication enhancer protein) and ac4 in the complementary-sense strand. The DNA-B encodes two genes, NSP (nuclear shuttle protein) in the virion-sense strand and MP (movement protein) in the complementary-sense strand. The two components share very low sequence identity, except for a small region of approximately 200 nucleotides known as the common region (CR). The CR includes the replication origin (ori), a conserved stem loop structure with the invariable nonanucleotide TAATATT//AC, and conserved repeat sequences (iterons) that are specifically recognized by Rep. The CR is important for maintaining the

integrity of the bipartite genome, allowing both components to be replicated, since Rep proteins show high specificity for their cognate ori.

In Brazil, the most important diseases caused by begomoviruses are bean golden mosaic caused by Bean golden mosaic virus (BGMV) and rugose mosaic and leaf curl in tomato caused by a complex of begomoviruses. Currently, eighteen species have been described infecting tomato plants in Brazil, however, only a few species are widely disseminated in the field. There are reports of begomoviruses infecting other crops, such as okra, pepper, sweetpotato, passionfruit and cotton, but epidemics and economic losses have not been reported. In addition, a large number of begomoviruses infect non-cultivated plants, especially those belonging to the families Fabaceae, Malvaceae and Solanaceae.

Although BGMV has been reported in Brazil since the 1960's , the emergence of begomoviruses infecting tomatoes and other crops followed the introduction of B. tabaci Middle East-Asia Minor 1 (MEAM1; previously known as B. tabaci biotype B) in the mid-1990's. B. tabaci MEAM1, unlike the indigenous B. tabaci New World (NW; previously known as biotype A), has a larger host range and readily colonizes tomato and other solanaceous crops. The introduction of B. tabaci MEAM1 in Brazil facilitated the transfer of indigenous begomoviruses from non-cultivated plants to economically important crops, something that has been surveyed extensively. In addition, it has been demonstrated that Brazilian begomovirus populations have a rapid evolutionary dynamic, driven primarily by mutation and recombination.

Begomovirus populations have a high degree of genetic variability, comparable to those of ssRNA viruses. High mutation rates coupled with the propensity to recombine are key factors that contribute to their high genetic variability, allowing begomoviruses to increase their adaptive potential to new hosts and new environmental conditions. In addition, the exchange of genomic components (pseudo-recombination or reassortment) may also occur between bipartite viruses.

Studies carried out by our research group have focused on the evolutionary dynamics of begomovirus populations that infect cultivated and non-cultivated plants. However, these studies have analyzed only the DNA-A component, which accounts for 50% of the bipartite begomovirus genome. In addition, most of the research on the evolutionary genetics of viral populations has focused on the theoretical understanding of variability by quantifying the contribution of individual evolutionary forces, with no studies exploring the molecular mechanism responsible for the high and biased mutation rates. Finally, very few studies have

been carried out in Brazil to understand the factors guiding the predominance and emergence of begomoviruses in cultivated plants.

The objectives of this work were: i) To conduct a parallel evolutionary analysis of the DNA-A and DNA-B components of NW begomoviruses from cultivated and non-cultivated hosts; ii) To verify the possible role of A. thaliana cytidine deaminases (AtCDAs) as a component of plant antiviral immunity and to study their contribution to the high evolutionary rates displayed by begomoviruses; iii) To quantify the replicative fitness of Tomato severe rugose virus (ToSRV) and Tomato yellow spot virus (ToYSV) in tomato and in non-cultivated with which each virus has been frequently reported, and to quantify the transmission fitness by B. tabaci MEAM1 and B. tabaci MED; iv) To evaluate whitefly diversity in cassava across Brazil.

CHAPTER 1

EVOLUTIONARY DYNAMICS OF BIPARTITE BEGOMOVIRUSES IN THE NEW WORLD REVEALED BY COMPLETE GENOME ANALYSIS

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., Zerbini, F.M. Evolutionary dynamics of bipartite begomoviruses in the New World revealed by complete genome analysis. Molecular Ecology (in preparation).

Evolutionary dynamics of bipartite begomoviruses in the New World revealed by complete genome analysis

César A.D. Xavier[1,2], Márcio T. Godinho[1], Talita B. Mar[1#], Camila G. Ferro[1], Osvaldo F.L. Sande[1,2], José C. Silva[2], Roberto Ramos-Sobrinho[1&], Renato N. Nascimento[3], Iraildes Assunção[3], Gaus S.A. Lima[3], Alison T.M. Lima[4], F.M. Zerbini[1,2]*

[1]Dep. de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa (UFV), Viçosa, MG 36570-900, Brazil

[2]National Research Institute for Plant-Pest Interactions, UFV

[3]Centro de Ciências Agrárias/Fitossanidade, Universidade Federal de Alagoas, Rio Largo, AL 57100-000, Brazil

[4]Instituto de Ciências Agrárias, Universidade Federal de Uberlândia, Uberlândia, MG 38400-902, Brazil

[#]Present address:

[&]Present address: School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA

*Corresponding author:

Phone: (+55-31) 3612-2423

E-mail: zerbini@ufv.br

Abstract

Several key evolutionary events marked the evolution of geminiviruses, culminating with the emergence of 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. We performed a parallel evolutionary analysis of the DNA-A and DNA-B components 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 analyzed. 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 components, differences in the genetic structure between DNA-A and DNA-B were also observed, with cognate components 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 components.

Introduction

Although structurally simple, viruses possess complex evolutionary histories (Dolja & Koonin, 2018; Koonin, Dolja, & Krupovic, 2015; Shi et al., 2018; Solé, 2016; Wolf et al., 2018). 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, Tan, & Katze, 2000). This diversity allows viral populations to be dynamic, with a high adaptive capacity, infecting hosts in the three domains of life (Koonin et al., 2015; 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 (Sicard, Michalakis, Gutierrez, & Blanc, 2016). The biological and evolutionary significance of genome segmentation remains unclear (Iranzo & Manrubia, 2012; Lucia-Sanz, Aguirre, & Manrubia, 2018; Michalakis & Blanc, 2018; Pressing & Reanney, 1984).

A special case of genome segmentation are viruses with multipartite genomes, which have their genome segments packed into separate particles. The existence of viruses with multipartite genomes presents further challenges to the biological and evolutionary purpose of genome segmentation (Lucia-Sanz & Manrubia, 2017). While each segment reaches complete independence (since there is no physical connection between them), they need to act together to complete the infectious cycle. Therefore, multipartite genomes can suffer a high adaptative cost, especially during intercellular movement and transmission between hosts, 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 (MOI) (Garcia-Arriaza et al., 2004; Gutierrez & Zwart, 2018; Lucia-Sanz & Manrubia, 2017; Sanchez-Navarro, Zwart, & Elena, 2013; Sicard et al., 2013). Alternatively, it has been suggested that rapid replicative kinetics associated with higher fidelity of replication of smaller genomic segments are factors that 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 components is a factor that can at least in part contribute to the maintenance of multipartite genomes.

A number of studies have been performed aiming to understand the evolutionary dynamics of the different components of multipartite viruses at the population level (Briddon et al., 2010; Grigoras et al., 2014; Hu et al., 2008; Lozano, Grande-Perez, & Navas-Castillo, 2009; Roossinck, 2002). Given the trade-offs between functional complementation and

independence among the distinct components in multipartite viruses, it would be expected that the different components would be in an intimate process of co-evolution, with similar evolutionary histories. However, the above-mentioned studies indicated that the different components of multipartite genomes may undergo distinct selection pressures, and thus experience different evolutionary histories (Briddon et al., 2010; Lozano et al., 2009; Roossinck, 2002). This more relaxed evolution of multipartite genomes may be one more reason to explain their maintenance.

The genus Begomovirus (family Geminiviridae) is comprised of viral species with one or two genomic components of circular, single-strand DNA (ssDNA) of about 2,700 nucleotides, encapsidated in geminate icosahedral particles and transmitted to dicotyledonous plants by whiteflies of the Bemisia tabaci cryptic species complex (Zerbini et al., 2017). Begomoviruses constitute an important group of plant pathogens responsible for severe diseases in several crops and ornamentals plants of economic importance worldwide (Rojas et al., 2018). In addition to their economic importance, two main characteristics make begomoviruses an attractive model to study the evolutionary dynamics of viral populations. First, begomoviruses evolve at rates which are comparable to those of ssRNA viruses, with estimates of substitution rates in the order of 10^{-3}-10^{-4} substitutions/site/year (Duffy & Holmes, 2008; Duffy & Holmes, 2009; Ge, Zhang, Zhou, & Li, 2007). Second, begomoviruses include viruses with non-segmented as well as 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 components.

The two genomic components of bipartite begomoviruses are referred to as DNA-A and DNA-B. The two components share no significant sequence identity, except for an intergenic region (IR) of approximately 200 nucleotides. The IR includes the replication origin (ori), a conserved stem-loop structure with the invariable nonanucleotide TAATATT//AC, and conserved repeat sequences (iterons) that are specifically recognized by the viral replication-associated protein, Rep (Arguello-Astorga, Guevara-González, Herrera-Estrella, & Rivera-Bustamante, 1994; Fontes et al., 1994; Lazarowtiz, Wu, Rogers, & Elmer, 1992). The IR is important for maintaining the integrity of the bipartite genome, allowing both components to be replicated by Rep, since Rep proteins show high specificity for their cognate ori (Lazarowtiz et al., 1992).

Although the mechanism by which segmented and multipartite genomes have emerged is unclear, evidence suggests they may originate by fragmentation of the genome of non-segmented progenitors, with defective segments becoming infectious by complementation

(Garcia-Arriaza et al., 2004). Specifically for begomoviruses, Briddon et al. (2010) suggested that the DNA-B could have originated from a satellite molecule captured by the monopartite progenitor of all begomoviruses. Possibly, this association provided greater adaptability to the monopartite progenitor, and consequently was maintained over the evolutionary process.

Reassortment (the exchange of genomic components between strains or related species, often called pseudorecombination by geminivirologists) is an important factor that contributes to the generation of variability in viral populations (Hu et al., 2008; Vijaykrishna, Mukerji, & Smith, 2015). Reassortment is common among bipartite begomoviruses (Andrade et al., 2006; Briddon et al., 2010; Faria et al., 1994; Garrido-Ramirez, Sudarshana, & Gilbertson, 2000; Gilbertson et al., 1993; Hou & Gilbertson, 1996; Silva et al., 2014). It has been proposed that reassortment 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).

Understanding the processes driving the patterns of population genetic differentiation and genetic variability of viruses has been a long-standing focus in the area of evolutionary virology. A number of studies have been conducted to determine and quantify the contribution of the mechanisms dictating the evolutionary dynamics of begomovirus populations (Duffy & Holmes, 2008; Duffy & Holmes, 2009; García-Andrés, Accotto, Navas-Castillo, & Moriones, 2007; Ge et al., 2007; Lefeuvre et al., 2010; Lima et al., 2017; Lima et al., 2013; Mar et al., 2017; Rocha et al., 2013; Sanz et al., 1999). However, these studies were based on analysis of non-segmented viruses or of the DNA-A component of bipartite viruses. Considering the important role played by the DNA-B in the infection process, evolutionary studies that consider this component 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 component of NW begomoviruses from cultivated and non-cultivated hosts. The two components respond differentially to evolutionary processes, with the DNA-B being more variable and more prone to recombination than the DNA-A. A clear geographic segregation was observed for both components, but differences between the genetic structure of cognate DNA-A and DNA-B components could be observed. Interestingly, DNA-B coding regions evolve at least under the same selection pressures than DNA-A coding regions. Our results indicate an interplay between reassortment and recombination acting at different levels across distinct subpopulations and genomic components.

Results

A total of 239 full-length DNA-B sequences were obtained in this study, all from samples of cultivated and non-cultivated plants from which a DNA-A had been previously cloned and analyzed (Lima et al., 2013; Ramos-Sobrinho et al., 2014; Rocha et al., 2013). These sequences were combined with 292 DNA-A and 76 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 Suppl. Tables S1 and S2. The final data set consisted of a total of 292 DNA-A and 315 DNA-B sequences: 117 DNA-A and 123 DNA-B sequences of Bean golden mosaic virus (BGMV), 30 DNA-A and 41 DNA-B sequences of Blainvillea yellow spot virus (BlYSV), 50 DNA-A and 53 DNA-B sequences of Euphorbia yellow mosaic virus, 21 DNA-A and 24 DNA-B sequences of Macroptilium yellow spot virus (MaYSV) and 74 DNA-A and 74 DNA-B sequences of Tomato severe rugose virus (ToSRV) (Figure 1). For each data set, the number of DNA-A and DNA-B sequences was similar (x_4^2 = 2.2432, P = 0.6911).

Phylogenetic analysis

Bayesian phylogenetic trees based on full-length DNA-A and DNA-B nucleotide sequences were constructed for all data sets (Figure 2). Although the PACo analysis provided significant evidence for global congruence between DNA-A and DNA-B trees (P≤0.001; Table 1), the residual square sum (m^2_{XY}) values were variable across the different data sets (0.1079 – 0.6039; Table 1), indicating that congruence levels are variable among them. Visual analysis of phylogenetic trees indicated the formation of large clades according to geographical region, for both components (Figure 2), providing initial evidence of population subdivision based on geography.

DNA-A and DNA-B phylogenetic trees for BGMV were perfectly symmetric, yielding "mirror image" trees (Figure 2A). In both trees it is possible to visualize two major clades supported by high posterior probability values and separated by long branches (Figure 2A). The first clade (DF/GO/MG) is comprised of isolates collected in Distrito Federal (DF), Goiás (GO) and Minas Gerais (MG) states, and the second clade (AL) is comprised of isolates sampled in Alagoas (AL) state. In addition, well-defined subclusters can be observed for both components, suggesting the existence of a substructure within a smaller geographical scale. The DF/GO/MG cluster can be subdivided into isolates from Florestal (MG) and from Paranoá (DF), Cristalina (GO), Santo Antônio de Goiás (GO) and Unaí (MG). The AL cluster also presented two

subclusters, one comprised of isolates sampled in Murici and another by isolates from Palmeira dos Índios (Figure 2A).

Conversely, DNA-A and DNA-B phylogenetic trees for ToSRV were not completely congruent (Figure 2B). A clear geographical clustering was observed for the DNA-B tree, with three clades comprised almost exclusively of isolates from Coimbra, Carandaí and Florestal/2014. Moreover, 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 Carandaí were not segregated (Figure 2B). 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 (Suppl. Table S3). This indicates that the exchange of components between isolates from Coimbra and Carandaí can occur.

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 2C). 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 Goiás (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 2D). For the DNA-A tree two well-supported major clades can be observed, while the DNA-B tree shows several poorly supported clades (Figure 2D). To minimize the effect of recombination, phylogenetic trees were reconstructed based on recombination-free DNA-A and DNA-B data sets (Figure 2E). Although MaYSV isolates were sampled in a smaller geographical scale (maximum distance among fields of approximately 200 kilometers) and a smaller number of sequences were obtained compared to the other data sets, two well-supported clades based in geographical location can be observed in the DNA-A tree, with isolates sampled in Pernambuco (PE) state clustered separately from those collected in neighboring AL (Figure 2E). Although two well-supported clades were also observed in the DNA-B tree, these were not perfectly congruent with the DNA-A tree (Figure 2E). 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.

Although the BlYSV data set contains isolates sampled in distant geographical regions (approximately 1800 km apart), phylogenetic trees do not display a clear and consistent pattern of geography-based clustering as noted for other data sets (Figure 2F, G). 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 components among isolates of the northeast with putative major parents in the southeast (Suppl. Table S3). For the DNA-B, isolates sampled in the southeast showed recombination events with putative minor parents in the northeast (Suppl. 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.

Genetic structure of populations

To compare the genetic structure based on DNA-A and DNA-B components, a nonparametric multivariate statistical analysis was performed using DAPC (Figure 3; Suppl. Table S4). Since DAPC requires predefined groups, the analysis was run sequentially with k values varying from 2 to 10, with the best number of genetic clusters chosen based on the minimum number of k after which the Bayesian Information Criterion (BIC) decreased by an insignificant amount. When a clear limit after which BIC values decreased by an insignificant amount could not be observed, the percentage of variation among genetic clusters inferred by k-means was plotted against BIC according to **a**nalysis of **mo**lecular **va**riance (AMOVA) (Suppl. Figure S1). For most data sets it was possible to define clear genetic clusters according to the criteria mentioned above. However, for BlYSV 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 (Suppl. Figure S1). Therefore, for these three data sets all individuals were considered to constitute a single population in later analyses.

In general, DAPC analysis across all data sets consistently discriminated all pre-defined clusters based on k-means (data not shown). Although a clear geographical segregation can be observed for the two components, especially for BGMV, ToSRV and EuYMV (consistent with phylogenetic analysis), slight differences between the genetic structure based on DNA-A and

DNA-B could be observed, with cognate components belonging to distinct genetic clusters (Figure 3).

For BGMV, DAPC analysis discriminated five and four genetics clusters for DNA-A and DNA-B, respectively, primarily based on geography (Figure 3). The clusters named AL-1, AL-2 and MG-1 were observed for both components, displaying a perfect pattern based on sampling location (Figure 3). 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 (Figure 3). 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 3). Although differences in the genetic structure between DNA-A and DNA-B could be due to the different number of clusters assigned in the k-means analysis, such differences were observed even when the same k number was tested for both components. As there is no evidence of host-specific adaptive selection (see below), the geographical distance between sampled sites is suggested as the main factor driving genetic differentiation between populations.

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

Similar to the patterns observed for BGMV and ToSRV, a clear geographical segregation was also observed for both EuYMV components, with only minor differences in the assignment of some DNA-A and DNA-B components to the three genetic clusters (South-1, South-2 and GO) (Figure 3).

Although it was not possible to define clear genetic clusters for MaYSV DNA-B components based on DAPC analysis, the DNA-A components were split into four groups

(Figure 3). Isolates from PE clustered separately from isolates sampled in AL (Figure 3). Moreover, isolates from AL grouped into three separate genetic clusters (AL-1, AL-2, AL-3) with no association with sampling location or host (Figure 3). When the data were analyzed without recombinant blocks, only two clusters were formed, with isolates sampled in AL forming a single cluster separately from those sampled in PE (Suppl. Figure S2).

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 2). In general, high values of pairwise Nst were observed among different clusters, indicating great differentiation and fully supporting the subdivision proposed by DAPC analysis.

Genetic variability

Although there is evidence that DNA-B components are more variable than DNA-A components (Briddon et al., 2010) a quantitative and comparative analysis using large intraspecific data sets has not yet been performed. Therefore, to address this issue, we analyzed cognate DNA-A and DNA-B components from begomoviruses infecting cultivated and non-cultivated hosts (Figure 1A). Initially, to verify the distribution of variation among genome components, the number of segregating sites (S) for the complete genome and the proportion of segregating sites corresponding to DNA-A and DNA-B were calculated and plotted for each data set (Figure 4A; Suppl. Tables S5 and S6). For almost all data sets, the DNA-B showed a higher proportion of segregating sites compared to the DNA-A (x_4^2=123.44, P<2.2x10^{-16}), indicating greater variability. The only exception was MaYSV, in which the number of segregating sites between the two components was similar (DNA-A S=506 and DNA-B S=503; x_4^2<0.09, P>0.92). Since S corresponds to the number of sites in which any mismatch is observed, not considering the number of mismatches at each site, we also calculated the average pairwise number of nucleotide differences (nucleotide diversity index, π). In agreement with the previous result, the π values for the DNA-B were statistically higher than those for the DNA-A for almost all data sets (Figure 4B; Suppl. Tables S5 and S6). Once again, the exception was MaYSV, for which the value for the DNA-A (π=0.07170) was statistically higher that for the DNA-B (π=0.05567) (Figure 4B; Suppl. Table S5 and S6). These results confirm the highest variability of the DNA-B in relation to the DNA-A at the populational level. However, this is not an absolute rule, as exemplified by MaYSV, which has a more variable DNA-A than DNA-B.

The nucleotide diversity index was also calculated on a sliding window across the DNA-A and DNA-B for each data set (Figure 4E). The results show an uneven distribution of variation along the components for almost all datasets analyzed. 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 4E). To verify the statistical significance of the uneven distribution of variation along the genomic components, the genome was partitioned and nucleotide sequences of the coding (CP, Rep, TrAP, REn ac4, NSP and MP) and intergenic regions (IR-A, LIR-B and SIR-B) were compared for each data set (Suppl. Figure S4). Unsurprisingly, non-coding regions were more variable compared to coding regions (Suppl. Figure S4). The SIR-B was more variable than the LIR-B, which in turn was more variable than the IR-A. In general, for coding regions there was no consistent pattern, with the most variable region being highly dependent on the data set analyzed (Suppl. Figure S4). The intragenic variation for nucleotide sequences of CP, Rep, NSP and MP and intraintergenic variation for IR-A and LIR-B were also analyzed (Suppl. Figure S5). As well as differences in the variation levels between genomic components and between regions within each component, there are clear differences in intragenic and intraintergenic variation levels (Suppl. Figure S5).

Recombination analysis

It is known that recombination plays an important role in the diversification and evolution of begomoviruses (Lefeuvre et al., 2007; Lima et al., 2017; Monci, Sanchez-Campos, Navas-Castillo, & Moriones, 2002). However, most studies were based on analysis of non-segmented viruses or of the DNA-A component of bipartite viruses, and little is known about the effect of recombination on the DNA-B component (Lefeuvre & Moriones, 2015). To investigate the effect of recombination on this component and possible differences in recombination patterns between the DNA-A and DNA-B, the intraspecific data sets were analyzed using RDP4.

There is a clear correlation between number of recombination events and genetic variability. BlYSV and MaYSV, which showed the higher number of events for both genomic components, 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 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 5A, B; Suppl. 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 3 and 7 unique events were detected for the DNA-A data sets from BlYSV and MaYSV, respectively (Figure 5A). The MaYSV DNA-B was the most prone to recombination, with 13 unique recombination events, followed by BlYSV, EuYMV, BGMV and ToSRV, with 9, 5, 4 and 1 unique events, respectively (Figure 5B). In DNA-A data sets the frequency of sequences that has at least one recombination event was higher for MaYSV (76.19%) followed by BlYSV (36.67%). In contrast, for the DNA-B the frequency of recombinant sequences was higher for BlYSV (97.56%), followed by MaYSV (75.00%) and EuYMV (39.62%). For BGMV and ToSRV DNA-B data sets, the frequency of recombinant sequences was similar (x_1^2=0.41648, P=0.5187) even though it was extremely low, with only 4.07% for BGMV and 1.35% for ToSRV. These results demonstrate an asymmetric distribution of recombination events between genomic components, with a higher propensity of the DNA-B to recombine compared to the DNA-A, and also suggest that populations of begomoviruses from non-cultivated hosts are more prone to recombination, as shown for MaYSV and BlYSV and to a lesser extent for EuYMV DNA-B.

Recombination breakpoints were mapped across of the DNA-A and DNA-B. Overall, a non-random distribution of recombination breakpoints was observed along both components. In agreement with previous studies (Lefeuvre et al., 2007), most recombination events detected in the DNA-A involved breakpoints located in the Rep gene (10 out 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 to the initiation codon of the MP gene. In contrast, the SIR-B showed only 2/64 recombination breakpoints. This low number of could be, at least in part, a consequence of its high nucleotide diversity, which impairs the capacity of RDP to detect reliable recombination events. More likely, it could simply be a recombination cold spot. Differently from the asymmetric breakpoint distribution observed for the CP and REP genes, the NSP and MP genes displayed a very homogeneous distribution of recombination breakpoints and a lower number of events compared to the intergenic regions (12/64 and 13/64 breakpoints, respectively). Despite approximately one third of the DNA-B length being comprised of intergenic regions, the number of breakpoints mapped to intergenic and coding regions was similar (x_1^2=3.0625, P=0.08012).

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

During an ORF finder analysis to verify the integrity of ORFs encoded by the DNA-B, we observed the presence of a large number of small ORFS in the LIR-B. To investigate the coding potential of these small ORFs, a BLASTn analysis was performed for all data sets. 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 the DNA-A (Figure 5C; Table 3). 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 3). tRep ORFs were detected in 10 MaYSV isolates (41.67%) collected in three different regions (Table 3), ruling out the possibility of them being cloning artifacts. 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 5C). Interestingly, all tRep ORFs were located upstream of the start codon and downstream of the promoter of the MP gene (Figure 5D), suggesting that they may be transcribed and translated, and possibly interfering with expression of the MP gene. The ratio of non-synonymous to synonymous substitutions (dN/dS) for homologous regions (105 nt/35 aa) of tRep ORFs was 0.261, indicating negative selection acting in this region and suggesting that these ORFs, if expressed, are functional. Although a recombination analysis was not performed (it is not possible to obtain a good alignment between non-homologous components), the presence of these tRep ORFs in the LIR-B is a strong indication of inter-component recombination. An analysis of the LIR-B from the other four begomoviruses failed to identify any truncated ORFs homologous to the N-terminal region of the Rep gene as observed in MaYSV.

Reassortment analysis

As mentioned above for phylogenetic reconstruction, PACo analysis provided significant evidence for the global congruence between DNA-A and DNA-B trees (Table 1). 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 variable across the different data sets (Table 1), ranging from 0.1079 for ToSRV to 0.6039 for MaYSV, indicating that congruence levels are variable among different 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 6). This approach

assumes that squared residuals between DNA-A and DNA-B links that are significantly high reflect phylogenetic incongruence, indicating the occurrence of reassortment. Analysis of the estimated jackknife squared residuals for each link clearly indicate the occurrence of reassortment in all analyzed datasets (Figure 6). Tanglegram plots depicting pairs of DNA-A and DNA-B phylogenetic trees also suggest the occurrence of reassortment in all data sets (Figure 7). Interestingly, some groups within of the same data set are apparently more prone to reassortment, as evidenced by a higher entanglement in corresponding clades of the phylogenies (Figure 7). To test this possibility, jackknife squared residuals values were grouped according to sampling location and compared by estimating their 95% bootstrap confidence intervals from 1,000 nonparametric simulations. Indeed, some groups showed significantly higher residuals values (Suppl. Figure S6), indicative of a higher contribution to incongruence and a higher frequency of reassortment genomes. For example, for the EuYMV data set, significantly higher residual values were observed for sequences sampled in Rio Grande do Sul and Mato Grosso do Sul compared to Goiás and Paraná (Figure 6). Conversely, for the BlYSV data set there was no difference in residual values between sequences sampled at different locations, indicating no differences in reassortment frequency (Figure 6). 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. For almost all data sets, most of the reassortment events detected by RDP were supported by PACo analysis, displaying high squared residuals (Figure 6). The frequency of reassortment sequences detected by RDP, in agreement with PACo global test, was higher for MaYSV (19/25; 76.00%), followed by EuYMV (29/55; 52.72%), BGMV (35/117; 29.41%) and BlYSV (5/24; 20.83%) (x_3^2=25.881, P=1.01x10^{-05}; Table 4). For ToSRV only two sequences were detected as putatively reassortment by RDP. However, PACo analysis suggests a greater number of reassortment events, with 16/67 links (23.88%) 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 (97.03-100% and 94.51-100% for DNA-A and DNA-B, respectively), impairing the capacity of RDP to detect reliable reassortment events.

Interestingly and surprisingly, when the BGMV and MaYSV data sets including recombinant blocks were analyzed, the global square sum was lower compared to the same data sets without recombinant blocks (Table 1; there was no difference in the case of the EuYMV and BlYSV data sets). Although the differences were small, this result suggests that recombination may somehow be restoring the congruence between phylogenies. To test the

consistency of this result, the topological congruence between the phylogenetic trees in data sets with and without recombinant blocks was evaluated using a topology-based method, normalized PH85 distance (nPH85) (Geoghegan, Duchene, & Holmes, 2017). While for MaYSV the nPH85 distance was also lower for recombinant data set ($nPH85_{recombination}$=0.789 vs. $nPH85_{without\ recombination}$=0.828), there was no difference for BGMV, EuYMV and BlYSV (data not shown).

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 non-synonymous to synonymous substitutions (dN/dS) was compared for each gene/genomic component. The dN/dS ratios were <1 for almost all genes (Table 5), indicating the predominance of negative selection acting on both components. In agreement with these results, a higher number of individual sites under negative selection was detected in all data sets (Table 5). Only the ac4 gene has a dN/dS ratio >1 for BGMV (2.53), EuYMV (1.15) and MaYSV (1.45) (Table 5).

Although negative selection is predominant, dN/dS ratios were highly variable among genes/genomic components (0.028-2.530 for the DNA-A, 0.0667-0.2311 for the DNA-B) (Table 5). These results indicate that different genes/genomic components 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. To compare 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 (Figure 8). Considering data grouped by genomic component, the average dN/dS values were statistically lower for DNA-B (0.1562) compared to DNA-A (0.4778) (Figure 8A). In addition, the number of individual sites under negative selection was statistically higher for the DNA-B (Figure 8B), suggesting a stronger negative selection acting on this component. However, these results should be treated with caution. Overlapping genes (which occur in the DNA-A but not in the DNA-B) may limit the accumulation of synonymous substitutions or a higher accumulation of non-synonymous substitution in one of the overlapping genes, leading to an increase in the dN/dS ratio (Torres et al., 2013). Thus, the higher dN/dS ratios in the DNA-A can be wrongly interpreted as evidence of positive selection or more relaxed evolution. Indeed, the ac4 and TrAP genes, which have the higher dN/dS ratios (Figure 8A), are both overlapping genes in the DNA-A. Therefore, the possibility of this result being an artifact cannot be excluded. 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.

Considering data grouped by genes, the dN/dS ratio was statistically higher for ac4 and TrAP (Figure 8A). The CP and MP genes showed the lowest dN/dS ratios (Figure 8A). Moreover, ac4 and TrAP (and also REn) showed a similar number of individual sites under negative selection, statistically lower than CP, Rep, MP and NSP (Figure 8B).

The number of sites under positive selection was very low for both components, with no significant difference (Figure 8C). In agreement with the high dN/dS ratio of the ac4 gene of BGMV and EuYMV (Table 5), 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 5). In addition, for BlYSV TrAP we found evidence of positive selection by PARRIS ($P<0.043$), and one (codon 90) and ten (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 5).

Discussion

Little is known about the significance of genome multipartition for viral biology and evolution (Gutierrez & Zwart, 2018; Lucia-Sanz et al., 2018; Lucia-Sanz & Manrubia, 2017; Sicard et al., 2016). Independence among segments, in addition to functional and structural differences, might heavily impact the evolutionary processes acting on the different segments, especially the generation and maintenance of diversity (Betancourt, Fereres, Fraile, & García-Arenal, 2008; Gallet et al., 2018; Grigoras et al., 2010; Lozano et al., 2009; Pita & Roossinck, 2013; Sicard et al., 2013).

Several key events marked the evolution of geminiviruses, culminating with the emergence of segmented (bipartite) genomes in viruses classified in the genus Begomovirus (Rojas, Hagen, Lucas, & Gilbertson, 2005). Expansion of the genome through the capture of a new component by an ancestral non-segmented virus, together with the loss of the V2 gene [present only in Old World (OW) begomoviruses] 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, García-Arenal, & Pagán, 2015).

Our results, based on analysis of five NW bipartite begomoviruses, indicate that DNA-A and DNA-B components 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.

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, leading to fitness trade-offs and negative selection. In addition, 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 component under stricter evolutionary constraints. However, using an interspecific data set, Ho, Kuchie, and Duffy (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 components 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 BlYSV MP and NSP genes are its most variable genes. We also observed 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 non-synonymous substitutions (Gardiner et al., 1988; Jeffrey, Pooma, & Petty, 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, being more permissive to variation (Padidam, Beachy, & Fauquet, 1996; Roshan et al., 2018; Rothenstein, Krenz, Selchow, & Jeske, 2007). Moreover, NW DNA-B components accommodate additional functions besides movement, such as suppression of plant defense 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, in absolute terms, when the whole component is

considered, the DNA-B may support a greater accumulation of variation compared to the DNA-A, since it has two IRs comprising approximately 1/3 of its length, in contrast to the DNA-A IR encompassing only 1/8 of its length. On this regard, our results show that the SIR-B is the most variable genome region, followed by the LIR-B and the IR-A. The SIR-B has no known cis-elements involved in replication or transcription, thus being more permissive to variation. On the contrary, the LIR-B and IR-A contain important cis-elements, therefore being under greater selective pressure. However, the LIR-B is almost three times longer than the IR-A, thus supporting a greater accumulation of variation without interference in regulatory cis-elements.

Asymmetric accumulation among the different genomic components in multipartite viruses has been shown, and 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, Zwart, Sánchez-Navarro, & Elena, 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). Furthermore, it has been proposed that segments accumulating at higher frequency would have a greater mutation load, and consequently would evolve faster than other components present at a lower frequency (Gutierrez & Zwart, 2018; Lozano et al., 2009). Lozano et al. (2009) suggested that the higher variability observed for the RNA-dependent RNA polymerase (RdRp) of Tomato chlorosis virus (ToCV; genus Crinivirus) is due to its location in the RNA1, which reaches a higher titer than the RNA2 and therefore accumulates a higher number of mutations in populational terms. A similar pattern was observed for the nanovirus FBNSV, where the high-titer DNA-N component is the most variable, whereas, DNA-R and DNA-S are the most conserved segments and accumulate at the lowest levels (Grigoras et al., 2010; Sicard et al., 2013). However, for other segments this relationship is not very clear, consistent with the joint action of other factors modulating the propensity to variation. Although no study has 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, Khatabi, & Fondong, 2019; V.B. Pinto and F.M. Zerbini, unpublished). Considering that the two components 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, Froissart, Michalakis, & Blanc, 2008; Moury, Fabre, & Senoussi, 2007). After severe bottlenecks, a large reduction in the effective population size (Ne) can favor 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 components could at least in part be influenced by genetic drift, due to differences in the strength of the bottlenecks experienced by each component specially during transmission by B. tabaci.

Recombination is considered one of the main forces driving begomovirus evolution (Lefeuvre, Lett, Varsani, & Martin, 2009; Lefeuvre et al., 2007; Lima et al., 2017; Lima et al., 2013; 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 components compared to DNA-A components in all data sets. Martin, van der Walt, Posada, and Rybicki (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 intra- genomic interactions established. Thus, the greater propensity of the DNA-B to recombine may be explained by its lowest organizational complexity (only two non-overlapping genes, compared to one non-overlapping and four overlapping genes in the DNA-A) and the lowest number of inter- and intragenomic interactions of its encoded proteins (Hanley-Bowdoin, Bejarano, Robertson, & Mansoor, 2013).

In contrast to our results, previous studies analyzing the evolutionary dynamics of cassava mosaic geminiviruses (CMG) reported a much smaller number of unique recombinations events for the DNA-B compared to the DNA-A (De Bruyn et al., 2016; De Bruyn et al., 2012). However, these results were based on the analysis of interspecific data sets with a higher number of DNA-A than DNA-B sequences. Since OW DNA-B components have

a higher degree of variation compared to the DNA-A (as discussed above), obtaining a reliable interspecific alignment may be difficult, impairing the capacity of RDP to detect reliable recombination events. In addition, sampling more sequences might increase the likelihood of parental and recombinant sequences being sampled together, increasing the number of events detected.

Rodelo-Urrego et al. (2015) observed no difference in the frequency of recombinant sequences between DNA-A and DNA-B components 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 BlYSV, 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 species, 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 components and viruses.

Although less frequent, intercomponent homologous recombination involving the exchange of replication origin among heterologous components in multipartite viruses has been previously reported (Grigoras et al., 2014; Hu et al., 2008; 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 begomoviruses from NW. These Rep fragments, ranging from 35 to 51 nt in length, were located in the viral-sense strand and it was suggested that they could be involved in the posttranscriptional regulation of the cognate Rep gene. This suggests that non-homologous recombination among heterologous components may also occur. We detected small truncated Rep (tRep) ORFs in the complementary-sense strand of the LIR-B of several MaYSV isolates, ranging from 35 to 95 amino acids (significantly longer than the short regions identified by Gregorio-Jorge et al. (2010)). The presence of these tRep ORFs is a strong indication of intercomponent recombination. The tRep ORFs were located downstream of the MP gene promoter, suggesting that they can be expressed. Moreover, they include the coding region for Motif I and the IRD, which are involved in the process of recognition and binding of Rep to the origin of replication (Arguello-Astorga et al., 1994). If the tRep ORFs are expressed, one possibility is that the

truncated proteins can compete with Rep for binding at the origin of replication, negatively interfering in the replication process. Another possibility is that tRep ORFs might interfere with the translation of MP, due to their location upstream of the MP initiation codon. However, these possibilities need to be experimentally tested. Although non-homologous 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 component such as the DNA-B and might accommodate the emergence of new functions, as suggested above.

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, Lefeuvre, Roumagnac, & Martin, 2018). We were not able to estimate and compare evolutionary rates between DNA-A and DNA-B data sets, since the sequences were sampled in a 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 components. 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 (mainly the presence of overlapping genes in the DNA-A) and the specialized functions of each component 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 component may evolve under at least 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, although molecular clock-based methods to estimate substitution rates consider evolution to be driven exclusively by mutational dynamics, recombination constitutes an important mechanism that may speed up begomovirus evolution (Lefeuvre et al., 2009; Lefeuvre et al., 2007). Differences in the propensity for recombination between DNA-A and DNA-B may drastically affect their evolution, guiding each component to distinct patterns of variation. Therefore, even with DNA-A and DNA-B working together as a single functional unit, with both components necessary to systemically infect the plant, only those regulatory elements required to maintain the integrity and functionality of genome may show similar evolutionary rates driven by reciprocal selection, unlike the remaining of the genome.

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; Ramos-Sobrinho et al., 2014; Rocha et al., 2013). 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 components of BGMV, ToSRV, and EuYMV, and to a lesser extent for the DNA-A of MaYSV. For BlYSV, 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.

Slight differences in the genetic structure between DNA-A and DNA-B were observed for BGMV, ToSRV and EuYMV, consistent a role of reassortment. Indeed, we were able to detect reassortment events across all data sets using two different methods. However, all events detected were intraspecific and most exchanged components were restricted among geographically and genetically close subpopulations, probably contributing little for the genetic makeup of the subpopulation. These results may reflect a limited dispersion capacity of the virus, which is primarily dependent on the vector. If this is the case, the genetic structure of begomovirus population would match the genetic structure of B. tabaci populations. Thus, studies addressing the genetic structure of both the virus and its vector might provide insights on the ecological and genetic processes that affect begomovirus populations.

The extent of evolutionary congruence among distinct genomic components of segmented and multipartite viruses depends on the interplay between evolutionary and biological process acting at both the inter- and intragenomic levels. Thus, different structural and functional constraints along genomic regions might drive such regions and even whole segments to distinct patterns of evolution. Although it is expected that different segments will follow a similar evolutionary way, given the functional dependence among segments, previous studies showed that different components in multipartite viruses might experiment distinct evolutionary routes (Briddon et al., 2010; Grigoras et al., 2014; Hu et al., 2008; Lozano et al., 2009; Ohshima et al., 2016; Roossinck, 2002). Our results demonstrate 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 component may behave as a single independent entity, exploring a larger portion of sequence space and allowing for faster adaptation to a constantly changing environment.

Methods

Cloning and sequencing of DNA-B components

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^{o}C. Total DNA was used as a template for rolling-circle amplification (RCA) of viral genomes as described by Inoue-Nagata, Albuquerque, Rocha, and Nagata (2004). To facilitate the cloning of DNA-B components, 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 component were chosen. Unit genome-length fragments (approximately 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.) by primer walking. Full-length begomovirus genomes were assembled using Geneious v. 8.1 (Kearse et al., 2012). Sequences were initially analyzed with the BLASTn algorithm (Altschul et al., 1990) to confirm that they corresponded to a DNA-B. The genomic organization and integrity of the open reading frames (ORF) were checked using the Fangorn Forest method (Silva, Carvalho, Fontes, & Cerqueira, 2017) implemented in Geminivirus Data Warehouse (Jose Cleydson F. Silva et al., 2017). Sequences with truncated ORFs or with atypical DNA-B organization were discarded.

Multiple sequence alignments and phylogenetic analysis

All genome sequences were organized to begin at the nicking site in the invariant nonanucleotide at the origin of replication (5'-TAATATT//AC-3'). Multiple sequence alignments were prepared for the full-length nucleotide sequences of DNA-A and DNA-B, of the CP (capsid protein), Rep (replication-associated protein), REn (replication enhancer protein), TrAP (trans-activating protein), AC4, MP (movement protein) and NSP (nuclear shuttle protein) 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). Alignments were manually checked and adjusted when necessary. The same alignments were used for all subsequently performed analyzes.

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). The program 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 (AIC). The analyses were carried out running 10,000,000 generations and excluding the first 2,500,000 generations as burn-in. 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 Development Core Team, 2007 8112).

Variability indices and genetic structure

The average pairwise number of nucleotide differences per site (nucleotide diversity, π) was calculated using a script (available from the authors upon request) written in R software. 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, Sánchez-DelBarrio, Messeguer, & Rozas, 2003).

Inferences about population genetic structure was performed using Discriminant Analysis of Principal Components (DAPC) (Jombart, Devillard, & Balloux, 2010), a nonparametric multivariate static 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 pre-defined 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). Since there is not always a clear limit at which the BIC value decreases by an insignificant amount, the percentage of variation among genetic clusters inferred by k-means was plotted with BIC according to **a**nalysis of **mo**lecular **va**riance

(AMOVA) performed with the Poppr package (Kamvar, Tabima, & Grunwald, 2014) in R software. After assigning individuals to each inferred genetic cluster, the optimum PC values to run discriminant analysis were investigated through visual analysis of the cumulative variance graph explained by PC and a-score optimization. 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).

Subpopulation will be used here to define a group of individuals sampled at the same location without previous knowledge of population structure, and genetic cluster as a set of individuals genetically similar defined based on the DAPC analysis.

Recombination and reassortment analysis

The occurrence of recombination was analyzed for the DNA-A and DNA-B separately using intraspecific data sets. Recombination analysis was performed using the Rdp, Geneconv, Bootscan, Maximum Chi Square, Chimaera, SisterScan and 3Seq methods implemented in Recombination Detection Program (RDP) 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 cut-off of 0.05. Only recombination events detected by at least four different methods were considered reliable.

To assess the occurrence of reassortment (or pseudorecombination), the topological congruence between the inferred Bayesian phylogenetic trees based on full-length nucleotide sequences of the DNA-A and DNA-B were analyzed with Procrustean Approach to Cophylogeny (PACo). PACo is a distance-based method which assesses the global congruence between trees and the contribution of each individual DNA-A and DNA-B association to the global congruence statistic (Balbuena, Miguez-Lozano, & Blasco-Costa, 2013). The method uses matrices of patristic distances calculated from inferred Bayesian DNA-A and DNA-B phylogenies, testing the null hypothesis of no congruence between trees, a through permutation test. Matrices of patristic distances from DNA-A and DNA-B phylogenies were calculated using the ape package (Paradis, Claude, & Strimmer, 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. Individual links that highly contribute to global congruence have a small

contribution to the global square sum. The statistical significance for global test 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. Since recombination can also lead to incongruence between phylogenies, analyzes were performed in data sets with and without recombinant blocks. In addition, reassortment was analyzed 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 components using RDP as described previously.

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) (Sergei L. Kosakovsky-Pond & Simon D. W. Frost, 2005) and partitioning for robust inference of selection (PARRIS) (Scheffler, Martin, & Seoighe, 2006), implemented in the DataMonkey webserver (S.L. Kosakovsky-Pond & S.D.W. Frost, 2005). The mean ratios of non-synonymous 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 webserver. 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.

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(17), 8345-8350.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215, 403-410.

Andrade, E. C., Manhani, G. G., Alfenas, P. F., Calegario, R. F., Fontes, E. P. B., & Zerbini, F. M. (2006). Tomato yellow spot virus, a tomato-infecting begomovirus from Brazil with a closer

relationship to viruses from Sida sp., forms pseudorecombinants with begomoviruses from tomato but not from Sida. Journal of General Virology, 87, 3687-3696.

Arguello-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 interative elements: a model for replication. Virology, 203, 90-100.

Balbuena, J. A., Miguez-Lozano, R., & Blasco-Costa, I. (2013). PACo: a novel procrustes application to cophylogenetic analysis. PLoS ONE, 8(4), e61048.

Baltimore, D. (1971). Expression of animal virus genomes. Bacteriological Reviews, 35(3), 235-241.

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(24), 12416-12421.

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.

Chao, L. (1991). Levels of selection, evolution of sex in RNA viruses, and the origin of life. Journal of Theoretical Biology, 153(2), 229-246.

Chen, K., Khatabi, B., & Fondong, V. N. (2019). The AC4 protein of a cassava geminivirus is required for virus infection. Molecular Plant-Microbe Interactions, 32.

De Bruyn, A., Harimalala, M., Zinga, I., Mabvakure, B. M., Hoareau, M., Ravigne, V., . . . Lefeuvre, P. (2016). Divergent evolutionary and epidemiological dynamics of cassava mosaic geminiviruses in Madagascar. BMC Evolutionary Biology, 16, 182.

De Bruyn, A., Villemot, J., Lefeuvre, P., Villar, E., Hoareau, M., Harimalala, M., . . . Lett, J. M. (2012). East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus. BMC Evolutionary Biology, 12, 228.

de Vienne, D. M., Refregier, G., Lopez-Villavicencio, M., Tellier, A., Hood, M. E., & Giraud, T. (2013). Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. New Phytologist, 198(2), 347-385.

Dolja, V. V., & Koonin, E. V. (2018). Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. Virus Research, 244, 36-52.

Duffy, S., & Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus Tomato yellow leaf curl virus. Journal of Virology, 82(2), 957-965.

Duffy, S., & Holmes, E. C. (2009). Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. Journal of General Virology, 90, 1539-1547.

Elena, S. F. (2016). Evolutionary transitions during RNA virus experimental evolution. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1701).

Faria, J. C., Gilbertson, R. L., Hanson, S. F., Morales, F. J., Ahlquist, P. G., Loniello, A. O., & Maxwell, D. P. (1994). Bean golden mosaic geminivirus type II isolates from the Dominican Republic and Guatemala: Nucleotide sequences, infectious pseudorecombinants, and phylogenetic relationships. Phytopathology, 84, 321-329.

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.

Gale, M., Jr., Tan, S. L., & Katze, M. G. (2000). Translational control of viral gene expression in eukaryotes. Microbiology and Molecular Biology Reviews 64(2), 239-280.

Gallet, R., Fabre, F., Thebaud, 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(14).

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(2), 302-312.

Garcia-Arriaza, J., Manrubia, S. C., Toja, M., Domingo, E., & Escarmis, C. (2004). Evolutionary transition toward defective RNAs that are infectious by complementation. Journal of Virology, 78(21), 11678-11685.

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.

Garrido-Ramirez, E. R., Sudarshana, M., & Gilbertson, R. L. (2000). Bean golden yellow mosaic virus from Chiapas, Mexico: Characterization, pseudorecombination with other bean-infecting geminiviruses and germ plasm screening. Phytopathology, 90, 1224-1232.

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(11), 5902-5907.

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 Pathogens, 13(2), e1006215.

Gilbertson, R. L., Hidayat, S. H., Paplomatas, E. J., Rojas, M. R., Hou, Y.-H., & Maxwell, D. P. (1993). Pseudorecombination between infectious cloned DNA components of tomato mottle and bean dwarf mosaic geminiviruses. Journal of General Virology, 74, 23-31.

Gregorio-Jorge, J., Bernal-Alcocer, A., Banuelos-Hernandez, B., Alpuche-Solis, A. G., Hernandez-Zepeda, C., Moreno-Valenzuela, O., . . . Arguello-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.

Grigoras, I., Ginzo, A. I., Martin, D. P., Varsani, A., Romero, J., Mammadov, A., . . . Gronenborn, B. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. Journal of General Virology, 95, 1178-1191.

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(18), 9105-9117.

Gutierrez, S., & Zwart, M. P. (2018). Population bottlenecks in multicomponent viruses: first forays into the uncharted territory of genome-formula drift. Current Opinion in Virology, 33, 184-190.

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

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(11), e111957.

Hou, Y. M., & Gilbertson, R. L. (1996). Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. Journal of Virology, 70, 5430-5436.

Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J., & Yeh, H. H. (2008). Reassortment and concerted evolution in Banana bunchy top virus genomes. Journal of Virology, 82(12), 6088-6088.

Hu, Z., Zhang, X., Liu, W., Zhou, Q., Zhang, Q., Li, G., & Yao, Q. (2016). Genome segments accumulate with different frequencies in Bombyx mori bidensovirus. Journal of Basic Microbiology, 56(12), 1338-1343.

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(2), 287-294.

Hutchinson, M. C., Cagua, E. F., Balbuena, J. A., Stouffer, D. B., & Poisot, T. (2017). paco: implementing Procrustean Approach to Cophylogeny in R. Methods in Ecology and Evolution, 8(8), 932-940.

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(2), 209-211.

Iranzo, J., & Manrubia, S. C. (2012). Evolutionary dynamics of genome segmentation in multipartite viruses. Proc Biol Sci, 279(1743), 3812-3819.

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(1), 208-218.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, 27(21), 3070-3071.

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(1), 94.

Kamvar, Z. N., Tabima, J. F., & Grunwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, 2, e281.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647-1649.

Koonin, E. V., Dolja, V. V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology, 479–480, 2-25.

Kosakovsky-Pond, S. L., & Frost, S. D. W. (2005). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics, 21(10), 2531-2533.

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. Molecular Biology and Evolution, 22, 1208-1222.

Lazarowtiz, S. G., Wu, L. C., Rogers, S. G., & Elmer, J. S. (1992). Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. Plant Cell, 4, 799-809.

Lefeuvre, P., Lett, J. M., Varsani, A., & Martin, D. P. (2009). Widely conserved recombination patterns among single-stranded DNA viruses. Journal of Virology, 83(6), 2697-2707.

Lefeuvre, P., Martin, D. P., Harkins, G., Lemey, P., Gray, A. J. A., Meredith, S., . . . Heydarnejad, J. (2010). The spread of tomato yellow leaf curl virus from the Middle East to the world. PLoS Pathogens, 6(10), e1001164.

Lefeuvre, P., Martin, D. P., Hoareau, M., Naze, F., Delatte, H., Thierry, M., . . . Lett, J. M. (2007). Begomovirus 'melting pot' in the south-west Indian Ocean islands: Molecular diversity and evolution through recombination. Journal of General Virology, 88(Pt 12), 3458-3468.

Lefeuvre, P., & Moriones, E. (2015). Recombination as a motor of host switches and virus emergence: Geminiviruses as case studies. Current Opinion in Virology, 10, 14-19.

Lima, A. T. M., Silva, J. C. F., Silva, F. N., Castillo-Urquiza, G. P., Silva, F. F., Seah, Y. M., . . . Zerbini, F. M. (2017). The diversification of begomovirus populations is predominantly driven by mutational dynamics. Virus Evolution, 3(1), vex005.

Lima, A. T. M., Sobrinho, R. R., Gonzalez-Aguilera, J., Rocha, C. S., Silva, S. J. C., Xavier, C. A. D., . . . 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.

Lister, R. M. (1966). Possible relationships of virus-specific products of tobacco rattle virus infections. Virology, 28(2), 350-353.

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(24), 12973-12983.

Lucia-Sanz, A., Aguirre, J., & Manrubia, S. (2018). Theoretical approaches to disclosing the emergence and adaptive advantages of multipartite viruses. Current Opinion in Virology, 33, 89-95.

Lucia-Sanz, A., & Manrubia, S. (2017). Multipartite viruses: adaptive trick or evolutionary treat? NPJ Systems Biology and Applications, 3, 34.

Mar, T. B., Xavier, C. A. D., Lima, A. T. M., Nogueira, A. M., Silva, J. C. F., Ramos-Sobrinho, R., . . . Zerbini, F. M. (2017). Genetic variability and population structure of the New World begomovirus Euphorbia yellow mosaic virus. Journal of General Virology, 98(6), 1537-1551.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution, 1, 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(4), e51.

Michalakis, Y., & Blanc, S. (2018). Editorial overview: Multicomponent viral systems. Current Opinion in Virology, 33, vi-ix.

Monci, F., Sanchez-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(2), 317-326.

Monsion, B., Froissart, R., Michalakis, Y., & Blanc, S. (2008). Large bottleneck size in cauliflower mosaic virus populations during host plant colonization. PLoS Pathogens, 4(10), 7.

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.

Nee, S. (1987). The evolution of multicompartmental genomes in viruses. Journal of Molecular Evolution, 25(4), 277-281.

Nylander, J. A. A. (2004). MrModeltest v2.

Ohshima, K., Matsumoto, K., Yasaka, R., Nishiyama, M., Soejima, K., Korkmaz, S., . . . Takeshita, M. (2016). Temporal analysis of reassortment and molecular evolution of Cucumber mosaic virus: Extra clues from its segmented genome. Virology, 487, 188-197.

Ojosnegros, S., Garcia-Arriaza, J., Escarmis, C., Manrubia, S. C., Perales, C., Arias, A., . . . Domingo, E. (2011). Viral genome segmentation can result from a trade-off between genetic content and particle stability. PLoS Genetics, 7(3), e1001344.

Padidam, M., Beachy, R. N., & Fauquet, C. M. (1996). The role of AV2 ("precoat") and coat protein in viral replication and movement in tomato leaf curl geminivirus. Virology, 224, 390-404.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics, 20(2), 289-290.

Peng, R. D. (2008). simpleboot: Simple Bootstrap Routines.

Pita, J. S., & Roossinck, M. J. (2013). Mapping viral functional domains for genetic diversity in plants. Journal of Virology, 87(2), 790-797.

Pond, S. L. K., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). GARD: a genetic algorithm for recombination detection. Bioinformatics, 22, 3096-3098.

Prangishvili, D. (2013). The wonderful world of archaeal viruses. Annual Review of Microbiology, 67, 565-585.

Prasanna, H. C., Sinha, D. P., Verma, A., Singh, M., Singh, B., Rai, M., & Martin, D. P. (2010). The population genomics of begomoviruses: Global scale population structure and gene flow. Virology Journal, 7, 220.

Pressing, J., & Reanney, D. C. (1984). Divided genomes and intrinsic noise. Journal of Molecular Evolution, 20(2), 135-146.

R Development Core Team. (2007). R: A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria.

Ramos-Sobrinho, R., Xavier, C. A. D., Pereira, H. M. B., Lima, G. S. A., Assunção, I. P., Mizubuti, E. S. G., . . . Zerbini, F. M. (2014). Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts. Journal of General Virology, 95(11), 2540-2552.

Rocha, C. S., Castillo-Urquiza, G. P., Lima, A. T. M., Silva, F. N., Xavier, C. A. D., Hora-Junior, B. T., . . . Zerbini, F. M. (2013). Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. Journal of Virology, 87(10), 5784-5799.

Rodelo-Urrego, M., García-Arenal, F., & Pagán, I. (2015). The effect of ecosystem biodiversity on virus genetic diversity depends on virus species: A study of chiltepin-infecting begomoviruses in Mexico. Virus Evolution, 1(1), vev004.

Rojas, M. R., Hagen, C., Lucas, W. J., & Gilbertson, R. L. (2005). Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. Annual Review of Phytopathology, 43, 361-394.

Rojas, M. R., Macedo, M. A., Maliano, M. R., Soto-Aguilar, M., Souza, J. O., Briddon, R. W., . . . Gilbertson, R. L. (2018). World management of geminiviruses. Annual Review of Phytopathology, 56, 637-677.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19(1), 1572-1574.

Roossinck, M. J. (2002). Evolutionary history of Cucumber mosaic virus deduced by phylogenetic analyses. Journal of Virology, 76(7), 3382-3387.

Roshan, P., Kulshreshtha, A., Kumar, S., Purohit, R., & Hallan, V. (2018). AV2 protein of tomato leaf curl Palampur virus promotes systemic necrosis in Nicotiana benthamiana and interacts with host Catalase2. Scientific Reports, 8(1), 1273.

Rothenstein, D., Krenz, B., Selchow, O., & Jeske, H. (2007). Tissue and cell tropism of Indian cassava mosaic virus (ICMV) and its AV2 (precoat) gene product. Virology, 359(1), 137-145.

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.

Saleem, H., Nahid, N., Shakir, S., Ijaz, S., Murtaza, G., Khan, A. A., . . . Nawaz-Ul-Rehman, M. S. (2016). Diversity, Mutation and Recombination Analysis of Cotton Leaf Curl Geminiviruses. PLoS ONE, 11(3), e0151161.

Sanchez-Navarro, J. A., Zwart, M. P., & Elena, S. F. (2013). Effects of the number of genome segments on primary and systemic infections with a multipartite plant RNA virus. Journal of Virology, 87(19), 10805-10815.

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. Journal of Molecular Evolution, 49, 672-681.

Saunders, K., Salim, N., Mali, V. R., Malathi, V. G., Briddon, R., Markham, P. G., & Stanley, J. (2002). Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. Virology, 293(1), 63-74.

Scheffler, K., Martin, D. P., & Seoighe, C. (2006). Robust inference of positive selection from recombining coding sequences. Bioinformatics, 22(20), 2493-2499.

Shi, M., Lin, X. D., Chen, X., Tian, J. H., Chen, L. J., Li, K., . . . Zhang, Y. Z. (2018). The evolutionary history of vertebrate RNA viruses. Nature, 556(7700), 197-202.

Sicard, A., Michalakis, Y., Gutierrez, S., & Blanc, S. (2016). The Strange Lifestyle of Multipartite Viruses. 12(11), e1005819.

Sicard, A., Yvon, M., Timchenko, T., Gronenborn, B., Michalakis, Y., Gutierrez, S., & Blanc, S. (2013). Gene copy number is differentially regulated in a multipartite virus. Nature Communications, 4, 2248.

Silva, F. N., Lima, A. T. M., Rocha, C. S., Castillo-Urquiza, G. P., Alves-Júnior, M., & Zerbini, F. M. (2014). Recombination and pseudorecombination driving the evolution of the begomoviruses Tomato severe rugose virus (ToSRV) and Tomato rugose mosaic virus (ToRMV): Two recombinant DNA-A components sharing the same DNA-B. Virology Journal, 11, 66.

Silva, J. C. F., Carvalho, T. F. M., Basso, M. F., Deguchi, M., Pereira, W. A., Sobrinho, R. R., . . . Fontes, E. P. B. (2017). Geminivirus data warehouse: a database enriched with machine learning approaches. BMC Bioinformatics, 18(1), 240.

Silva, J. C. F., Carvalho, T. F. M., Fontes, E. P. B., & Cerqueira, F. R. (2017). Fangorn Forest (F2): a machine learning approach to classify genes and genera in the family Geminiviridae. BMC Bioinformatics, 18, 431.

Solé, R. (2016). The major synthetic evolutionary transitions. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1701), 20160175.

Szathmary, E. (1992). Natural selection and dynamical coexistence of defective and complementing virus segments. Journal of Theoretical Biology, 157(3), 383-406.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution, 30(12), 2725-2729.

Torres, C., Fernandez, M. D., Flichman, D. M., Campos, R. H., & Mbayed, V. A. (2013). Influence of overlapping genes on the evolution of human hepatitis B virus. Virology, 441(1), 40-48.

Turner, P. E. (2003). Searching for the advantages of virus sex. Origins of Life and Evolution of the Biosphere, 33(1), 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.

Vijaykrishna, D., Mukerji, R., & Smith, G. J. D. (2015). RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion. PLoS Pathogens, 11(7), e1004902.

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. FEMS Microbiology Reviews, 28(2), 127-181.

Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucia-Sanz, A., Kuhn, J. H., Krupovic, M., . . . Koonin, E. V. (2018). Origins and Evolution of the Global RNA Virome. Mbio, 9(6).

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(1), 5004.

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(13), 5762-5768.

Zerbini, F. M., Briddon, R. W., Idris, A., Martin, D. P., Moriones, E., Navas-Castillo, J., . . . ICTV Consortium. (2017). ICTV Virus Taxonomy Profile: Geminiviridae. Journal of General Virology, 98(2), 131-133.

Zorzatto, C., Machado, J. P., Lopes, K. V., Nascimento, K. J., Pereira, W. A., Brustolini, O. J., . . . Fontes, E. P. (2015). NIK1-mediated translation suppression functions as a plant antiviral immunity mechanism. Nature, 520(7549), 679-682.

Table 1. Results of co-phylogeny analysis between begomovirus DNA-A and DNA-B segments using Procrustean Approach to Cophylogeny (PACo) for 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) data sets.

Virus	Global goodness-of-fit	
	m^2_{xy}*	P-value
BGMV#	0.4992	0.001
BGMV&	0.5411	0.001
BlYSV#	0.1682	0.001
BlYSV&	0.1623	0.001
EuYMV#	0.3059	0.001
EuYMV&	0.2970	0.001
MaYSV#	0.6039	0.001
MaYSV&	0.6613	0.001
ToSRV	0.1079	0.001

* Residual square sum for global test, which is inversely proportional to the level of global congruence between phylogenies.

Alignment includes recombinant blocks.

& Alignment does not include recombinant blocks.

Table 2. 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 componente					
	DNA-A*			DNA-B*		
	Cluster 1	Cluster 2	Nst#	Cluster 1	Cluster 2	Nst
BGMV	MW-2	MW-5	0.704	MW	MG-1	0.892
	MW-2	MG-1	0.922	MW	AL-2	0.868
	MW-2	AL-2	0.966	MW	AL-1	0.925
	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	0.911
	MG-1	AL-1	0.981	MG-1	AL-1	0.962
	AL-2	AL-1	0.695	AL-2	AL-1	0.680
EuYMV	GO	SO-1	0.487	GO	SO-2	0.363
	GO	SO-2	0.640	GO	SO-1	0.448
	SO-1	SO-2	0.442	SO-2	SO-1	0.215
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	0.316
	MG-3	MG-1	0.718	MG-3	MG-1	0.830
	MG-3	MG-2	0.623	MG-3	MG-2	0.778
	MG-4	MG-1	0.752	MG-4	MG-1	0.785
	MG-4	MG-2	0.703	MG-4	MG-2	0.709
	MG-1	MG-2	0.776	MG-1	MG-2	0.811

* Although there are subpopulations with the same names in the DNA-A and DNA-B data sets of some viral species, 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.

Table 3. 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)
			Start	End			
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:Sti3:11	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	$1e^{-61}/1e^{-23}$

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

Table 4. Reassortment events detected by RDP analysis of concatenated genomic components and proportion of reassortment sequences in 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) data sets.

Virus	Event*	Number of reassortment sequences by event/total number of sequences (%)	Number of reassortment sequences/total number of sequences (%)
BGMV	1	4/119 (3.36)	35/119 (29.41)
	2	31/119 (26.05)	
BlYSV	1	1/24 (4.16)	5/24 (20.83)
	2	1/24 (4.16)	
	3	1/24 (4.16)	
	4	2/24 (8.33)	
EuYMV	1	4/55 (7.28)	29/55 (52.72)
	2	1/55 (1.82)	
	3	3/55 (5.46)	
	4	2/55 (3.63)	
	5	5/55 (9.10)	
	6	2/55 (3.63)	
	7	3/55 (5.46)	
	8	1/55 (1.82)	
	9	1/55 (1.82)	
	10	1/55 (1.82)	
	11	5/55 (9.10)	
	12	1/55 (1.82)	
MaYSV	1	9/25 (36.00)	19/25 (76.00)
	2	10/25 (40.00)	
ToSRV	1	1/67 (1.49)	2/67 (2.98)
	2	1/67 (1.49)	

* For detailed information of the reassortment events see Suppl. Table S7.

Table 5. Non-synonymous to synonymous substitution ratios (dN/dS) and number of negatively and positively selected sites in each gene of the DNA-A and DNA-B segments 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).

Virus	Gene	dN/Ds	Negative selection				Positive selection			
			SLAC	FEL	REL	Total	SLAC	FEL	REL	Total[&]
BGMV	CP	0.1037	0	39	86	86	0	0	0	0
	Rep	0.2490	8	41	3	41	0	0	2	2
	TrAP	0.4007	0	6	4	10	0	0	0	0
	REn	0.4261	1	6	1	6	0	0	0	0
	AC4	2.5300	0	3	3	3	0	1	0	**1**
	MP	0.1183	20	67	-*	67	0	0	-*	0
	NSP	0.2290	14	48	4	48	0	0	6	6
BlYSV	CP	0.0282	23	56	96	96	0	0	0	0
	Rep	0.1837	21	59	27	67	0	2	5	5
	TrAP	0.4378	5	13	2	13	0	1	10	**10**
	REn	0.1902	7	15	33	33	0	0	0	0
	AC4	0.8575	1	4	12	12	0	0	0	0
	MP	0.0667	63	103	47	104	0	1	7	7
	NSP	0.1803	48	83	36	51	0	1	1	1
EuYMV	CP	0.0611	22	47	-[#]	47	0	0	0	0
	Rep	0.2589	21	36	14	36	1	3	14	14
	TrAP	0.6470	2	8	0	10	0	1	1	2
	REn	0.2710	2	11	26	26	0	0	0	0
	AC4	1.1500	1	5	8	8	0	2	0	**2**
	MP	0.0835	35	65	27	72	0	0	1	1
	NSP	0.2284	28	56	8	56	0	1	3	3
MaYSV	CP	0.0943	11	41	78	78	0	0	0	0
	Rep	0.2284	13	79	70	106	0	2	6	6
	TrAP	0.5240	2	7	4	11	0	1	0	1
	REn	0.3078	3	9	3	9	0	0	3	3
	AC4	1.4500	1	6	0	6	0	1	0	1
	MP	0.0952	16	41	10	41	0	1	4	4
	NSP	0.1983	11	38	8	38	0	2	12	12
ToSRV	CP	0.2610	2	13	0	13	0	0	6	6
	Rep	0.1644	8	38	0	38	0	0	6	6
	TrAP	0.5467	0	2	0	2	0	0	3	3
	REn	0.3366	1	7	1	7	0	0	1	1
	AC4	0.2377	0	7	12	12	0	0	0	0
	MP	0.1319	4	26	0	26	0	0	3	3
	NSP	0.2311	5	22	2	22	0	0	0	0
Total			399	1107	625	1301	1	20	94	87

* Analysis was not performed as the number of haplotypes was above the program's limit).

[#] No rates with dN>dS were inferred for this data set, suggesting that all sites are under negative selection.
[&] Numbers in bold indicate evidence of positive selection at P<0.1 according to PARRIS.

Figure legends

Figure 1. (**A**) Begomovirus sequences analyzed in this study. The map of Brazil (**B**) shows the origin of the sequences sampled in different regions of country. Circles indicate the location where samples were collected. Circle size represents the number of sequences (DNA-A and DNA-B) obtained from samples collected at each location. AL, Alagoas; BA, Bahia; DF, Distrito Federal; GO, Goiás; MG, Minas Gerais; MS, Mato Grosso do Sul; PE, Pernambuco; PR, Paraná; RS, Rio Grande do Sul.

Figure 2. Midpoint-rooted Bayesian phylogenetic trees based on full-length nucleotide sequences of the DNA-A and DNA-B segments of (**A**) Bean golden mosaic virus (BGMV), (**B**) Tomato severe rugose virus (ToSRV), (**C**) Euphorbia yellow mosaic virus (EuYMV), (**D**, **E**) Macroptilium yellow spot virus (MaYSV), and (**F**, **G**) Blainvillea yellow spot virus (BlYSV). **E**, **G.** Phylogenetic trees inferred using a data set in which recombinant blocks were excluded. Nodes with posterior probability values between 0.50 and 0.80 are indicated by empty circles and nodes with values equal to or greater than 0.81 are indicated by filled circles. The scale bar represents the number of nucleotide substitutions per site. Isolate color indicates sampling location.

Figure 3. Analysis of population structure using Discriminant Analysis of Principal Components (DAPC) for Bean golden mosaic virus (BGMV), Euphorbia yellow mosaic virus (EuYMV), Macroptilium yellow spot virus (MaYSV) and Tomato severe rugose virus (ToSRV) data sets. DAPC scatterplots (left panels) show the first two principal components with ellipses representing genetic clusters and dots representing each isolate. Isolates in the Bayesian phylogenetic trees (center panels) are colored based on genetic clusters identified in the DAPC analysis. Pie-charts in the maps (right panels) represent the genetic composition of subpopulations according to DAPC analysis. Pie-chart size represents the number of sequences obtained from samples collected at each location. Although there are genetics clusters with the same names in the DNA-A and DNA-B data sets of some viral species, the composition of isolates in these clusters is different.

Figure 4. Genetic variability based on full-length nucleotide sequences of the DNA-A and DNA-B segments 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). (**A**) Percentage of segregating sites (S) in DNA-A (black) and DNA-B (gray) data sets. (**B**-**D**) Statistical significance of the differences between the average pairwise number of nucleotide differences per site (nucleotide diversity - π) calculated for (**B**) the DNA-A and DNA-B of each virus, (**C**) the DNA-A of the different viruses and (**D)** the DNA-B of the different viruses. Ninety-five percent bootstrap confidence intervals (CIs) for the difference between π values were estimated from 1,000 nonparametric simulations. Confidence intervals which include the value "zero" denote no statistically significant difference between the means. (**E**) Average pairwise number of nucleotide differences per site (nucleotide diversity - π) calculated on a 100-nucleotide sliding window with a step size of 10 nucleotides across the full-length nucleotide sequences of the DNA-A and DNA-B. Horizontal dotted lines represent the average pairwise number of nucleotide differences per site. Genome maps for the DNA-A and DNA-B are presented at the bottom.

Figure 5. 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 Suppl. 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 gray). (**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 gray 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).

Figure 6. Evidence of reassortment events viewed through individual contributions of DNA-A-DNA-B links for the global congruence statistic. Estimated jackknife squared residuals (bars) and 95% confidence intervals (error bars) for 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). X-axis labels refer to isolate names for each pair of DNA-A and DNA-B segments. Horizontal dotted lines indicate the median squared residual values. Light gray bars correspond to sequences containing potential reassortment events detected by concatenated analysis in RDP. Box plots represent the jackknifed squared residuals grouped by sampling location. Ninety-five percent bootstrap confidence intervals (CIs) for the jackknifed squared residuals were estimated from 1,000 nonparametric simulations. Box plots with the same letter indicate that the confidence interval includes the value "zero", denoting no statistically significant difference between means. The horizontal line inside the box corresponds to the median. This analysis was performed using data sets without recombinant blocks.

Figure 7. Tanglegram plots depicting phylogenetic relationships between inferred Bayesian phylogenetic trees of the DNA-A and DNA-B 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). The association among cognate DNA-A and DNA-B segments is showed as connecting lines colored according to sampling location. For BGMV, Midwest represent isolates sampled in Paranoá, Santo Antônio de Goiás, Cristalina and Unaí. AL, Alagoas; BA, Bahia; GO, Goiás; MS, Mato Grosso do Sul; PE, Pernambuco; PR, Paraná; RS, Rio Grande do Sul.

Figure 8. Statistical significance of the difference between selection pressure parameters calculated for DNA-A and DNA-B data sets, considering data grouped by genomic segment and coding region. Box-plots correspond to: **(A)** non-synonymous to synonymous substitution ratios (dN/dS); **(B)** the number of negatively selected sites; **(C)** the number of positively selected sites. Ninety-five percent bootstrap confidence intervals (CIs) for the difference between mean dN/dS values, number of negatively and positively selected sites were estimated from 1,000 nonparametric simulations. Box plots with the same letter indicate that the confidence interval includes the value "zero", denoting no statistically significant difference between means. The horizontal line inside the box corresponds to the median.

Figure 1

A

Specie	State	Host	Number of sequences			Total number of sequences by species	
			DNA-A (retrieve from GenBank)	DNA-B			
				This study	GenBank	DNA-A	DNA-B
BGMV	Distrito Federal (DF)	*Phaseolus vulgaris*	30	30	-	117	123
	Goiás (GO)	*Phaseolus vulgaris*	29	27	-		
	Minas Gerais (MG)	*Phaseolus vulgaris*	16	19	-		
		Macroptilium lathyroides	13	13	-		
	Alagoas (AL)	*Phaseolus lunatus*	29	34	-		
BIYSV	Minas Gerais (MG)	*Blainvillea rhomboidea*	22	29	8	30	41
	Alagoas (AL)	*Blainvillea rhomboidea*	7	-	2		
	Bahia (BA)	*Blainvillea rhomboidea*	1	-	2		
EuYMV	Goiás (GO)	*Euphorbia heterophylla*	16	-	15	50	53
	Mato Grosso do Sul (MS)	*Euphorbia heterophylla*	2	-	2		
	Paraná (PR)	*Euphorbia heterophylla*	7	-	9		
	Rio Grande do Sul (RS)	*Euphorbia heterophylla*	25	-	27		
MaYSV	Alagoas (AL)	*Phaseolus vulgaris*	3	4	-	21	24
		Phaseolus lunatus	9	12	-		
		Macroptilium lathyroides	5	4	-		
	Pernambuco (PE)	*Desmodium glabrum*	4	-	4		
ToSRV	Minas Gerais (MG)	*Solanum lycopersicum*	74	67	7	74	74
Total			292	239	76	292	315

B

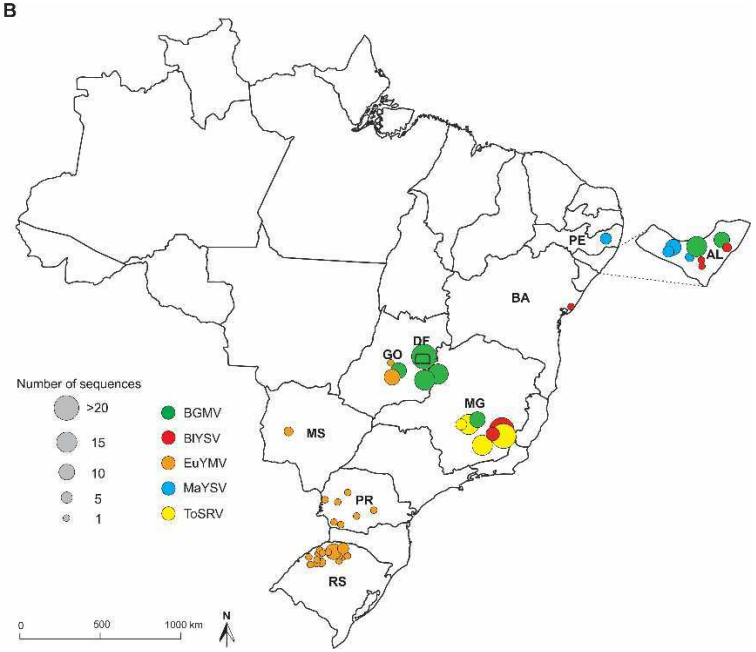


Figure 2A

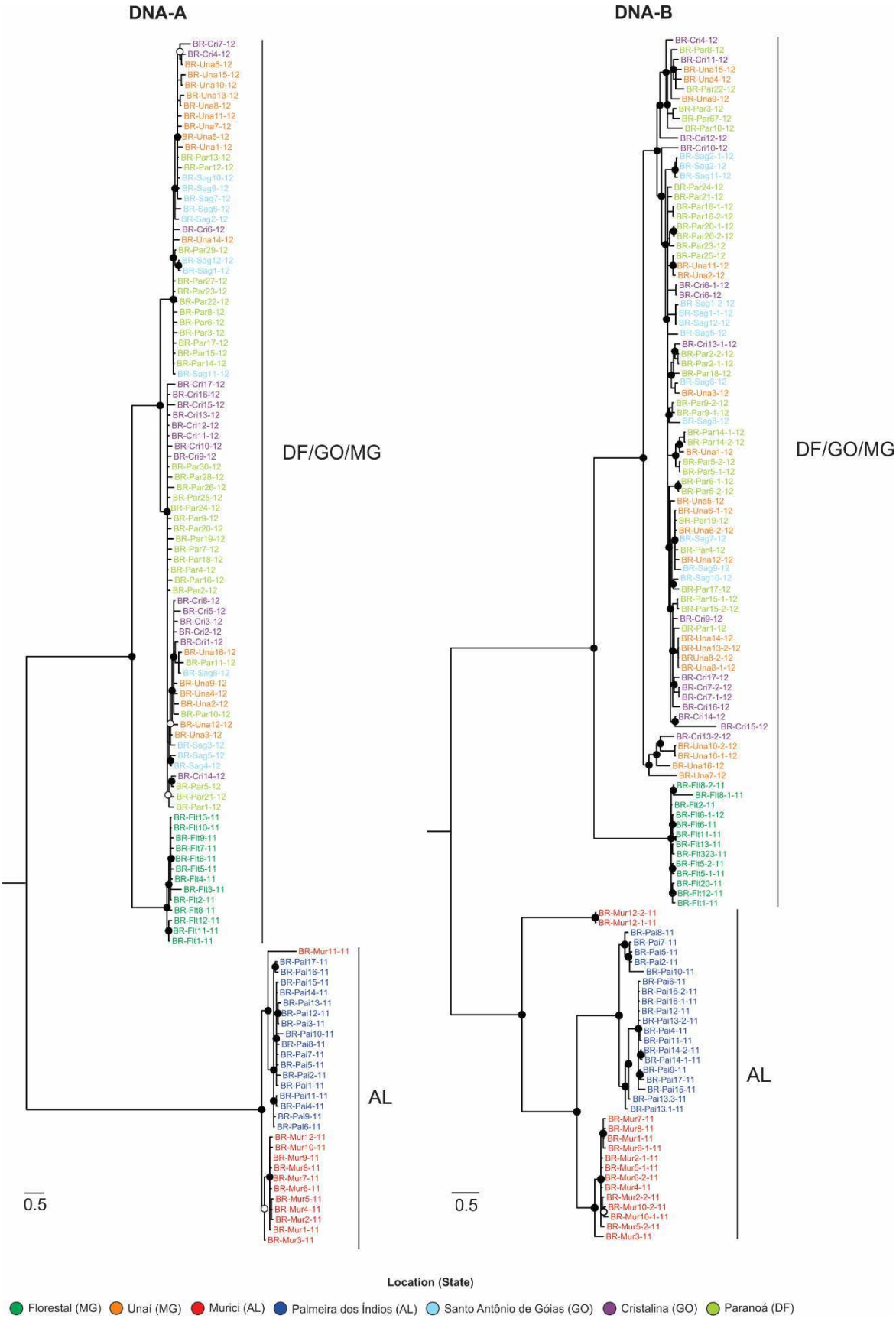

DNA-A
DNA-B
DF/GO/MG
AL
0.5
Location (State)
Florestal (MG)
Unaí (MG)
Murici (AL)
Palmeira dos Índios (AL)
Santo Antônio de Góias (GO)
Cristalina (GO)
Paranoá (DF)

Figure 2B

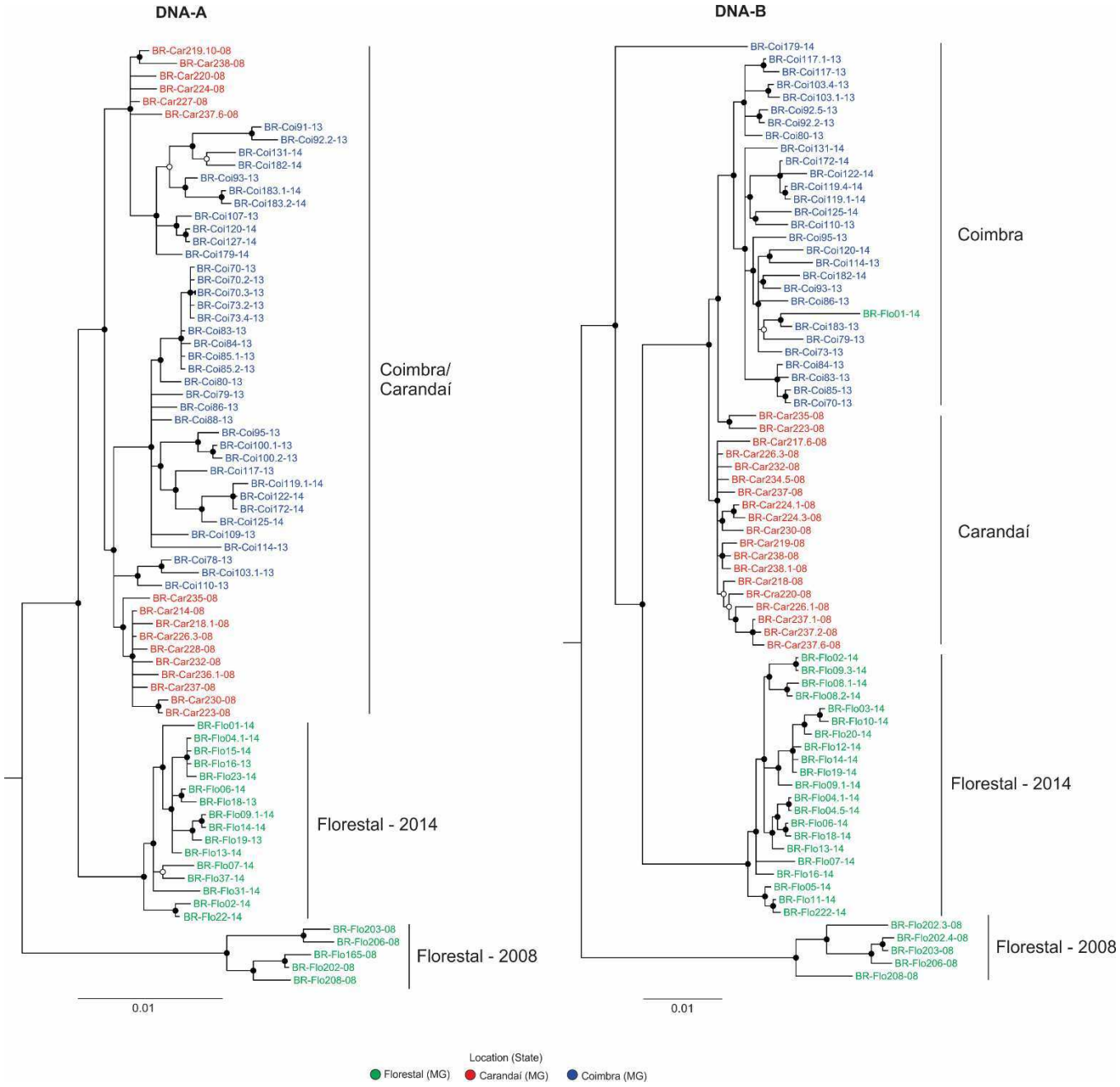

DNA-A
BR-Car219.10-08
BR-Car238-08
BR-Car220-08
BR-Car224-08
BR-Car227-08
BR-Car237.6-08
BR-Coi91-13
BR-Coi92.2-13
BR-Coi131-14
BR-Coi182-14
BR-Coi93-13
BR-Coi183.1-14
BR-Coi183.2-14
BR-Coi107-13
BR-Coi120-14
BR-Coi127-14
BR-Coi179-14
BR-Coi70-13
BR-Coi70.2-13
BR-Coi70.3-13
BR-Coi73.2-13
BR-Coi73.4-13
BR-Coi83-13
BR-Coi84-13
BR-Coi85.1-13
BR-Coi85.2-13
BR-Coi80-13
BR-Coi79-13
BR-Coi86-13
BR-Coi88-13
BR-Coi95-13
BR-Coi100.1-13
BR-Coi100.2-13
BR-Coi117-13
BR-Coi119.1-14
BR-Coi122-14
BR-Coi172-14
BR-Coi125-14
BR-Coi109-13
BR-Coi114-13
BR-Coi78-13
BR-Coi103.1-13
BR-Coi110-13
BR-Car235-08
BR-Car214-08
BR-Car218.1-08
BR-Car226.3-08
BR-Car228-08
BR-Car232-08
BR-Car236.1-08
BR-Car237-08
BR-Car230-08
BR-Car223-08
BR-Flo01-14
BR-Flo04.1-14
BR-Flo15-14
BR-Flo16-13
BR-Flo23-14
BR-Flo06-14
BR-Flo18-13
BR-Flo09.1-14
BR-Flo14-14
BR-Flo19-13
BR-Flo13-14
BR-Flo07-14
BR-Flo37-14
BR-Flo31-14
BR-Flo02-14
BR-Flo22-14
BR-Flo203-08
BR-Flo206-08
BR-Flo165-08
BR-Flo202-08
BR-Flo208-08
Coimbra/
Carandaí
Florestal - 2014
Florestal - 2008
0.01
DNA-B
BR-Coi179-14
BR-Coi117.1-13
BR-Coi117-13
BR-Coi103.4-13
BR-Coi103.1-13
BR-Coi92.5-13
BR-Coi92.2-13
BR-Coi80-13
BR-Coi131-14
BR-Coi172-14
BR-Coi122-14
BR-Coi119.4-14
BR-Coi119.1-14
BR-Coi125-14
BR-Coi110-13
BR-Coi95-13
BR-Coi120-14
BR-Coi114-13
BR-Coi182-14
BR-Coi93-13
BR-Coi86-13
BR-Flo01-14
BR-Coi183-13
BR-Coi79-13
BR-Coi73-13
BR-Coi84-13
BR-Coi83-13
BR-Coi85-13
BR-Coi70-13
BR-Car235-08
BR-Car223-08
BR-Car217.6-08
BR-Car226.3-08
BR-Car232-08
BR-Car234.5-08
BR-Car237-08
BR-Car224.1-08
BR-Car224.3-08
BR-Car230-08
BR-Car219-08
BR-Car238-08
BR-Car238.1-08
BR-Car218-08
BR-Cra220-08
BR-Car226.1-08
BR-Car237.1-08
BR-Car237.2-08
BR-Car237.6-08
BR-Flo02-14
BR-Flo09.3-14
BR-Flo08.1-14
BR-Flo08.2-14
BR-Flo03-14
BR-Flo10-14
BR-Flo20-14
BR-Flo12-14
BR-Flo14-14
BR-Flo19-14
BR-Flo09.1-14
BR-Flo04.1-14
BR-Flo04.5-14
BR-Flo06-14
BR-Flo18-14
BR-Flo13-14
BR-Flo07-14
BR-Flo16-14
BR-Flo05-14
BR-Flo11-14
BR-Flo222-14
BR-Flo202.3-08
BR-Flo202.4-08
BR-Flo203-08
BR-Flo206-08
BR-Flo208-08
Coimbra
Carandaí
Florestal - 2014
Florestal - 2008
0.01
Location (State)
Florestal (MG)
Carandaí (MG)
Coimbra (MG)

Figure 2C

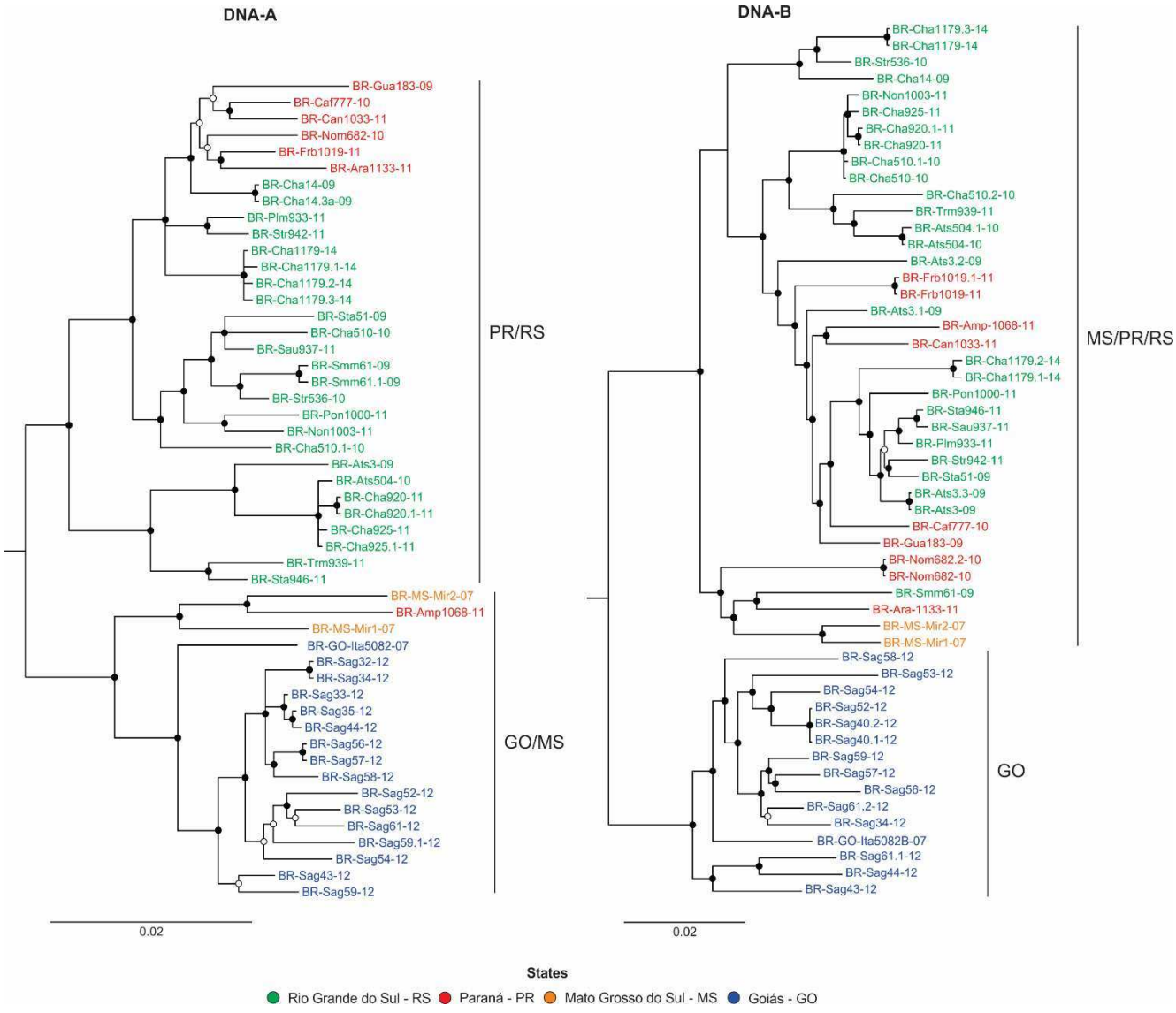

DNA-A
BR-Gua183-09
BR-Caf777-10
BR-Can1033-11
BR-Nom682-10
BR-Frb1019-11
BR-Ara1133-11
BR-Cha14-09
BR-Cha14.3a-09
BR-Plm933-11
BR-Str942-11
BR-Cha1179-14
BR-Cha1179.1-14
BR-Cha1179.2-14
BR-Cha1179.3-14
BR-Sta51-09
BR-Cha510-10
BR-Sau937-11
BR-Smm61-09
BR-Smm61.1-09
BR-Str536-10
BR-Pon1000-11
BR-Non1003-11
BR-Cha510.1-10
BR-Ats3-09
BR-Ats504-10
BR-Cha920-11
BR-Cha920.1-11
BR-Cha925-11
BR-Cha925.1-11
BR-Trm939-11
BR-Sta946-11
PR/RS
BR-MS-Mir2-07
BR-Amp1068-11
BR-MS-Mir1-07
BR-GO-Ita5082-07
BR-Sag32-12
BR-Sag34-12
BR-Sag33-12
BR-Sag35-12
BR-Sag44-12
BR-Sag56-12
BR-Sag57-12
BR-Sag58-12
BR-Sag52-12
BR-Sag53-12
BR-Sag61-12
BR-Sag59.1-12
BR-Sag54-12
BR-Sag43-12
BR-Sag59-12
GO/MS
0.02
DNA-B
BR-Cha1179.3-14
BR-Cha1179-14
BR-Str536-10
BR-Cha14-09
BR-Non1003-11
BR-Cha925-11
BR-Cha920.1-11
BR-Cha920-11
BR-Cha510.1-10
BR-Cha510-10
BR-Cha510.2-10
BR-Trm939-11
BR-Ats504.1-10
BR-Ats504-10
BR-Ats3.2-09
BR-Frb1019.1-11
BR-Frb1019-11
BR-Ats3.1-09
BR-Amp-1068-11
BR-Can1033-11
BR-Cha1179.2-14
BR-Cha1179.1-14
BR-Pon1000-11
BR-Sta946-11
BR-Sau937-11
BR-Plm933-11
BR-Str942-11
BR-Sta51-09
BR-Ats3.3-09
BR-Ats3-09
BR-Caf777-10
BR-Gua183-09
BR-Nom682.2-10
BR-Nom682-10
BR-Smm61-09
BR-Ara-1133-11
BR-MS-Mir2-07
BR-MS-Mir1-07
MS/PR/RS
BR-Sag58-12
BR-Sag53-12
BR-Sag54-12
BR-Sag52-12
BR-Sag40.2-12
BR-Sag40.1-12
BR-Sag59-12
BR-Sag57-12
BR-Sag56-12
BR-Sag61.2-12
BR-Sag34-12
BR-GO-Ita5082B-07
BR-Sag61.1-12
BR-Sag44-12
BR-Sag43-12
GO
0.02
States
Rio Grande do Sul - RS
Paraná - PR
Mato Grosso do Sul - MS
Goiás - GO

Figure 2D,E

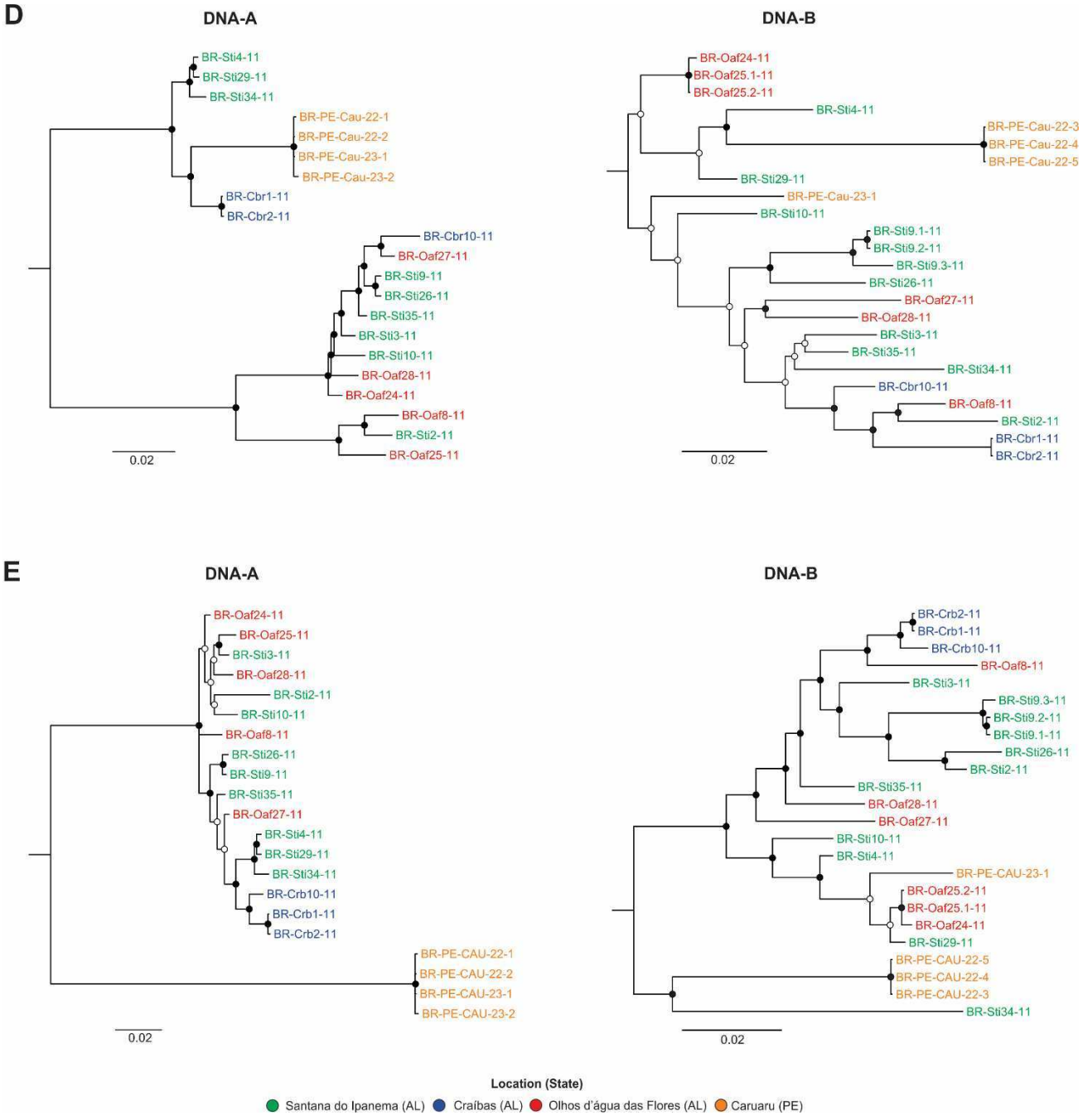

D
DNA-A
BR-Sti4-11
BR-Sti29-11
BR-Sti34-11
BR-PE-Cau-22-1
BR-PE-Cau-22-2
BR-PE-Cau-23-1
BR-PE-Cau-23-2
BR-Cbr1-11
BR-Cbr2-11
BR-Cbr10-11
BR-Oaf27-11
BR-Sti9-11
BR-Sti26-11
BR-Sti35-11
BR-Sti3-11
BR-Sti10-11
BR-Oaf28-11
BR-Oaf24-11
BR-Oaf8-11
BR-Sti2-11
BR-Oaf25-11
0.02
DNA-B
BR-Oaf24-11
BR-Oaf25.1-11
BR-Oaf25.2-11
BR-Sti4-11
BR-PE-Cau-22-3
BR-PE-Cau-22-4
BR-PE-Cau-22-5
BR-Sti29-11
BR-PE-Cau-23-1
BR-Sti10-11
BR-Sti9.1-11
BR-Sti9.2-11
BR-Sti9.3-11
BR-Sti26-11
BR-Oaf27-11
BR-Oaf28-11
BR-Sti3-11
BR-Sti35-11
BR-Sti34-11
BR-Cbr10-11
BR-Oaf8-11
BR-Sti2-11
BR-Cbr1-11
BR-Cbr2-11
0.02
E
DNA-A
BR-Oaf24-11
BR-Oaf25-11
BR-Sti3-11
BR-Oaf28-11
BR-Sti2-11
BR-Sti10-11
BR-Oaf8-11
BR-Sti26-11
BR-Sti9-11
BR-Sti35-11
BR-Oaf27-11
BR-Sti4-11
BR-Sti29-11
BR-Sti34-11
BR-Crb10-11
BR-Crb1-11
BR-Crb2-11
BR-PE-CAU-22-1
BR-PE-CAU-22-2
BR-PE-CAU-23-1
BR-PE-CAU-23-2
0.02
DNA-B
BR-Crb2-11
BR-Crb1-11
BR-Crb10-11
BR-Oaf8-11
BR-Sti3-11
BR-Sti9.3-11
BR-Sti9.2-11
BR-Sti9.1-11
BR-Sti26-11
BR-Sti2-11
BR-Sti35-11
BR-Oaf28-11
BR-Oaf27-11
BR-Sti10-11
BR-Sti4-11
BR-PE-CAU-23-1
BR-Oaf25.2-11
BR-Oaf25.1-11
BR-Oaf24-11
BR-Sti29-11
BR-PE-CAU-22-5
BR-PE-CAU-22-4
BR-PE-CAU-22-3
BR-Sti34-11
0.02
Location (State)
Santana do Ipanema (AL)
Craíbas (AL)
Olhos d'água das Flores (AL)
Caruaru (PE)

Figure 2F,G

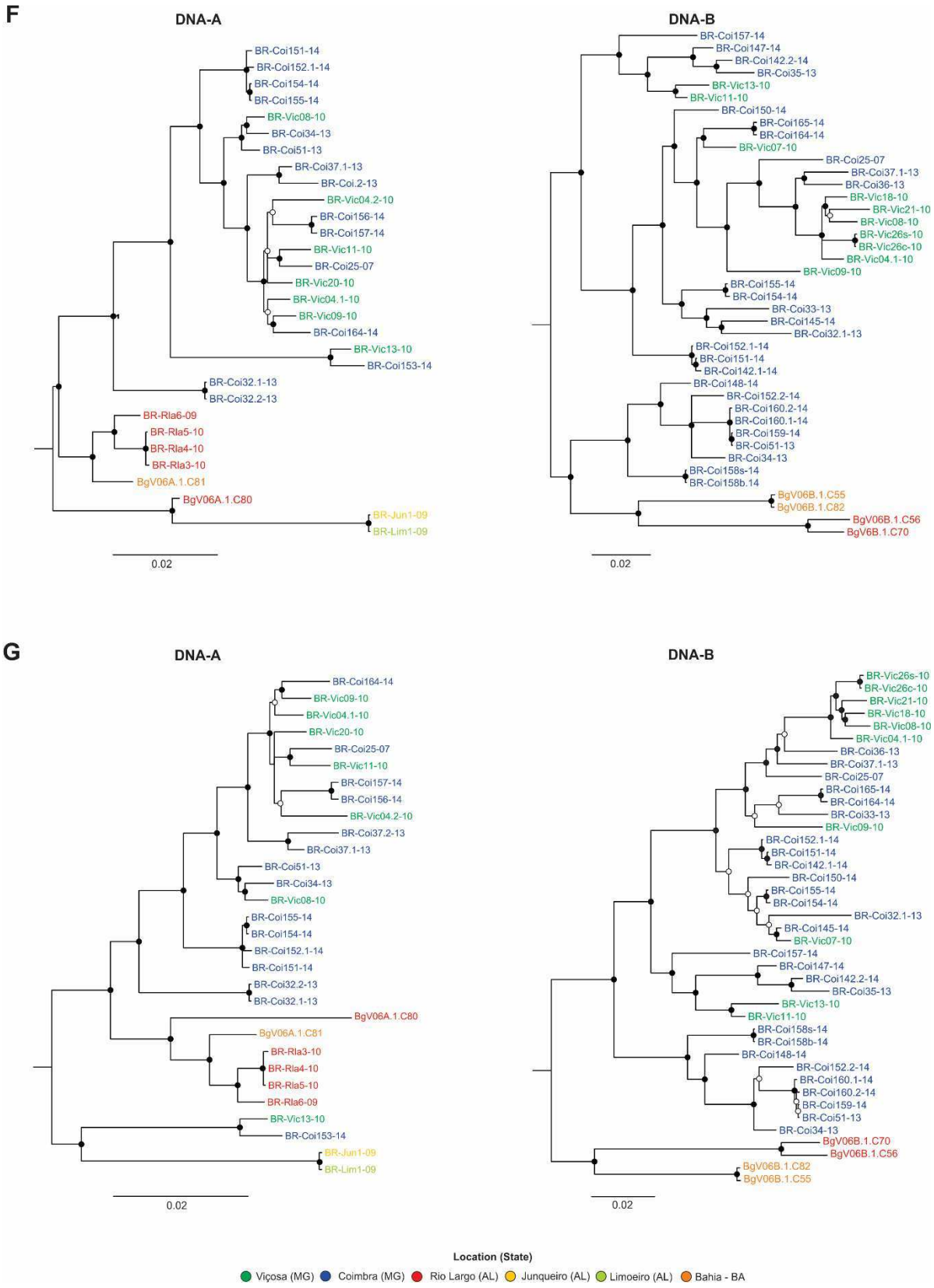

F
DNA-A
BR-Coi151-14
BR-Coi152.1-14
BR-Coi154-14
BR-Coi155-14
BR-Vic08-10
BR-Coi34-13
BR-Coi51-13
BR-Coi37.1-13
BR-Coi.2-13
BR-Vic04.2-10
BR-Coi156-14
BR-Coi157-14
BR-Vic11-10
BR-Coi25-07
BR-Vic20-10
BR-Vic04.1-10
BR-Vic09-10
BR-Coi164-14
BR-Vic13-10
BR-Coi153-14
BR-Coi32.1-13
BR-Coi32.2-13
BR-Rla6-09
BR-Rla5-10
BR-Rla4-10
BR-Rla3-10
BgV06A.1.C81
BgV06A.1.C80
BR-Jun1-09
BR-Lim1-09
0.02
DNA-B
BR-Coi157-14
BR-Coi147-14
BR-Coi142.2-14
BR-Coi35-13
BR-Vic13-10
BR-Vic11-10
BR-Coi150-14
BR-Coi165-14
BR-Coi164-14
BR-Vic07-10
BR-Coi25-07
BR-Coi37.1-13
BR-Coi36-13
BR-Vic18-10
BR-Vic21-10
BR-Vic08-10
BR-Vic26s-10
BR-Vic26c-10
BR-Vic04.1-10
BR-Vic09-10
BR-Coi155-14
BR-Coi154-14
BR-Coi33-13
BR-Coi145-14
BR-Coi32.1-13
BR-Coi152.1-14
BR-Coi151-14
BR-Coi142.1-14
BR-Coi148-14
BR-Coi152.2-14
BR-Coi160.2-14
BR-Coi160.1-14
BR-Coi159-14
BR-Coi51-13
BR-Coi34-13
BR-Coi158s-14
BR-Coi158b.14
BgV06B.1.C55
BgV06B.1.C82
BgV06B.1.C56
BgV6B.1.C70
0.02
G
DNA-A
BR-Coi164-14
BR-Vic09-10
BR-Vic04.1-10
BR-Vic20-10
BR-Coi25-07
BR-Vic11-10
BR-Coi157-14
BR-Coi156-14
BR-Vic04.2-10
BR-Coi37.2-13
BR-Coi37.1-13
BR-Coi51-13
BR-Coi34-13
BR-Vic08-10
BR-Coi155-14
BR-Coi154-14
BR-Coi152.1-14
BR-Coi151-14
BR-Coi32.2-13
BR-Coi32.1-13
BgV06A.1.C80
BgV06A.1.C81
BR-Rla3-10
BR-Rla4-10
BR-Rla5-10
BR-Rla6-09
BR-Vic13-10
BR-Coi153-14
BR-Jun1-09
BR-Lim1-09
0.02
DNA-B
BR-Vic26s-10
BR-Vic26c-10
BR-Vic21-10
BR-Vic18-10
BR-Vic08-10
BR-Vic04.1-10
BR-Coi36-13
BR-Coi37.1-13
BR-Coi25-07
BR-Coi165-14
BR-Coi164-14
BR-Coi33-13
BR-Vic09-10
BR-Coi152.1-14
BR-Coi151-14
BR-Coi142.1-14
BR-Coi150-14
BR-Coi155-14
BR-Coi154-14
BR-Coi32.1-13
BR-Coi145-14
BR-Vic07-10
BR-Coi157-14
BR-Coi147-14
BR-Coi142.2-14
BR-Coi35-13
BR-Vic13-10
BR-Vic11-10
BR-Coi158s-14
BR-Coi158b-14
BR-Coi148-14
BR-Coi152.2-14
BR-Coi160.1-14
BR-Coi160.2-14
BR-Coi159-14
BR-Coi51-13
BR-Coi34-13
BgV06B.1.C70
BgV06B.1.C56
BgV06B.1.C82
BgV06B.1.C55
0.02
Location (State)
Viçosa (MG)
Coimbra (MG)
Rio Largo (AL)
Junqueiro (AL)
Limoeiro (AL)
Bahia - BA

Figure 3

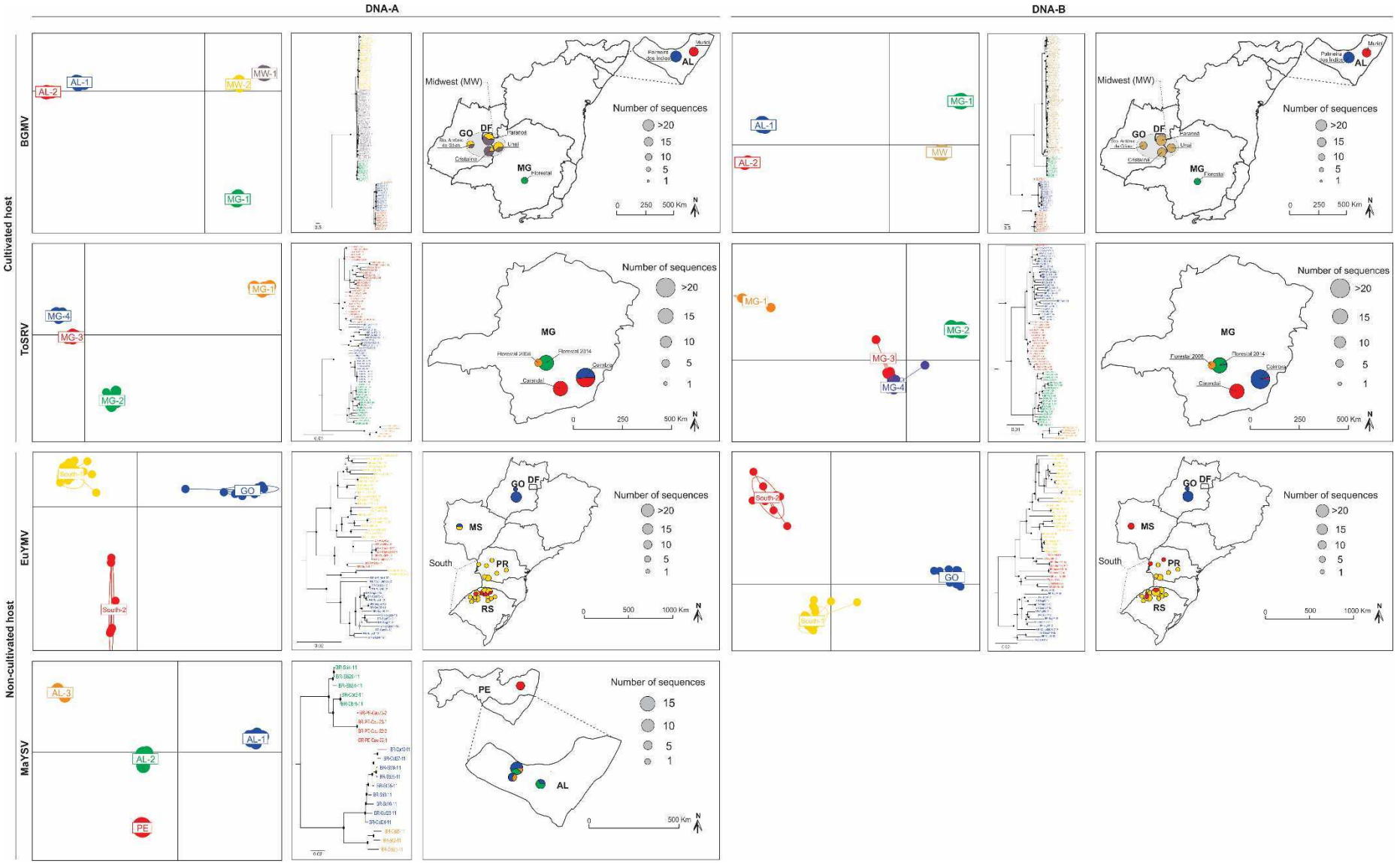

DNA-A
DNA-B
Cultivated host
Non-cultivated host
BGMV
ToSRV
EuYMV
MaYSV
AL-1
AL-2
MW-1
MW-2
MG-1
MG-2
MG-3
MG-4
South-1
South-2
GO
AL-3
PE
Midwest (MW)
Number of sequences
>20
15
10
5
1
0 250 500 Km
MG
Florestal 2008
Florestal 2014
Corrêntina
Carandaí
0 500 1000 Km
South
MS
PR
RS
DF
0 500 Km

Figure 4

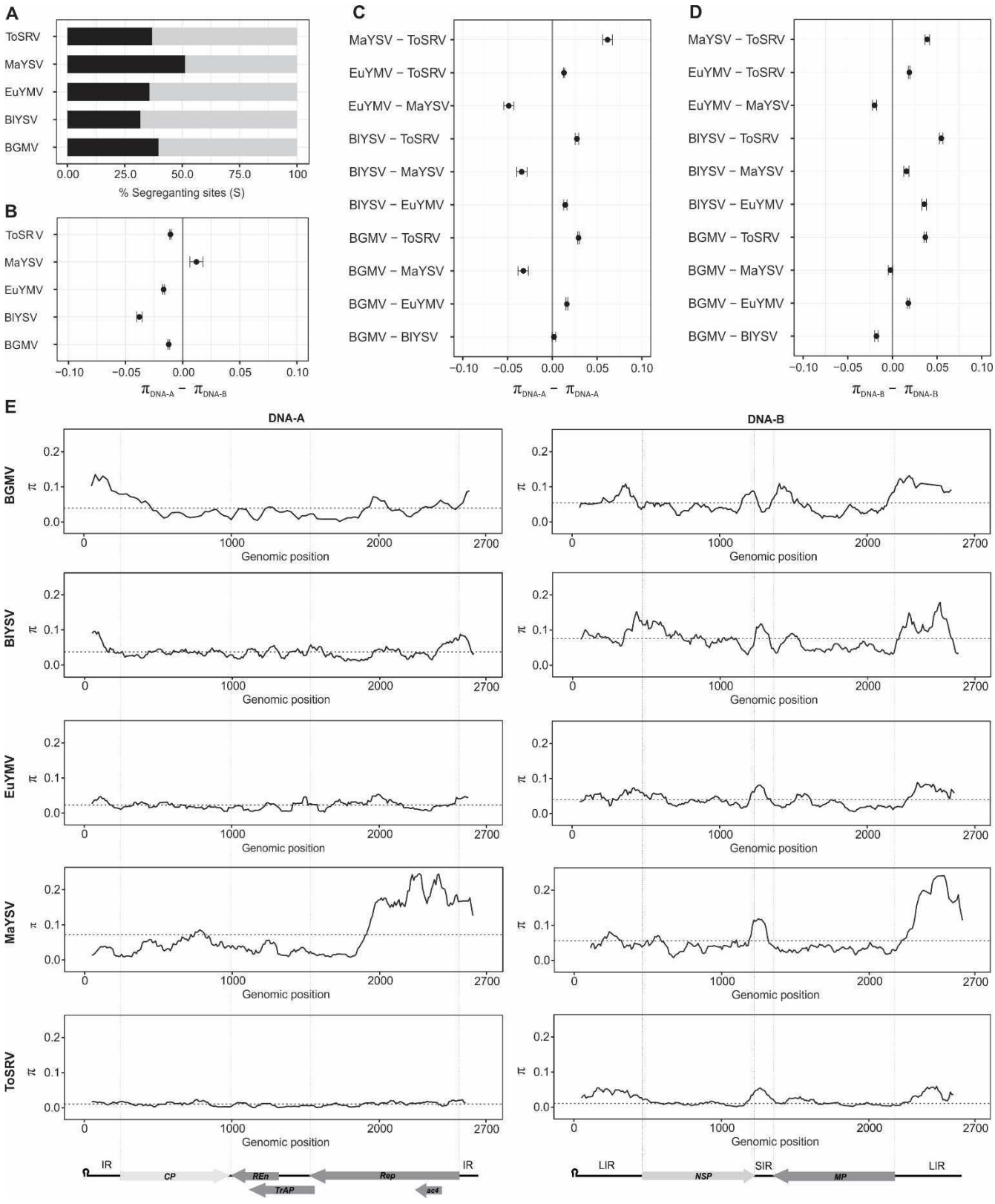

A
ToSRV
MaYSV
EuYMV
BlYSV
BGMV
0.00
25.0
50.0
75.0
100
% Segreganting sites (S)
B
ToSRV
MaYSV
EuYMV
BlYSV
BGMV
−0.10
−0.05
0.00
0.05
0.10
$\pi_{DNA-A} - \pi_{DNA-B}$
C
MaYSV − ToSRV
EuYMV − ToSRV
EuYMV − MaYSV
BlYSV − ToSRV
BlYSV − MaYSV
BlYSV − EuYMV
BGMV − ToSRV
BGMV − MaYSV
BGMV − EuYMV
BGMV − BlYSV
$\pi_{DNA-A} - \pi_{DNA-A}$
D
$\pi_{DNA-B} - \pi_{DNA-B}$
E
DNA-A
DNA-B
BGMV
BlYSV
EuYMV
MaYSV
ToSRV
π
Genomic position
IR
CP
REn
Rep
TrAP
ac4
LIR
NSP
SIR
MP

Figure 5

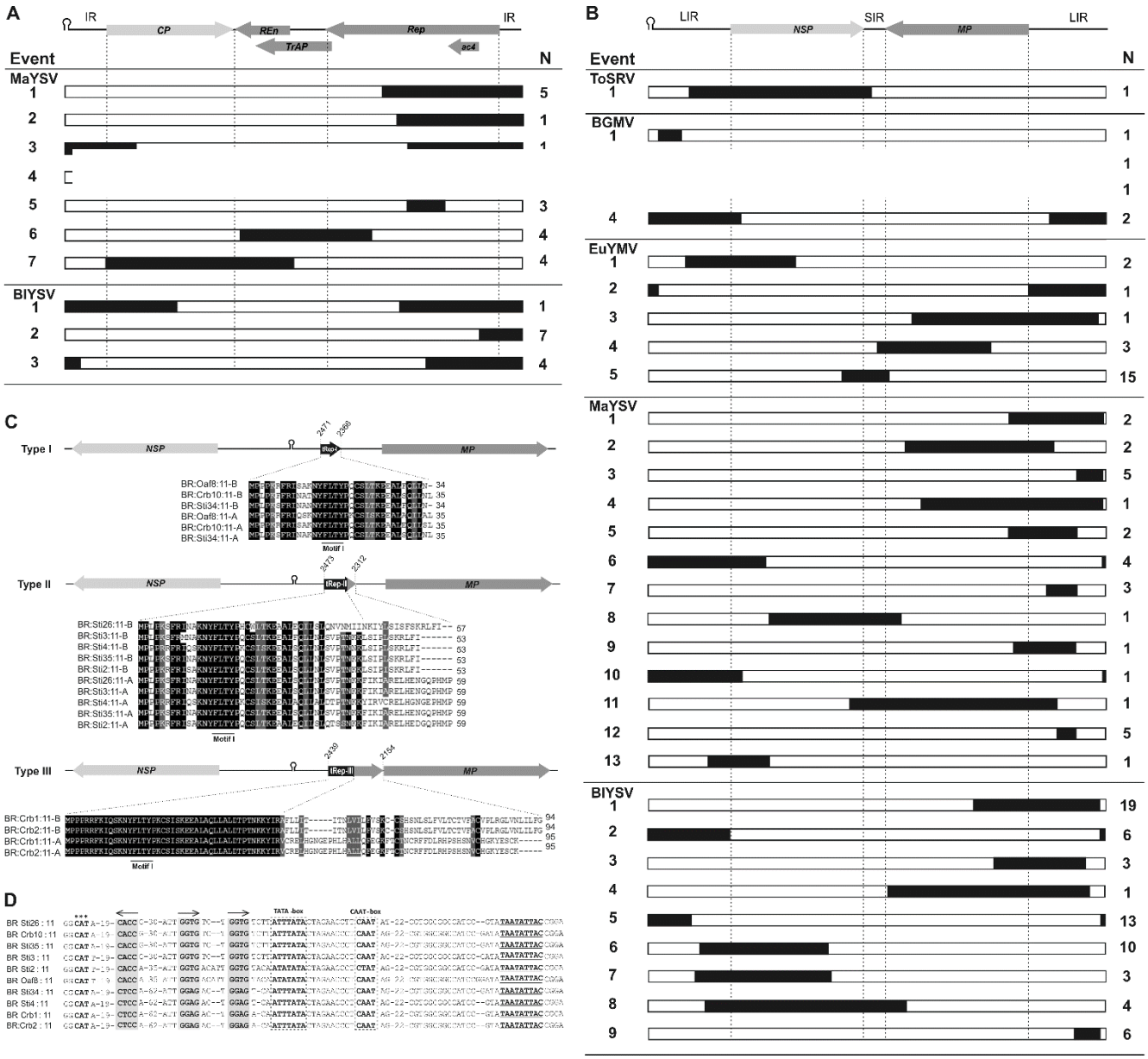

A
IR
CP
REn
Rep
IR
TrAP
ac4
Event
N
MaYSV
1
2
3
4
5
6
7
BlYSV
1
2
3
B
LIR
NSP
SIR
MP
LIR
Event
N
ToSRV
1
BGMV
1
EuYMV
1
2
3
4
5
MaYSV
1
2
3
4
5
6
7
8
9
10
11
12
13
BlYSV
1
2
3
4
5
6
7
8
9
C
Type I
NSP
MP
iRep-I
Type II
NSP
MP
iRep-II
Type III
NSP
MP
iRep-II
Motif I
D
TATA-box
CAAT-box

Figure 6

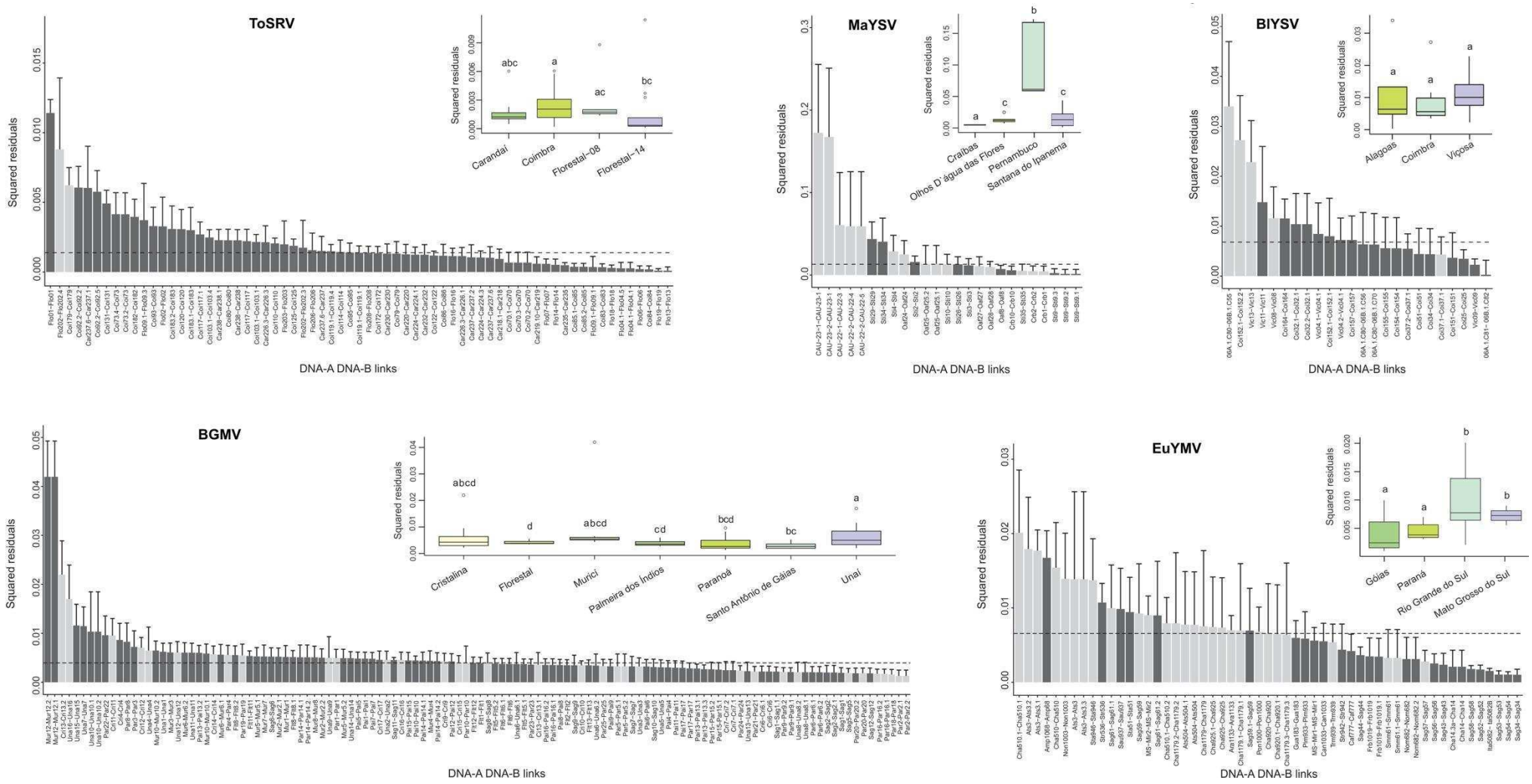

ToSRV
Squared residuals
DNA-A DNA-B links
abc
a
ac
bc
Carandaí
Coimbra
Florestal−08
Florestal−14
MaYSV
Squared residuals
DNA-A DNA-B links
a
c
b
c
Craíbas
Olhos D´água das Flores
Pernambuco
Santana do Ipanema
BIYSV
Squared residuals
DNA-A DNA-B links
a
a
a
Alagoas
Coimbra
Viçosa
BGMV
Squared residuals
DNA-A DNA-B links
abcd
d
abcd
cd
bcd
bc
a
Cristalina
Florestal
Murici
Palmeira dos Índios
Paranoá
Santo Antônio de Goiás
Unaí
EuYMV
Squared residuals
DNA-A DNA-B links
a
a
b
b
Góias
Paraná
Rio Grande do Sul
Mato Grosso do Sul

Figure 7

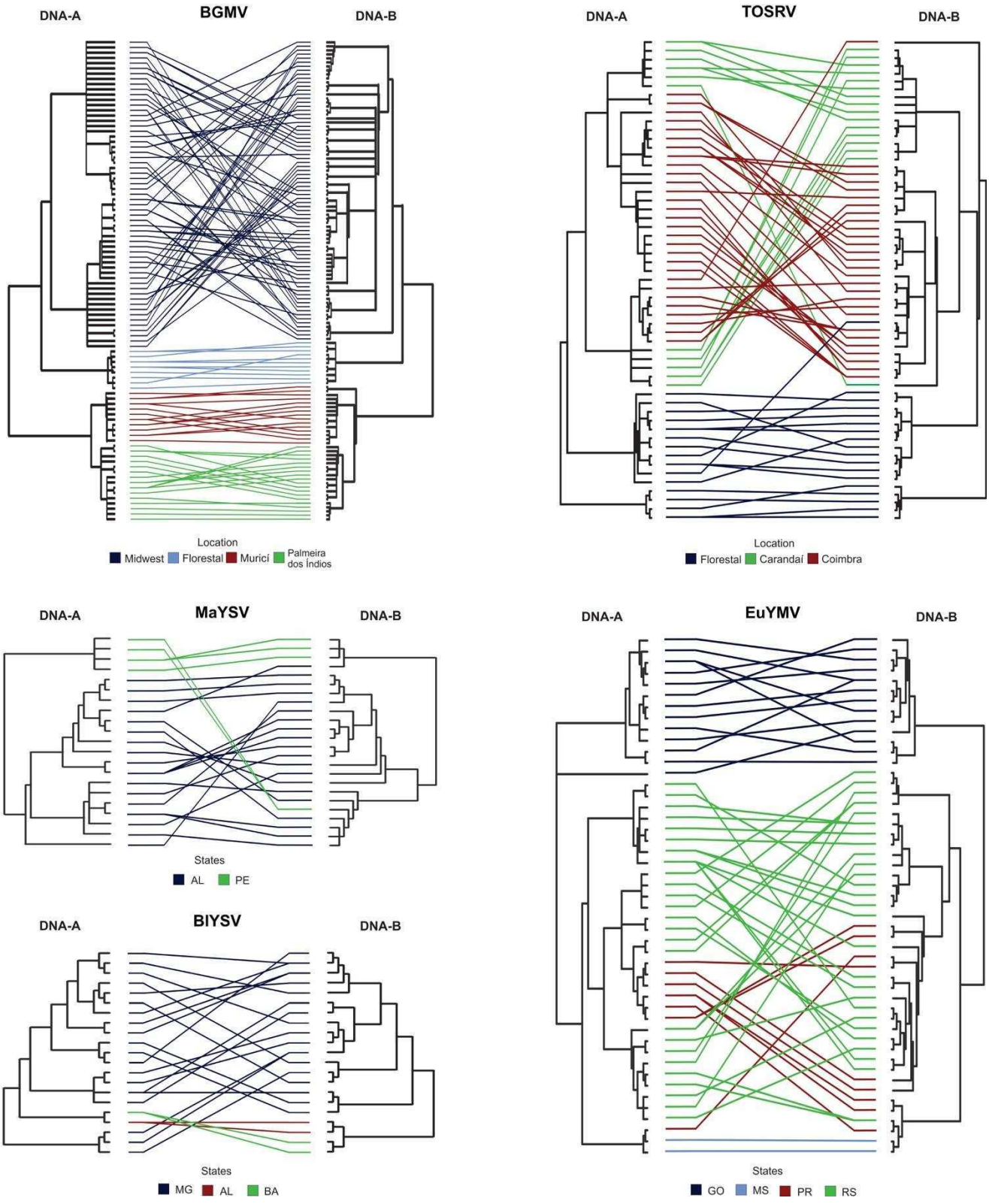
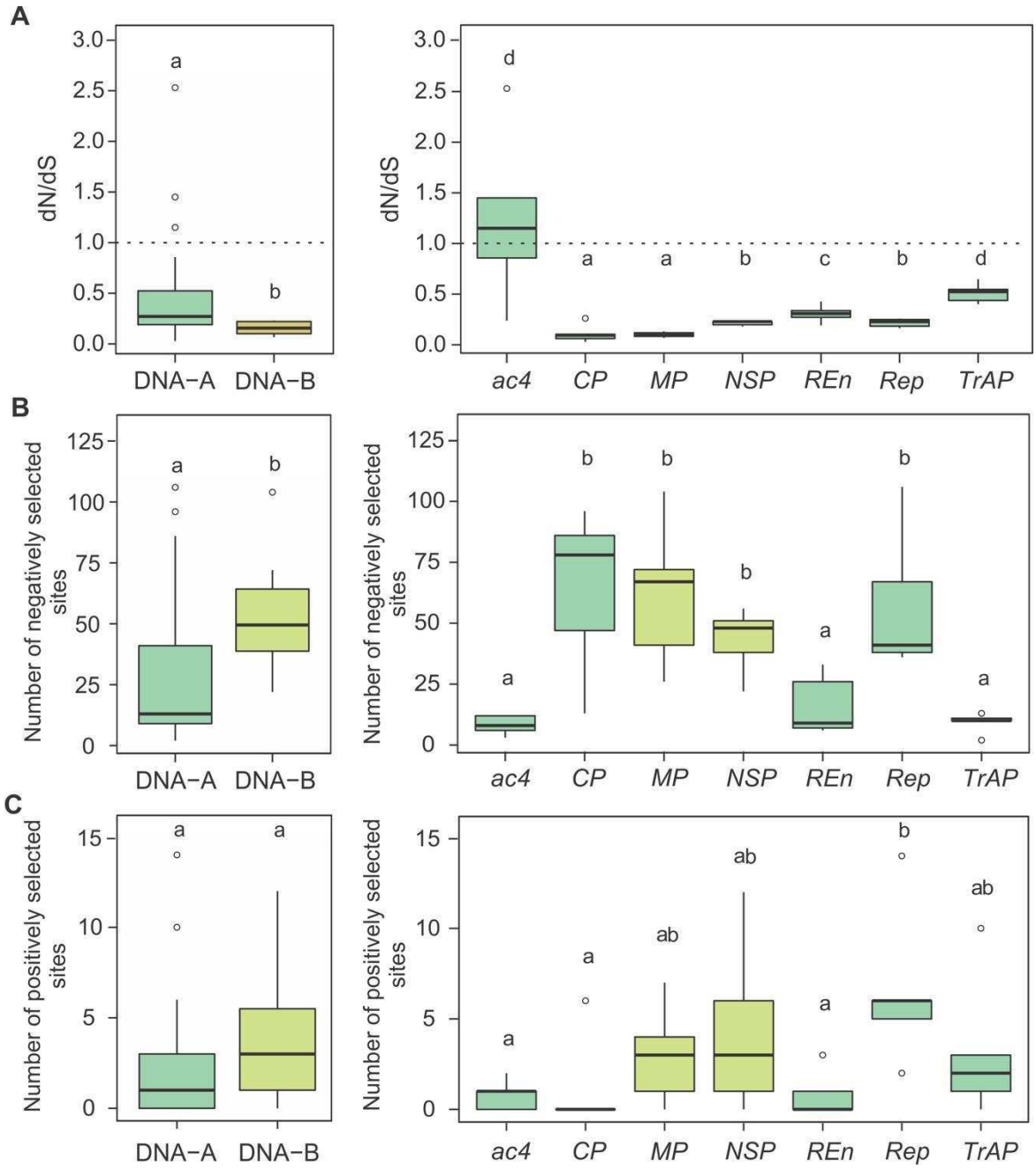

DNA-A
BGMV
DNA-B
Location
Midwest
Florestal
Muricí
Palmeira dos Índios
DNA-A
TOSRV
DNA-B
Location
Florestal
Carandaí
Coimbra
DNA-A
MaYSV
DNA-B
States
AL
PE
DNA-A
BIYSV
DNA-B
States
MG
AL
BA
DNA-A
EuYMV
DNA-B
States
GO
MS
PR
RS

Figure 8


A
dN/dS
3.0
2.5
2.0
1.5
1.0
0.5
0.0
a
b
DNA−A
DNA−B
d
a
a
b
c
b
d
ac4
CP
MP
NSP
REn
Rep
TrAP
B
Number of negatively selected sites
125
100
75
50
25
0
a
b
b
b
a
b
a
C
Number of positively selected sites
15
10
5
0
a
a
ab
b
ab

Supplementary Table S1. Begomovirus isolates obtained in this study.

Sample code	Date of collection	Location	Geographical coordinates		Host	Enzyme	Isolate name	GenBank access number
Blainvillea yellow spot virus (BlYSV)								
HV4	May, 2010	Viçosa, MG*	S20°45'22.9"	W42°50'58.0"	Blainvillea rhomboidea	SacI	BR:Vic04.1:10	
HV9	May, 2010	Viçosa, MG	S20°45'22.9"	W42°50'58.0"	Blainvillea rhomboidea	SacI	BR:Vic09:10	
HV13	May, 2010	Viçosa, MG	S20°45'22.9"	W42°50'58.0"	Blainvillea rhomboidea	SacI	BR:Vic13:10	
OS32	July, 2013	Coimbra, MG	S20°51'29.9"	W042°51'42.2"	Blainvillea rhomboidea	ClaI	BR:Coi32.1:13	
OS33	July, 2013	Coimbra, MG	S20°51'29.9"	W042°51'42.2"	Blainvillea rhomboidea	SacI	BR:Coi33:13	
OS34	July, 2013	Coimbra, MG	S20°51'29.9"	W042°51'42.2"	Blainvillea rhomboidea	SacI	BR:Coi34:13	
OS35	July, 2013	Coimbra, MG	S20°51'29.9"	W042°51'42.2"	Blainvillea rhomboidea	SacI	BR:Coi35:13	
OS36	July, 2013	Coimbra, MG	S20°51'29.9"	W042°51'42.2"	Blainvillea rhomboidea	SacI	BR:Coi36:13	
OS37	July, 2013	Coimbra, MG	S20°51'29.9"	W042°51'42.2"	Blainvillea rhomboidea	XbaI	BR:Coi37.1:13	
OS51	July, 2013	Coimbra, MG	S20°51'29.9"	W042°51'42.2"	Blainvillea rhomboidea	SacI	BR:Coi51:13	
OS142	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi142.1:14	
						SacI	BR:Coi142.2:14	
OS145	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi145:14	
OS147	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi147:14	
OS148	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi148:14	
OS150	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi150:14	
OS151	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi151:14	
OS152	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi152.1:14	
						SacI	BR:Coi152.2:14	
OS154	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi154:14	
OS155	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi155:14	
OS157	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi157:14	
OS158	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi158s:14	
						BamHI	BR:Coi158b:14	
OS159	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi159:14	
OS160	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi160.1:14	
						SacI	BR:Coi160.2:14	
OS164	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi164:14	
OS165	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Blainvillea rhomboidea	SacI	BR:Coi165:14	
Macroptilium yellow spot virus (MaYSV)								
RC18	July, 2011	Craíbas, AL	S09°40'37.9"	W036°46'37.8"	Phaseolus vulgaris	PstI	BR:Crb1:11	
RC19	July, 2011	Craíbas, AL	S09°40'39.1"	W036°46'38.4"	Phaseolus vulgaris	PstI	BR:Crb2:11	

RC29	July, 2011	Craíbas, AL	S09°40'39.8"	W036°46'38.4"	Phaseolus vulgaris	PstI	BR:Crb10:11
RC77	July, 2011	Olho D'água das Flores, AL	S09°32'28.1"	W037°17'26.8"	Phaseolus vulgaris	PstI	BR:Oaf8:11
46AL	July, 2011	Olho D'água das Flores, AL	S09°32'27.8"	W037°17'22.1"	Macroptilium lathyroides	PstI	BR:Oaf28:11
61AL	July, 2011	Olho D'água das Flores, AL	S09°32'26.9"	W037°17'17.6"	Macroptilium lathyroides	PstI	BR:Oaf27:11
RC94	July, 2011	Olho D'água das Flores, AL	S09°32'59.4"	W037°18'43.5"	Macroptilium lathyroides	PstI	BR:Oaf24:11
RC95	July, 2011	Olho D'água das Flores, AL	S09°32'59.2"	W037°18'43.0"	Macroptilium lathyroides	PstI	BR:Oaf25.1:11
						PstI	BR:Oaf25.2:11
97AL	July, 2011	Santana do Ipanema, AL	S09°23'25.1"	W037°12'46.9"	Phaseolus lunatus	PstI	BR:Sti34:11
99AL	July, 2011	Santana do Ipanema, AL	S09°23'24.7"	W037°12'48.0"	Phaseolus lunatus	BglII	BR:Sti2:11
100AL	July, 2011	Santana do Ipanema, AL	S09°23'24.8"	W037°12'48.0"	Phaseolus lunatus	SpeI	BR:Sti3:11
102AL	July, 2011	Santana do Ipanema, AL	S09°23'24.9"	W037°12'48.0"	Phaseolus lunatus	BglII	BR:Sti4:11
105AL	July, 2011	Santana do Ipanema, AL	S09°23'24.9"	W037°12'48.3"	Phaseolus lunatus	PstI	BR:Sti35:11
111AL	July, 2011	Santana do Ipanema, AL	S09°23'24.9"	W037°12'47.5"	Phaseolus lunatus	BglII	BR:Sti9.1:11
						BglII	BR:Sti9.2:11
						SpeI	BR:Sti9.3:11
112AL	July, 2011	Santana do Ipanema, AL	S09°23'24.9"	W037°12'47.6"	Phaseolus lunatus	PstI	BR:Sti10:11
128AL	July, 2012	Santana do Ipanema, AL	S09°23'24.9"	W037°12'47.6"	Phaseolus lunatus	ApaI	BR:Sti26:11
152AL	July, 2011	Santana do Ipanema, AL	S09°23'24.8"	W037°12'48.6"	Phaseolus lunatus	XbaI	BR:Sti29:11
Bean golden mosaic virus (BGMV)							
MG-146	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una1:12
MG-147	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una2:12
MG-148	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una3:12
MG-149	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una4:12
MG-150	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una5:12
MG-152	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una6.1:12
						SpeI	BR:Una6.2:12
MG-153	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una7:12
MG-154	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una8.1:12
						SpeI	BRUna8.2:12
MG-155	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una9:12
MG-156	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una10.1:12
						SpeI	BR:Una10.2:12
MG-157	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una11:12
MG-158	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una12:12
MG-161	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una13:12
MG-162	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una14:12

MG-163	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una15:12
MG-165	June, 2012	Unaí, MG	S16°43'02.2"	W046°45'14"	Phaseolus vulgaris	SpeI	BR:Una16:12
GO-2	June, 2012	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	Phaseolus vulgaris	SpeI	BR:Sag11:12
GO-3	June, 2012	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	Phaseolus vulgaris	SpeI	BR:Sag1.1:12
						SpeI	BR:Sag1.2:12
GO-4	June, 2012	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	Phaseolus vulgaris	SpeI	BR:Sag2:12
						SpeI	BR:Sag2.1:12
GO-9	June, 2012	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	Phaseolus vulgaris	SpeI	BR:Sag7:12
GO-15	June, 2012	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	Phaseolus vulgaris	SpeI	BR:Sag5:12
GO-22	June, 2012	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	Phaseolus vulgaris	SpeI	BR:Sag6:12
GO-25	June, 2012	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	Phaseolus vulgaris	SpeI	BR:Sag12:12
GO-26	June, 2013	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	Phaseolus vulgaris	SpeI	BR:Sag8:12
GO-27	June, 2012	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	Phaseolus vulgaris	SpeI	BR:Sag9:12
GO-29	June, 2012	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	Phaseolus vulgaris	SpeI	BR:Sag10:12
GO-63	June, 2012	Paranoá, DF	S16°00'47"	W047°33'21.5"	Phaseolus vulgaris	SpeI	BR:Par14.1:12
						HindIII	BR:Par14.2:12
GO-64	June, 2012	Paranoá, DF	S16°00'47"	W047°33'21.5"	Phaseolus vulgaris	SpeI	BR:Par1:12
GO-65	June, 2012	Paranoá, DF	S16°00'47"	W047°33'21.5"	Phaseolus vulgaris	SpeI	BR:Par2.1:12
						SpeI	BR:Par2.2:12
GO-66	June, 2012	Paranoá, DF	S16°00'47"	W047°33'21.5"	Phaseolus vulgaris	SpeI	BR:Par15.1:12
						HindIII	BR:Par15.2:12
GO-67	June, 2012	Paranoá, DF	S16°00'47"	W047°33'21.5"	Phaseolus vulgaris	HindIII	BR:Par67:12
GO-68	June, 2012	Paranoá, DF	S16°00'47"	W047°33'21.5"	Phaseolus vulgaris	SpeI	BR:Par16.1:12
						HindIII	BR:Par16.2:12
GO-70	June, 2012	Paranoá, DF	S16°00'47"	W047°33'21.5"	Phaseolus vulgaris	SpeI	BR:Par3:12
GO-71	June, 2012	Paranoá, DF	S16°00'47"	W047°33'21.5"	Phaseolus vulgaris	SpeI	BR:Par17:12
GO-72	June, 2012	Paranoá, DF	S16°00'47"	W047°33'21.5"	Phaseolus vulgaris	SpeI	BR:Par4:12
GO-73	June, 2012	Paranoá, DF	S16°00'47"	W047°33'21.5"	Phaseolus vulgaris	HindIII	BR:Par5.1:12
						SpeI	BR:Par5.2:12
GO-75	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par6.1:12
						HindIII	BR:Par6.2:12
GO-76	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par18:12
GO-78	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par19:12
GO-79	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par8:12
GO-80	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par20.1:12
						HindIII	BR:Par20.2:12

G0-81	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par21:12
GO-82	June, 2013	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par22:12
GO-85	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par23:12
GO-88	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par9.1:12
						SpeI	BR:Par9.2:12
GO-89	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par10:12
GO-91	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par24:12
GO-92	June, 2012	Paranoá, DF	S15°57'57.4"	W047°31'35.8"	Phaseolus vulgaris	SpeI	BR:Par25:12
GO-200	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri9:12
GO-201	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri10:12
GO-204	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri11:12
GO-207	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri12:12
GO-214	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri4:12
GO-221	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri6:12
						SpeI	BR:Cri6.1:12
GO-222	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri7.1:12
						SpeI	BR:Cri7.2:12
GO-227	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri13.1:12
						SpeI	BR:Cri13.2:12
GO-228	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri14:12
GO-236	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri15:12
GO-238	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri16:12
GO241	June, 2012	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	Phaseolus vulgaris	SpeI	BR:Cri17:12
159AL	June, 2012	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai2:11
162AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai4:11
163AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai5:11
164AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai6:11
165AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai7:11
166AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai8:11
167AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai9:11
168AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai10:11
169AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SacI	BR:Pai11:11
170AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai12:11
173AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai13.1:11
						SpeI	BR:Pai13.2:11
						SpeI	BR:Pai13.3:11

174AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SpeI	BR:Pai14.1:11
						SpeI	BR:Pai14.2:11
176AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SacI	BR:Pai15:11
177AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SacI	BR:Pai16.1:11
						SpeI	BR:Pai16.2:11
178AL	July, 2011	Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	Phaseolus lunatus	SacI	BR:Pai17:11
254AL	July, 2011	Murici, AL	S09°17'10.0"	W035°57'43.3"	Phaseolus lunatus	SpeI	BR:Mur1:11
255AL	July, 2011	Murici, AL	S09°17'10.0"	W035°57'43.3"	Phaseolus lunatus	SpeI	BR:Mur6.1:11
						SpeI	BR:Mur6.2:11
257AL	July, 2011	Murici, AL	S09°17'10.0"	W035°57'43.3"	Phaseolus lunatus	SacI	BR:Mur7:11
258AL	July, 2011	Murici, AL	S09°17'10.0"	W035°57'43.3"	Phaseolus lunatus	SpeI	BR:Mur8:11
262AL	July, 2011	Murici, AL	S09°17'10.0"	W035°57'43.3"	Phaseolus lunatus	SpeI	BR:Mur10.1:11
						SpeI	BR:Mur10.2:11
265AL	July, 2011	Murici, AL	S09°17'10.0"	W035°57'43.3"	Phaseolus lunatus	SacI	BR:Mur2.1:11
						SpeI	BR:Mur2.2:11
269AL	July, 2011	Murici, AL	S09°17'10.0"	W035°57'43.3"	Phaseolus lunatus	SacI	BR:Mur3:11
272AL	July, 2011	Murici, AL	S09°17'10.0"	W035°57'43.3"	Phaseolus lunatus	SpeI	BR:Mur12.1:11
						SpeI	BR:Mur12.2:11
274AL	July, 2011	Murici, AL	S09°17'10.0"	W035°57'43.3"	Phaseolus lunatus	SpeI	BR:Mur4:11
275AL	July, 2011	Murici, AL	S09°17'10.0"	W035°57'43.3"	Phaseolus lunatus	SpeI	BR:Mur5.1:11
						SpeI	BR:Mur5.2:11
310MG	March, 2011	Florestal, MG	S19°52'58.4"	W44°25'14.6"	Macroptilium lathyroides	XbaI	BR:Flt1:11
311MG	March, 2011	Florestal, MG	S19°52'58.4"	W44°25'14.6"	Macroptilium lathyroides	XbaI	BR:Flt2:11
321MG	March, 2011	Florestal, MG	S19°52'58.4"	W44°25'14.6"	Macroptilium lathyroides	HindIII	BR:Flt20:11
323MG		Florestal, MG	S19°52'58.4"	W44°25'14.6"	Macroptilium lathyroides	XbaI	BR:Flt323:11
325MG	March, 2011	Florestal, MG	S19°52'58.4"	W44°25'14.6"	Macroptilium lathyroides	HindIII	BR:Flt5.1:11
						XbaI	BR:Flt5.2:11
326MG	March, 2011	Florestal, MG	S19°52'58.4"	W44°25'14.6"	Macroptilium lathyroides	HindIII	BR:Flt6:11
						XbaI	BR:Flt6.1:12
330MG	March, 2011	Florestal, MG	S19°52'58.4"	W44°25'14.6"	Macroptilium lathyroides	XbaI	BR:Flt8.1:11
						XbaI	BR:Flt8.2:11
332MG	March, 2011	Florestal, MG	S19°52'58.4"	W44°25'14.6"	Macroptilium lathyroides	XbaI	BR:Flt13:11
337MG	March, 2011	Florestal, MG	S19°52'58.4"	W44°25'14.6"	Macroptilium lathyroides	XbaI	BR:Flt11:11
338MG	March, 2011	Florestal, MG	S19°52'58.4"	W44°25'14.6"	Macroptilium lathyroides	XbaI	BR:Flt12:11
Tomato severe rugose virus (ToSRV)							
DV202	July, 2008	Florestal, MG	S19°52'25.4"	W44°25'00.6"	Solanum lycopersicum	KpnI	BR:Flo202.3:08

						KpnI	BR:Flo202.4:08
DV203	July, 2008	Florestal, MG	S19°52'25.4"	W44°25'00.6"	Solanum lycopersicum	KpnI	BR:Flo203:08
DV206	July, 2008	Florestal, MG	S19°52'25.4"	W44°25'00.6"	Solanum lycopersicum	KpnI	BR:Flo206:08
DV208	July, 2008	Florestal, MG	S19°52'25.4"	W44°25'00.6"	Solanum lycopersicum	KpnI	BR:Flo208:08
DV218	July, 2008	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	Solanum lycopersicum	KpnI	BR:Car218:08
DV219	July, 2008	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	Solanum lycopersicum	KpnI	BR:Car219:08
DV220	July, 2008	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	Solanum lycopersicum	KpnI	BR:Car220:08
DV224	July, 2008	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	Solanum lycopersicum	KpnI	BR:Car224.1:08
						KpnI	BR:Car224.3:08
DV226	July, 2008	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	Solanum lycopersicum	KpnI	BR:Car226.1:08
						KpnI	BR:Car226.3:08
DV230	July, 2008	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	Solanum lycopersicum	KpnI	BR:Car230:08
DV232	July, 2008	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	Solanum lycopersicum	KpnI	BR:Car232:08
DV237	July, 2008	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	Solanum lycopersicum	KpnI	BR:Car237.1:08
						KpnI	BR:Car237.2:08
DV238	July, 2008	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	Solanum lycopersicum	KpnI	BR:Car238.1:08
OS70	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi70:13
OS73	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi73:13
OS79	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi79:13
OS80	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi80:13
OS83	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi83:13
OS84	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi84:13
OS85	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi85:13
OS86	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi86:13
OS92	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi92.2:13
						KpnI	BR:Coi92.5:13
OS93	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi93:13
OS95	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi95:13
OS103	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi103.1:13
						KpnI	BR:Coi103.4:13
OS110	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi110:13
OS114	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi114:13
OS117	July, 2013	Coimbra, MG	S20°51'43.2"	W042°51'27"	Solanum lycopersicum	KpnI	BR:Coi117:13
						KpnI	BR:Coi117.1:13
OS119	February, 2014	Coimbra, MG	S20°36'39.3 "	W042°25'58.9"	Solanum lycopersicum	KpnI	BR:Coi119.1:13
						KpnI	BR:Coi119.4:13

OS120	February, 2014	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	Solanum lycopersicum	KpnI	BR:Coi120:14
OS122	February, 2014	Coimbra, MG	S20°36'39.3"	W042°49'16.9"	Solanum lycopersicum	KpnI	BR:Coi122:14
OS125	February, 2014	Coimbra, MG	S20°50'51 "	W042°52'41.2"	Solanum lycopersicum	KpnI	BR:Coi125:14
OS131	February, 2014	Coimbra, MG	S20°50'49.4"	W042°52'41.9"	Solanum lycopersicum	KpnI	BR:Coi131:14
OS172	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Solanum lycopersicum	KpnI	BR:Coi172:14
OS179	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Solanum lycopersicum	KpnI	BR:Coi179:14
OS182	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Solanum lycopersicum	KpnI	BR:Coi182:14
OS183	February, 2014	Coimbra, MG	S20°50'58.3"	W042°53'14.4"	Solanum lycopersicum	KpnI	BR:Coi183:14
OS201	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo01:14
OS202	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo02:14
OS203	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo03:14
OS204	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo04.1:14
						KpnI	BR:Flo04.5:14
OS205	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo05:14
OS206	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo06:14
OS207	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo07:14
OS208	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo08.1:14
						KpnI	BR:Flo08.2:14
OS209	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo09.1:14
						KpnI	BR:Flo09.3:14
OS210	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo10:14
OS211	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo11:14
OS212	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo12:14
OS213	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo13:14
OS214	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo14:14
OS216	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo16:14
OS218	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo18:14
OS219	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo19:14
OS220	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo20:14
OS222	June, 2014	Florestal, MG	S19°55'55.8"	W044°23'52.4"	Solanum lycopersicum	KpnI	BR:Flo222:14

*State sigla: AL, Alagoas; BA, Bahia; DF, Federal District; GO, Goiás; MG, Minas Gerais; MS, Mato Grosso do Sul; PE, Pernambuco; PR, Paraná; RS, Rio Grande do Sul

Supplementary Table S2. Begomovirus sequences retrieved from GenBank.

Isolate name	GenBank access number		Location	Geographicalcoordinates		Date	Host	Reference
	DNA-A	DNA-B		Latitude	Longitude			
Blainvillea yellow spot virus (BlYSV)								
BR:Vic04.1:10	KC706516.1		Viçosa, MG*	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic04.2:10	KC706517.1		Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic07:10		KC706523.1	Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic08:10	KC706518.1	KC706524.1	Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic09:10	KC706519.1		Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic11:10	KC706520.1	KC706525.1	Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic13:10	KC706521.1		Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic18:10		KC706526.1	Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic20:10	KC706522.1		Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic21:10		KC706527.1	Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic26s:10		KC706529.1	Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Vic26c:10		KC706528.1	Viçosa, MG	S20°45'22.9"	W42°50'58.0"	May, 2010	Blainvillea rhomboidea	Rocha et al., 2013
BR:Coi25:07	EU710756.1	EU710757.1	Coimbra, MG	S20°51'29.9"	W042°51'42.2"	July, 2007	Blainvillea rhomboidea	Rocha et al., 2013
BR:Coi32.1:13			Coimbra, MG	S20°51'29.9"	W042°51'42.2"	July, 2013	Blainvillea rhomboidea	Sande, 2014
BR:Coi32.2:13			Coimbra, MG	S20°51'29.9"	W042°51'42.2"	July, 2013	Blainvillea rhomboidea	Sande, 2014
BR:Coi34:13			Coimbra, MG	S20°51'29.9"	W042°51'42.2"	July, 2013	Blainvillea rhomboidea	Sande, 2014
BR:Coi37.1:13			Coimbra, MG	S20°51'29.9"	W042°51'42.2"	July, 2013	Blainvillea rhomboidea	Sande, 2014
BR:Coi37.2:13			Coimbra, MG	S20°51'29.9"	W042°51'42.2"	July, 2013	Blainvillea rhomboidea	Sande, 2014
BR:Coi51:13			Coimbra, MG	S20°51'29.9"	W042°51'42.2"	July, 2013	Blainvillea rhomboidea	Sande, 2014
BR:Coi151:14			Coimbra, MG	S20°50'58.3"	W042°53'14.4"	February, 2014	Blainvillea rhomboidea	Sande, 2014
BR:Coi152.1:14			Coimbra, MG	S20°50'58.3"	W042°53'14.4"	February, 2014	Blainvillea rhomboidea	Sande, 2014
BR:Coi153:14			Coimbra, MG	S20°50'58.3"	W042°53'14.4"	February, 2014	Blainvillea rhomboidea	Sande, 2014
BR:Coi154:14			Coimbra, MG	S20°50'58.3"	W042°53'14.4"	February, 2014	Blainvillea rhomboidea	Sande, 2014
BR:Coi155:14			Coimbra, MG	S20°50'58.3"	W042°53'14.4"	February, 2014	Blainvillea rhomboidea	Sande, 2014
BR:Coi156:14			Coimbra, MG	S20°50'58.3"	W042°53'14.4"	February, 2014	Blainvillea rhomboidea	Sande, 2014
BR:Coi157:14			Coimbra, MG	S20°50'58.3"	W042°53'14.4"	February, 2014	Blainvillea rhomboidea	Sande, 2014
BR:Coi164:14			Coimbra, MG	S20°50'58.3"	W042°53'14.4"	February, 2014	Blainvillea rhomboidea	Sande, 2014
BR:Jun1:09	JX871394.1		Junqueiro, AL	S9°54'24.6"	W36°28'30.1"	November, 2009	Blainvillea rhomboidea	Tavares et al., 2012
BR:Lim1:09	JX871393.1		Limoeiro, AL	S9°46'05.4"	W36°30'41.6"	November, 2009	Blainvillea rhomboidea	Tavares et al., 2012
BR:Rla6:09	JX871392.1		Rio Largo, AL	S9°28'43.7"	W35°49'44.7"	November, 2009	Blainvillea rhomboidea	Tavares et al., 2012
BR:Rla5:10	JX871391.1		Rio Largo, AL	S9°28'43.7"	W35°49'44.7"	January, 2010	Blainvillea rhomboidea	Tavares et al., 2012
BR:Rla4:10	JX871390.1		Rio Largo, AL	S9°28'43.7"	W35°49'44.7"	January, 2010	Blainvillea rhomboidea	Tavares et al., 2012
BR:Rla3:10	JX871389.1		Rio Largo, AL	S9°28'43.7"	W35°49'44.7"	January, 2010	Blainvillea rhomboidea	Tavares et al., 2012
BgV06A.1.C80	JF694468.1		Rio Largo, AL	S9°28'43.7"	W35°49'44.7"	February, 2012	Blainvillea rhomboidea	Wyant et al., 2012
BgV06A.1.C81	JF694476.1		BA	S12°56'42.6"	W38°26'52.2"	February, 2012	Blainvillea rhomboidea	Wyant et al., 2012
BgV06B.1.C55		JF694470	BA	S12°56'42.6"	W38°26'52.2"	February, 2012	Blainvillea rhomboidea	Wyant et al., 2012

BgV06B.1.C56		JF694477.1	Rio Largo, AL	S9°28'43.7"	W35°49'44.7"	February, 2012	Blainvillea rhomboidea	Wyant et al., 2012
BgV06B.1.C70		JF694478.1	Rio Largo, AL	S9°28'43.7"	W35°49'44.7"	February, 2012	Blainvillea rhomboidea	Wyant et al., 2012
BgV06B.1.C82		JF694469.1	BA	S12°56'42.6"	W38°26'52.2"	February, 2012	Blainvillea rhomboidea	Wyant et al., 2012
Bean golden mosaic virus (BGMV)								
BR:Pai1:11	KJ939737.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai2:11	KJ939738.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai3:11	KJ939739.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai4:11	KJ939740.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai5:11	KJ939741.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai6:11	KJ939742.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai7:11	KJ939743.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai8:11	KJ939744.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai9:11	KJ939745.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai10:11	KJ939746.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai11:11	KJ939747.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai12:11	KJ939748.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai13:11	KJ939749.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai14:11	KJ939750.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai15:11	KJ939751.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai16:11	KJ939752.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Pai17:11	KJ939753.1		Palmeira dos Índios, AL	S09°27'11.3"	W036°36'34.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur1:11	KJ939754.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur2:11	KJ939755.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur3:11	KJ939756.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur4:11	KJ939757.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur5:11	KJ939758.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur6:11	KJ939759.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur7:11	KJ939760.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur8:11	KJ939761.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur9:11	KJ939762.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur10:11	KJ939763.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur11:11	KJ939764.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Mur12:11	KJ939765.1		Murici, AL	S09°17'10.0"	W035°57'43.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Flt1:11	KJ939766.1		Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt2:11	KJ939767.1		Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt3:11	KJ939768.1		Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt4:11	KJ939769.1		Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt5:11	KJ939770.1		Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt6:11	KJ939771.1		Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt7:11	KJ939772.1		Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt8:11	KJ939773.1		Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014

BR:Flt9:11	KJ939774.1	Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt10:11	KJ939775.1	Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt11:11	KJ939776.1	Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt12:11	KJ939777.1	Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Flt13:11	KJ939778.1	Florestal, MG	S19°52'58.4"	W44°25'14.6"	March, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Sag1:12	KJ939779.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag2:12	KJ939780.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag3:12	KJ939781.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag4:12	KJ939782.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag5:12	KJ939783.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag6:12	KJ939784.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag7:12	KJ939785.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag8:12	KJ939786.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag9:12	KJ939787.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag10:12	KJ939788.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag11:12	KJ939789.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Sag12:12	KJ939790.1	Santo Antônio de Goiás, GO	S16°30'22.5"	W049°17'06.8"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par14:12	KJ939791.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par1:12	KJ939792.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par2:12	KJ939793.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par15:12	KJ939794.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par16:12	KJ939795.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par3:12	KJ939796.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par17:12	KJ939797.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par4:12	KJ939798.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par5:12	KJ939799.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par6:12	KJ939800.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par18:12	KJ939801.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par7:12	KJ939802.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par19:12	KJ939803.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par8:12	KJ939804.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par20:12	KJ939805.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par21:12	KJ939806.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par22:12	KJ939807.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par23:12	KJ939808.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par9:12	KJ939809.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par10:12	KJ939810.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par24:12	KJ939811.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par25:12	KJ939812.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par26:12	KJ939813.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par27:12	KJ939814.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014

BR:Par11:12	KJ939815.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par28:12	KJ939816.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par29:12	KJ939817.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par30:12	KJ939818.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par12:12	KJ939819.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Par13:12	KJ939820.1	Paranoá, DF	S16°00'47"	W047°33'21.5"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una1:12	KJ939821.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una2:12	KJ939822.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una3:12	KJ939823.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una4:12	KJ939824.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una5:12	KJ939825.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una6:12	KJ939826.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una7:12	KJ939827.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una8:12	KJ939828.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una9:12	KJ939829.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una10:12	KJ939830.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una11:12	KJ939831.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una12:12	KJ939832.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una13:12	KJ939833.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una14:12	KJ939834.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una15:12	KJ939835.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Una16:12	KJ939836.1	Unaí, MG	S16°43'02.2"	W046°45'14"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri1:12	KJ939837.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri2:12	KJ939838.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri3:12	KJ939839.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri9:12	KJ939840.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri10:12	KJ939841.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri11:12	KJ939842.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri12:12	KJ939844.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri4:12	KJ939844.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri5:12	KJ939845.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri6:12	KJ939846.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri7:12	KJ939847.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri8:12	KJ939848.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri13:12	KJ939849.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri14:12	KJ939850.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri15:12	KJ939851.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri16:12	KJ939852.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Cri17:12	KJ939853.1	Cristalina, GO	S17°05'29.2"	W047°37'01.4"	June, 2012	Phaseolus vulgaris	Sobrinho et al., 2014
Tomato severe rugose virus (ToSRV)							
BR:Coi70.1:13		Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014

BR:Coi70.2:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi70.3:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi73.2:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi73.4:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi78:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi79:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi80:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi83:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi84:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi85.1:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi85.2:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi86:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi88:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi91:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi92.2:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi93:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi95:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi100.1:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi100.2:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi103.1:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi107:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi109:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi110:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi114:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi117:13	Coimbra, MG	S20°51'43.2"	W042°51'27"	July, 2013	Solanum lycopersicum	Sande, 2014
BR:Coi119.1:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Coi120:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Coi122:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Coi125:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Coi127:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Coi131:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Coi172:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Coi179:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Coi182:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Coi183.1:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Coi183.2:14	Coimbra, MG	S20°36'39.3"	W042°25'58.9"	February, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo01:14	Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo02:14	Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo04.1:14	Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo06:14	Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo07:14	Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014

BR:Flo09.1:14			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo13:14			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo14:14			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo15:14			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo16:13			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo18:13			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo19:13			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo22:14			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo23:14			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo31:14			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo37:14			Florestal, MG	S19°55'55.8"	W044°23'52.4"	June, 2014	Solanum lycopersicum	Sande, 2014
BR:Flo165:08	KC004070.1		Florestal, MG	S19°52'25.4"	W44°25'00.6"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Flo202:08	KC004071.1		Florestal, MG	S19°52'25.4"	W44°25'00.6"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Flo203:08	KC004072.1		Florestal, MG	S19°52'25.4"	W44°25'00.6"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Flo206:08	KC004073.1		Florestal, MG	S19°52'25.4"	W44°25'00.6"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Flo208:08	KC004074.1		Florestal, MG	S19°52'25.4"	W44°25'00.6"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car214:08	KC004075.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car218.1:08	KC004076.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car219.10:08	KC004077.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car220:08	KC004078.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car224:08	KC004079.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car226.3:08	KC004080.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car227:08	KC004081.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car228:08	KC004082.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Sida sp.	Rocha et al., 2013
BR:Car230:08	KC004083.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car232:08	KC004084.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car233:08	KC004085.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car235:08	KC004086.1	KC706625.1	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car236.1:08	KC004087.1		Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car237.6:08	KC004088.1	KC706626.1	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car238:08	KC004089.1	KC706627.1	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car237:08	KC706620.1	KC706624.1	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car217.6:08		KC706621.1	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car223:08		KC706622.1	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
BR:Car234.5:08		KC706623.1	Carandaí, MG	S20°56'56.5"	W43°47'42.2"	July, 2008	Solanum lycopersicum	Rocha et al., 2013
Macroptilium yellow spot virus (MaYSV)								
BR:Crb1:11	KC004111.1		Craíbas, AL	S09°40'37.9"	W036°46'37.8"	July, 2011	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Crb2:11	KC004116.1		Craíbas, AL	S09°40'39.1"	W036°46'38.4"	July, 2011	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Crb10:11	KC004093.1		Craíbas, AL	S09°40'39.8"	W036°46'38.4"	July, 2011	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Oaf8:11	KC004121.1		Olho D'água das Flores, AL	S09°32'28.1"	W037°17'26.8"	July, 2011	Phaseolus vulgaris	Sobrinho et al., 2014
BR:Oaf28:11	KJ939857.1		Olho D'água das Flores, AL	S09°32'27.8"	W037°17'22.1"	July, 2011	Macroptilium lathyroides	Sobrinho et al., 2014

BR:Oaf27:11	KJ939856.1		Olho D'água das Flores, AL	S09°32'26.9"	W037°17'17.6"	July, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Oaf24:11	KC004132.1		Olho D'água das Flores, AL	S09°32'59.4"	W037°18'43.5"	July, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Oaf25:11	KC004133.1		Olho D'água das Flores, AL	S09°32'59.2"	W037°18'43.0"	July, 2011	Macroptilium lathyroides	Sobrinho et al., 2014
BR:Sti34:11	KJ939891.1		Santana do Ipanema, AL	S09°23'25.1"	W037°12'46.9"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Sti2:11	KJ939860.1		Santana do Ipanema, AL	S09°23'24.7"	W037°12'48.0"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Sti3:11	KJ939861.1		Santana do Ipanema, AL	S09°23'24.8"	W037°12'48.0"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Sti4:11	KJ939862.1		Santana do Ipanema, AL	S09°23'24.9"	W037°12'48.0"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Sti35:11	KJ939892.1		Santana do Ipanema, AL	S09°23'24.9"	W037°12'48.3"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Sti9:11	KJ939867.1		Santana do Ipanema, AL	S09°23'24.9"	W037°12'47.5"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Sti10:11	KJ939868.1		Santana do Ipanema, AL	S09°23'24.9"	W037°12'47.6"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Sti26:11	KJ939883.1		Santana do Ipanema, AL	S09°23'24.9"	W037°12'47.6"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:Sti29:11	KJ939886.1		Santana do Ipanema, AL	S09°23'24.8"	W037°12'48.6"	July, 2011	Phaseolus lunatus	Sobrinho et al., 2014
BR:PE:CAU:22_1	KT779562.1		Caruaru, PE	S8°17'47.6"	W35°58'56.1"	August, 2013	Desmodium glabrum	Fontenele et al., 2016
BR:PE:CAU:22_2	KT779561.1		Caruaru, PE	S8°17'47.6"	W35°58'56.1"	August, 2013	Desmodium glabrum	Fontenele et al., 2016
BR:PE:CAU:22_3		KT779560.1	Caruaru, PE	S8°17'47.6"	W35°58'56.1"	August, 2013	Desmodium glabrum	Fontenele et al., 2016
BR:PE:CAU:22_4		KT779559.1	Caruaru, PE	S8°17'47.6"	W35°58'56.1"	August, 2013	Desmodium glabrum	Fontenele et al., 2016
BR:PE:CAU:22_5		KT779558.1	Caruaru, PE	S8°17'47.6"	W35°58'56.1"	August, 2013	Desmodium glabrum	Fontenele et al., 2016
BR:PE:CAU:23_1	KT779565.1	KT779563.1	Caruaru, PE	S8°17'47.6"	W35°58'56.1"	August, 2013	Desmodium glabrum	Fontenele et al., 2016
BR:PE:CAU:23_2	KT779564.1	KT779563.1	Caruaru, PE	S8°17'47.6"	W35°58'56.1"	August, 2013	Desmodium glabrum	Fontenele et al., 2016
Euphorbia yellow mosaic virus (EuYMV)								
BR:GO:Ita5082:07	JF756671	JF756677	Itaberaí, GO	S16°01'23.9"	W49°47'31.0"	2007	Euphorbia heterophylla	Mar et al., 2017
BR:Sag32:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag33:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag35:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag34:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag40.1:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag40.2:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag43:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag44:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag52:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag53:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag54:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag56:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag57:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag58:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag59:12 2A			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag61.1:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Sag61.2:12			Santo Antônio de Goiás, GO	S16°30'22.50"	W49°17'06.80"	July, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:MS:Mir2:07	FN435997	FN435998	Miranda, MS	S20°15'41.9"	W56°22'14.9"	2007	Euphorbia heterophylla	Mar et al., 2017
BR:MS:Mir:1:07	FN435995	FN435996	Miranda, MS	S20°15'41.9"	W56°22'14.9"	2007	Euphorbia heterophylla	Mar et al., 2017
BR:Gua183:09			Guamiranga, PR	S25°11'15.28"	W50°52'36.98"	September, 2009	Euphorbia heterophylla	Mar et al., 2017

BR:Nom682:10			Nova Mercedes, PR	S24°30'54.70"	W54°07'0.22"	June, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Caf777:10			Cafelândia, PR	S24°39'21.30"	W53°13'07.40"	September, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Frb1019:11			Francisco Beltrão, PR	S26°04'01"	W53°00'42"	April, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Frb1019.1:11			Francisco Beltrão, PR	S26°04'01"	W53°00'42"	April, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Can1033:11			Candói, PR	S25°32'07.74"	W52°00'03.55"	April, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Amp1068:11			Ampére, PR	S25°57'06.90"	W53°24'25"	April, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Ara1133:11			Araruna, PR	S24°03'50"	W52°33'52"	April, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Ats03:09			Almirante Tamandaré, RS	S28°06'35.02"	W52°54'29.57"	September, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Ats3.1:09			Almirante Tamandaré, RS	S28°06'35.02"	W52°54'29.57"	September, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Ats3.2:09			Almirante Tamandaré, RS	S28°06'35.02"	W52°54'29.57"	September, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Ats3.3:09			Almirante Tamandaré, RS	S28°06'35.02"	W52°54'29.57"	September, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Cha14:09 2A			Chapada, RS	S28°02'39.56"	W53°04'53.95"	March, 2009	Euphorbia heterophylla	Mar et al., 2017
BR:Sta51:09			Santo Ângelo, RS	S28°22'54.70"	W54°18'17.23"	March, 2009	Euphorbia heterophylla	Mar et al., 2017
BR:Smm61:09 2A	KY559509.1	KY559614.1	São Miguel das Missões, RS	S28°29'35.59"	W54°33'37.15"	March, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Ats504:10	KY559517.1	KY559615.1	Almirante Tamandaré, RS	S28°06'35.02305"	52°54'29.57"	March, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Ats504.1:10		KY559616.1	Almirante Tamandaré, RS	S28°06'35.02305"	52°54'29.57"	March, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Cha510:10	KY559518.1	KY559617.1	Chapada, RS	S28°02'39"	W53°04'53"	March, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Cha510.1:10	KY559519.1	KY559618.1	Chapada, RS	S28°02'39"	W53°04'53"	March, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Cha510.2:10		KY559619.1	Chapada, RS	S28°02'39"	W53°04'53"	March, 2012	Euphorbia heterophylla	Mar et al., 2017
BR:Str536:10	KY559522.1	KY559620.1	Santa Rosa, RS	S27°52'52.94"	W54°26'02.50"	March, 2010	Euphorbia heterophylla	Mar et al., 2017
BR:Cha920:11	KY559525.1	KY559621.1	Chapada, RS	S28°02'39.56"	W53°04'53.95"	March, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Cha920.1:11	KY559526.1	KY559622.1	Chapada, RS	S28°02'39.56"	W53°04'53.95"	March, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Cha925:11	KY559530.1	KY559623.1	Chapada, RS	S28°01'08.80"	W53°05'52.31"	March, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Plm933:11	KY559531.1	KY559624.1	Palmeira das Missões, RS	S27°45'57.35"	W53°27'29.49"	March, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Sau937:11	KY559532.1	KY559625.1	Santo Augusto, RS	S27°44'12.44"	W53°51'27.91"	March, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Trm939:11	KY559533.1	KY559626.1	Três de Maio, RS	S27°45'39.44"	W54°15'42.87"	March, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Str942:11	KY559534.1	KY559627.1	Santa Rosa, RS	S27°45'39.44"	W54°15'42.87"	March, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Sta946:11	KY559535.1	KY559628.1	Santo Ângelo, RS	S28°22'54.70"	W54°18'17.23"	March, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Pon1000:11	KY559540.1	KY559631.1	Pontão, RS	S27°58'09.60"	W52°43'47.90"	April, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Non1003:11	KY559541.1	KY559632.1	Nonoai, RS	S27°29'49.71"	W52°54'07.23"	April, 2011	Euphorbia heterophylla	Mar et al., 2017
BR:Cha1179:14	KY559543.1	KY559633.1	Chapada, RS	S28°02'39.56"	W53°04'53.95"	July, 2014	Euphorbia heterophylla	Mar et al., 2017
BR:Cha1179.1:14	KY559544.1	KY559634.1	Chapada, RS	S28°02'39.56"	W53°04'53.95"	July, 2014	Euphorbia heterophylla	Mar et al., 2017
BR:Cha1179.2:14	KY559545.1	KY559635.1	Chapada, RS	S28°02'39.56"	W53°04'53.95"	July, 2014	Euphorbia heterophylla	Mar et al., 2017
BR:Cha1179.3:14	KY559546.1	KY559636.1	Chapada, RS	S28°02'39.56"	W53°04'53.95"	July, 2014	Euphorbia heterophylla	Mar et al., 2017

*State sigla: AL, Alagoas; BA, Bahia; DF, Federal District; GO, Goiás; MG, Minas Gerais; MS, Mato Grosso do Sul; PE, Pernambuco; PR, Paraná; RS, Rio Grande do Sul

Supplementary Table S3. Recombination events detected in the DNA-A and DNA-B components of Bean golden mosaic virus (BGMV), Tomato severe rugose virus (ToSRV), Euphorbia yellow mosaic virus (EuYMV), Macroptilium yellow spot virus (MaYSV) and Blainvillea yellow spot virus (BlYSV).

Species	Component	Event	Recombinant	Recombination breakpoints*		Parents		Method#	P-value&
				Begin	End	Minor	Major		
BGMV	DNA-B	1	BR:Cri15:12	179	53	Unknown	BR:Cri14:12	GMCS**3**	5.147 x 10^{-30}
		2	BR:Cri13.2:12	1013	2379	BR:Una10.2:12	BR:Par8:12	MC**S**3	2.712 x 10^{-11}
		3	BR:Una16:12	988	2368	BR:Una10.2:12	BR:Una2:12	MCS**3**	1.629 x 10^{-03}
		4	BR:Mur12.1:11, BR:Mur12.2:11	544	2330	Unknown	BR:Mur3:11	RGMCS**3**	1.748 x 10^{-18}
ToSRV	DNA-B	1	BR:Flo01:14	232	1276	BR:Coi183:13	BR:Flo18:14	GMC**3**	2.156 x 10^{-14}
EuYMV	DNA-B	1	BR:Frb1019:11, BR:Frb1019.1:11,	212	833	BR:Nom682.2:10	BR:Str942:11	RGMSC**3**	2.478 x10^{-08}
		2	BR:Can1033:11	58	2215	Unknown	BR:Gua183:09	MCS**3**	1.892 x 10^{-05}
		3	BR:Sag61.1:12	1494	2564	BR:Sag44:12	BR:Sag40.2:12	RM**S**3	4.046 x 10^{-08}
		4	BR:Cha1179.1:14, BR:Cha510.2:10, BR:Cha1179.2:14,	1306	1956	Unknown	BR:Sau937:11	RGMC**S**3	2.632 x 10^{-08}
		5	BR:Sag34:12, BR:Sag40.1:12, BR:Sag40.2:12, BR:Sag43:12, BR:Sag44:12, BR:Sag52:12, BR:Sag53:12, BR:Sag54:12, BR:Sag56:12, BR:Sag57:12, BR:Sag58:12, BR:Sag59:12, BR:Sag61.1:12, BR:Sag61.2:12, BR:Ats3.2:09	1101	1369	Unknown	BR:Trm939:11	GM**S**3	1.116 x 10^{-03}
MaYSV	DNA-A	1	BR:Crb2:11, BR:Crb1:11, BR:Sti4:11, BR:Sti29:11, BR:Sti34:11	1853	2643	BR:PE:CAU:23:2	BR:Crb10:11	RGBMC**S**3	1.032 x 10^{-35}
		2	BR:Oaf25:11	1893	2619	Unknown	BR:Sti3:11	RGMC**S**3	5.847 x10^{-35}

		3	BR:Oaf8:11	1941	401	Unknown	BR:Oaf24:11	RGMC**S**3	9.771 x 10^{-33}
		4	BR:Sti2:11	2626	1979	Unknown	BR:Crb10:11	RGMCS**3**	1.540 x 10^{-11}
		5	BR:Oaf8:11, BR:Oaf25:11, BR:Sti2:11	1941(?)	2157	Unknown	BR:Sti10:11	RGM**C**3	5.057 x 10^{-11}
		6	BR:PE:CAU:23:2, BR:PE:CAU:22:1, BR:PE:CAU:22:2, BR:PE:CAU:23:1	988	1763	Unknown	BR:Crb10:11	RGBMC**S**3	2.581 x 10^{-41}
		7	BR:Sti35:11, BR:Sti3:11, BR:Sti9:11, BR:Sti26:11	1326	233	BR:Oaf27:11	BR:Sti10:11	MC**S**3	8.640 x 10-07
	DNA-B	1	BR:Crb2:11, BR:Crb1:11	2596	2075	Unknown	BR:Crb10:11	RGMCS**3**	1.035 x 10^{-24}
		2	BR:Sti34:11, BR:Sti4:11	1452	2506	BR:Oaf8:11	Unknown	RGMC**S**3	1.379 x 10^{-13}
		3	BR:Sti29:11, BR:Sti4:11, BR:PE:CAU:22:3, BR:PE:CAU:22:4, BR:PE:CAU:22:5	2427	2596	Unknown	BR:Oaf24:11	RGMCS**3**	9.343 x 10^{-19}
		4	BR:Sti2:11	1553	2595	BR:Oaf8:11	BR:Sti26:11		
		5	BR:Sti4:11, BR:Sti10:11	2061	2444(?)	BR:Sti35:11	BR:Sti29:11	RMS**3**	1.977 x 10^{-09}
		6	BR:Sti29:11, BR:Oaf24:11, BR:Oaf25.1:11, BR:Oaf25.2:11	2597(?)	673	BR:Sti2:11	BR:Sti4:11	RGBMCS**3**	1.397 x 10^{-07}
		7	BR:Sti9.3:11, BR:Sti9.1:11, BR:Sti9.2:11	2329	2423	BR:Oaf27:11	Unknown	**R**GMCS3	6.316 x 10^{-09}
		8	BR:Sti34:11	406	1451(?)	BR:Oaf27:11	Unknown	RGMCS**3**	1.511 x 10^{-09}
		9	BR:PE:CAU:23:1	1996	2434(?)	BR:PE:CAU:22:5	BR:Oaf24:11	RMCS**3**	2.469 x10^{-09}
		10	BR:Sti9.3:11	2597	537	Unknown	BR:Sti9.2:11	RMCS**3**	1.046 x 10^{-08}
		11	BR:Oaf28:11	1143	2328	BR:Sti3:11	BR:Oaf24:11	RMS**3**	5.766 x 10^{-08}
		12	BR:Oaf28:11, BR:Oaf24:11, BR:Oaf25.1:11,BR:Oaf27:11, BR:Oaf25.2:11	2329(?)	2430	Unknown	BR:Sti3:11	RG**M**S3	6.789 x 10^{-06}
		13	BR:Sti10:11	340	685	BR:Oaf8:11	BR:Sti4:11	RGMCS**3**	4.572 x 10^{-06}
BlYSV	DNA-A	1	BgV06A.1.C80	642	2059	BR:Lim1:09	BgV06A.1.C81	GMC**S**3	5.488 x 10^{-06}

	2	BR:Coi32.2:13, BR:Coi32.1:13, BR:Rla6:09, BR:Rla5:10, BR:Rla4:10, BR:Rla3:10, BgV06A.1.C81	2403	2641	Unknown	BR:Vic04.2:10	RGMC**S**3	5.346 x10^{-07}
	3	BgV06A.1.C80, BR:Jun1:09, BR:Lim1:09, BR:Rla6:09	2060(?)	202	Unknown	BR:Coi37.2:13	GMC**S**3	5.488 x 10^{-06}
DNA-B	1	BR:Vic07:10, BR:Vic04.1:10, BR:Vic08:10, BR:Vic09:10, BR:Vic18:10, BR:Vic21:10, BR:Coi25:07, BR:Vic26c:10, BR:Vic26s:10, BR:Coi165:14, BR:Coi164:14, BR:Coi155:14, BR:Coi154:14, BR:Coi150:14, BR:Coi145:14, BR:Coi37.1:13, BR:Coi36:13, BR:Coi33:13, BR:Coi32.1:13	2578	1865	Unknown	BR:Vic13:10	RGMC**S**3	3.364x10^{-22}
	2	BR:Coi33:13, BR:Coi142.1:14, BR:Coi151:14, BR:Coi152.1:14, BR:Coi155:14, BR:Coi154:14	2577(?)	464	BR:Coi158s:14	BR:Vic07:10	RGMC**3**	1.345 x 10^{-07}
	3	BR:Coi148:14, BR:Coi158b:14, BR:Coi158s:14	1970	2485	BR:Vic11:10	BR:Coi152.2:14	**R**GMCS3	1.250 x 10^{-09}
	4	BR:Coi157:14	1372	2512	BR:Coi160.2:14	BR:Coi35:13	RGMC**S**3	3.330 x 10^{-14}
	5	BR:Vic08:10, BR:Vic04.1:10, BR:Vic07:10, BR:Vic09:10, BR:Vic18:10, BR:Vic21:10, BR:Coi25:07, BR:Vic26c:10, BR:Vic26s:10, BR:Coi36:13, BR:Coi37.1:13, BR:Coi164:14, BR:Coi165:14	2579(?)	248	BgV06B.1.C56	BR:Coi158s:14	**R**GMC	5.498 x 10^{-05}
	6	BR:Vic11:10, BR:Vic13:10, BR:Coi34:13, BR:Coi35:13, BR:Coi142.2:14, BR:Coi160.2:14, BR:Coi157:14,	292	1029	Unknown	BR:Coi164:14	GMC**S**3	3.187 x 10^{-19}

	BR:Coi152.2:14, BR:Coi148:14, BR:Coi147:14						
7	BR:Vic07:10, BR:Coi33:13, BR:Coi164:14	249	1034	BR:Coi25:07	BR:Coi145:14	R**M**S3	7.821 x 10^{-06}
8	BR:Coi145:14, BR:Coi32.1:13, BR:Coi150:14, BR:Coi154:14	320	1483(?)	Unknown	BR:Coi158s:14	GMC**3**	2.106 x10^{-03}
9	BR:Coi160.2:14, BR:Coi34:13, BR:Coi51:13, BR:Coi157:14, BR:Coi160.1:14, BR:Coi159:14	2422	2575(?)	Unknown	BR:Coi152.2:14	RMC**S**3	4.049 x 10^{-05}

[*] Numbering starts at the first nucleotide after the cleavage site at the origin of replication and increase clockwise. (?), Breakpoints could not be accurately located.

[#] R, Rdp; G, Geneconv; B, Boostcan; M, Maxichi; C, Chimaera; S, Siscan; 3, 3Seq.

[&] The reported P-value is from the method in bold and underlined, and is the lowest P-value calculated for the featured event.

Supplementary Table S4. Parameters used in the Discriminant Analysis of Principal Components (DAPC) analysis for Bean golden mosaic virus (BGMV), Euphorbia yellow mosaic virus (EuYMV), Macroptilium yellow spot virus (MaYSV) and Tomato severe rugose virus (ToSRV) data sets.

			Discriminant Analysis (DA)		
Virus	**Component**	**K***	**Number of retained principal components (PCs)**	**Number of retained discriminant functions**	**Proportion of conserved variance (%)**
BGMV	DNA-A	5	6	3	96.97
	DNA-B	4	6	3	90.93
EuYMV	DNA-A	3	4	2	61.32
	DNA-B	3	10	2	67.51
MaYSV	DNA-A[#]	4	3	3	83.75
	DNA-A[&]	2	5	1	86.45
ToSRV	DNA-A	4	8	3	74.12
	DNA-B	4	7	3	74.73

* Number of clusters inferred by the k-means algorithm based on the Bayesian Information Criterion (BIC) used for DAPC analysis.

[#] Analysis performed considering recombination events.

[&] Analysis performed excluding recombination events.

Supplementary Table S5. Genetic variability indices of the DNA-A data sets 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).

Population	N*	h	Hd	S	DNA-A π	CP π	Rep π	Trap π	Ren π	AC4 π	IR-A π
BGMV (Total)	117	110	0.998	439	0.03965	0.03758	0.03090	0.02156	0.02337	0.02337	0.09498
MG-1	13	10	0.923	-	0.00183	0.00204	0.00073	0.00118	0.00347	0.00347	0.00237
MW-1	42	41	0.999	-	0.00323	0.00211	0.00221	0.00134	0.00084	0.00084	0.01172
MW-2	33	33	1.000	-	0.00283	0.00142	0.00205	0.00031	0.00131	0.00131	0.01103
AL-1	18	17	0.993	-	0.00300	0.00274	0.00205	0.00199	0.00306	0.00306	0.00732
AL-2	11	9	0.945	-	0.00107	0.00178	0.00102	0.00000	0.00046	0.00046	0.00056
BlYSV (Total)	30	26	0.991	433	0.03742	0.03334	0.03408	0.03616	0.03006	0.02831	0.05812
EuYMV (Total)	50	46	0.995	407	0.02319	0.01949	0.02546	0.01752	0.01510	0.01740	0.03648
GO	17	15	0.985	-	0.01067	0.00984	0.01112	0.01127	0.00944	0.00879	0.02800
SO-1	25	25	1.000	-	0.01738	0.01761	0.01867	0.01193	0.01139	0.01223	0.03147
SO-2	8	7	0.964	-	0.01120	0.00952	0.01023	0.00952	0.01038	0.01495	0.03539
MaYSV (Total)	21	20	0.995	506	0.07170	0.04451	0.11122	0.02773	0.02991	0.18584	0.06927
AL-1	9	9	1.000	-	0.01712	0.01745	0.01258	0.02399	0.02283	0.00991	0.02207
AL-2	3	3	1.000	-	0.02683	0.01058	0.03438	0.03248	0.03175	0.02067	0.03272
AL-3	5	5	1.000	-	0.01472	0.01534	0.00802	0.01538	0.02155	0.00388	0.02515
PE	4	3	0.833	-	0.00077	0.00066	0.00048	0.00128	0.00125	0.00194	0.00153
ToSRV (Total)	74	62	0.993	268	0.01062	0.01111	0.00965	0.00548	0.00759	0.01208	0.01808
MG-1	5	5	1.000	-	0.00803	0.00767	0.00812	0.00256	0.00301	0.00455	0.01584
MG-2	16	13	0.967	-	0.00315	0.00306	0.00278	0.00263	0.00203	0.00574	0.00542
MG-3	23	16	0.945	-	0.00558	0.00282	0.00455	0.00518	0.00506	0.00680	0.01630
MG-4	30	28	0.995	-	0.00450	0.00320	0.00384	0.00367	0.00590	0.00515	0.00897

* N, Number of sequences; h, haplotype number; Hd, haplotype diversity; S, number of segregating sites; π, average pairwise number of nucleotide differences per site (nucleotide diversity).

Supplementary Table S6. Genetic variability indices of the DNA-B data sets 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).

Population	N*	h	Hd	S	DNA-B π	MP π	NSP π	LIR-B π	SIR-B π
BGMV (Total)	123	104	0.997	628	0.05454	0.04015	0.0408	0.08662	0.11303
MG-1	13	11	0.962	-	0.00226	0.00174	0.0002	0.00374	0.00237
MW	76	62	0.991	-	0.01061	0.00679	0.0092	0.01779	0.02761
AL-1	19	17	0.982	-	0.00623	0.00480	0.0096	0.01065	0.00936
AL-2	15	14	0.990	-	0.01647	0.01775	0.0126	0.01496	0.05714
BlYSV (Total)	41	39	0.998	840	0.07552	0.05053	0.07761	0.09226	0.13834
EuYMV (Total)	53	48	0.996	677	0.04000	0.02561	0.03650	0.05375	0.09018
GO	15	13	0.971	-	0.02683	0.02139	0.01794	0.03870	0.05090
South-1	28	27	0.997	-	0.02757	0.01816	0.02720	0.03729	0.04309
South-2	10	8	0.956	-	0.03953	0.02717	0.03891	0.05224	0.05333
MaYSV (Total)	24	20	0.982	503	0.05567	0.03237	0.03948	0.08973	0.13541
ToSRV (Total)	74	72	0.999	434	0.02143	0.01134	0.01073	0.03886	0.05075
MG-1	5	5	1.000	-	0.01224	0.0066	0.01608	0.01557	0.00488
MG-2	22	20	0.991	-	0.00709	0.00391	0.00247	0.01454	0.00565
MG-3	19	19	1.000	-	0.00452	0.0376	0.00349	0.01267	0.01429
MG-4	28	28	1.000	-	0.01012	0.00532	0.00466	0.01563	0.02202

* N, Number of sequences; h, haplotype number; Hd, haplotype diversity; S, number of segregating sites; π, average pairwise number of nucleotide differences per site (nucleotide diversity).

Supplementary Table S7. Reassortment events detected in 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) data sets.

Species	Event	Breakpoints*		Reassortants	Parents		Method#	P-value&
		Begin	End		Minor	Major		
BGMV	1	2618	403	BR:Cri13.2:12, BR:Una4:12, BR:Una16:12, BR:Cri11:12	Unknown	BR-Par4-12	MC**S**3	1.98×10^{-10}
	2	35	2593	BR:Par19:12, BR:Una2:12, BR:Una3:12, BR:Una9:12, BR:Una12:12, BR:Cri9:12, BR:Cri10:12, BR:Cri12:12, BR:Cri13.1:12, BR:Cri14:12, BR:Cri15:12, BR:Cri16:12, BR:Sag5:12, BR:Sag8:12, BR:Par1:12, BR:Par2.1:12, BR:Par2.2:12, BR:Par4:12, BR:Par5.1:12, BR:Par5.2:12, BR:Par9.1:12, BR:Par9.2:12, BR:Par10:12, BR:Par16.1:12, BR:Par16.2:12, BR:Par18:12, BR:Par20.1:12, BR:Par20.2:12, BR:Par21:12, BR:Par24:12, BR:Par25:12	Unknown	BR-Una6.1-12	RGBMCS**3**	4.65×10^{-6}
BlYSV	1	5305	2657	BgV06A.1.C80:1	BgV06A.1.C80:2	BgV06A.1.C81	GBMCS**3**	5.85×10^{-69}
	2	5282	2658	BR:Vic13:10	Unknown	BR:Vic11:10	RGBMC**S**	1.19×10^{-42}
	3	5201	2594	BR:Coi152.2:14	BR:Coi152.1:14	BR:Coi51:13	RGBMCS**3**	1.53×10^{-32}
	4	1	2399	BR:Vic08:10, BR:Coi37.2:13	BR:Coi51:13	BR:Vic04.2:10	RGBMC**S**3	1.19×10^{-9}
EuYMV	1	2683	9	BR:Cha925:11, BR:Cha920:11, BR:Cha920.1:11, BR:Cha925.1:11	BR:Cha510:10	BR:Ats504:10	RGBMCS**3**	6.3×10^{-33}
	2	2678	7	BR:Cha510.2:10	Unknown	BR:Cha510.1:10	GBMCS**3**	5.69×10^{-27}
	3	2618	5018	BR:Ats3.1:09, BR:Ats3.2:09, BR:Ats3.4:09	BR:Plm933:11	BR:Ats3.3:09	RGBMCS**3**	2.05×10^{-19}

	4	2702	106	BR:Cha1179.3:14, BR:Cha1179:14	BR:Str536:10	BR:Cha1179.2:14	MCS**3**	1.05 x 10^{-16}
	5	25	2620	BR:Sta946:11, BR:Ats3.3:09, BR:Ats504:10, BR:Ats504.1:10, BR:Trm939:11	Unknown	BR:Sau937:11	RGBM**C**S3	2.36 x 10^{-9}
	6	42	2629	BR:Cha510:10, BR:Non1003:11	BR:Sau937:11	BR:Cha510.1:10	R**M**C3	1.36 x 10^{-9}
	7	5168	2689	BR:Smm61:09, BR:Ara1133:11, BR:Smm61.1:09	BR:Str536:10	Unknown	RGBMC**S**	1.45 x 10^{-14}
	8	2851	4980	BR:MS:Mir2:07	BR:MS:Mir1:07	Unknown	RGBMC**S**	3.05 x 10^{-17}
	9	2526	44	BR:Sag59:12	BR:Sag59.1:12	BR:Sag43:12	RG**B**M3	6.02 x 10^{-8}
	10	40	2676	BR:Sag61.1:12	BR:Sag61.2:12	Unknown	RGBMC**S**3	1.06 x 10^{-8}
	11	5176	2692	BR:Str942:11, BR:Plm933:11, BR:Pon1000:11, BR:Cha1179.1:14, BR:Cha1179.2:14	BR:Cha14.3a:09	BR:Sta51:09	RM**C**3	1.20 x 10^{-4}
	12	35	2750	BR:Sag57:12	BR:Sag56:12	BR:Sag61.2:12	GBMC**S**3	1.01 x 10^{-6}
MaYSV	1	2629	5085	BR:Sti29:11, BR:Crb1:11, BR:Crb2:11, BR:Sti4:11, BR:PE:CAU:22:1, BR:PE:CAU:22:2, BR:PE:CAU:22:3, BR:PE:CAU:23:1, BR:PE:CAU:23:2	BR:Oaf25.1:11	BR:Sti34:11	RGBMCS**3**	3.78 x 10^{-9}
	2	2865	5100	BR:Oaf27:11, BR:Oaf24:11, BR:Oaf25.1:11, BR:Oaf25.2:11, BR:Oaf28:11, BR:Sti9.1:11, BR:Sti9.2:11, BR:Sti9.3:11, BR:Sti10:11, BR:Sti35:11	Unknown	BR:Sti26:11	RGBMCS**3**	1.04 x 10^{-18}
ToSRV	1	2516	41	BR:Flo202.2:08	BR:Flo203:08	BR:Flo202.1:08	BMCS**3**	1.06 x 10^{-11}
	2	2620	5134	BR:Coi179:14	Unknown	BR:Coi93:13	MC**S**3	7.20 x 10^{-9}

* Breakpoints correspond to the artificial joint between the two concatenated genomic components.

R, Rdp; G, Geneconv; B, Boostcan; M, Maxichi; C, Chimaera; S, Siscan; 3, 3Seq.

& The reported P-value is from the method in bold and underlined, and is the lowest P-value calculated for the featured event.

Supplementary figure legends

Supplementary Figure S1. Inference of the number of genetic clusters (k) using the k-means algorithm and analysis of molecular variance (AMOVA) for 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) data sets. The k-means algorithm was run sequentially with values of k varying from 2 to 10. The choice of the number of clusters (indicated by the vertical dotted line) was primarily based on the minimum number of k after which the Bayesian Information Criterion (BIC, blue lines) decreases by an insignificant amount, as proposed by Jombart et al. (2010). In addition, the percentage of variation among genetic clusters was plotted according to AMOVA analysis (red lines). A number of plots do not have the vertical dotted line due to unclear genetic cluster identification.

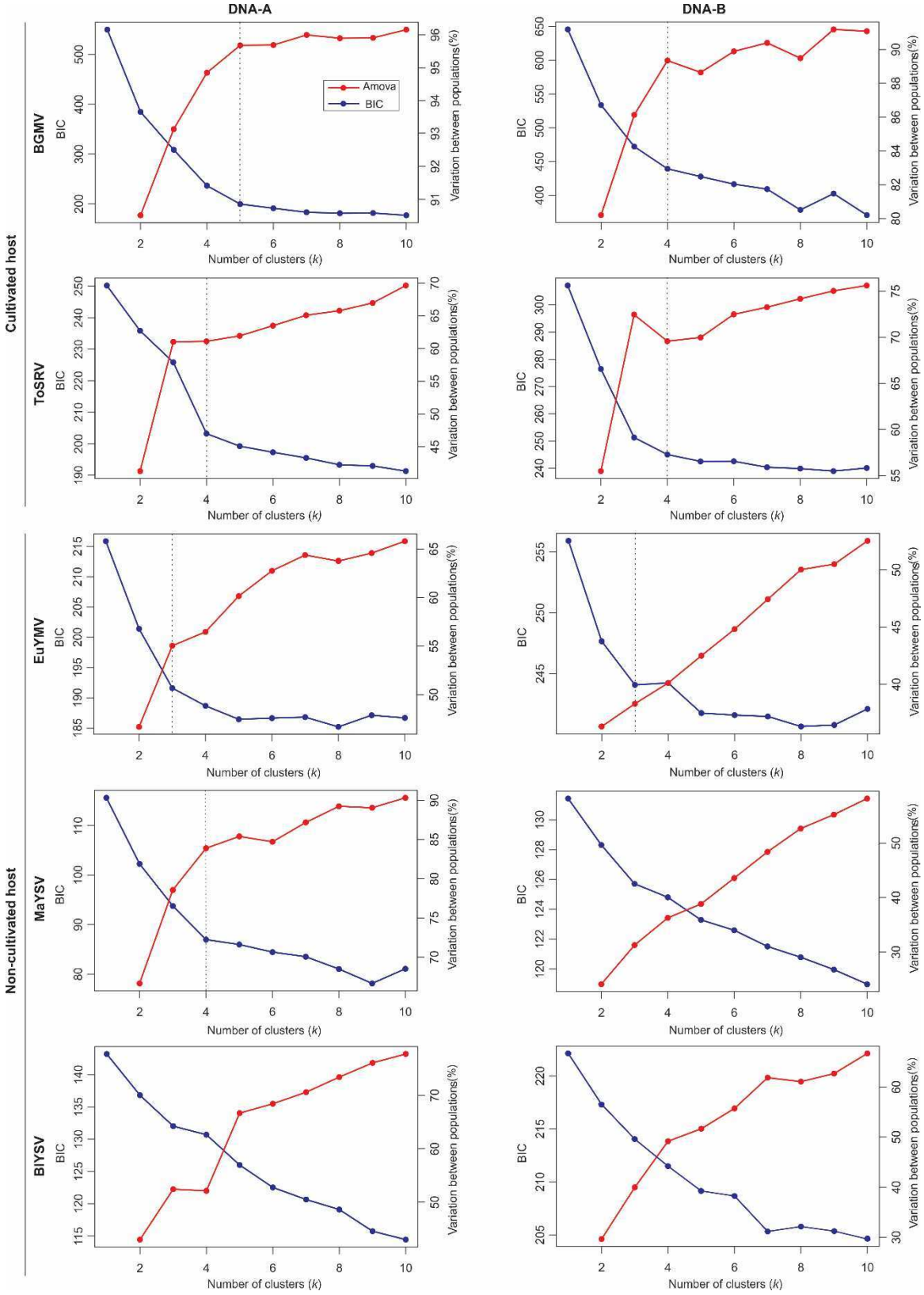
Supplementary Figure S2. Analysis of population subdivision using Discriminant Analysis of Principal Components (DAPC) for Macroptilium yellow spot virus (MaYSV) DNA-A using a data set without recombinants blocks. (**A**) Number of genetic clusters (k) inferred using the k-means algorithm. The algorithm was run sequentially with values of k varying from 2 to 10. The vertical dotted line indicates the number of genetic clusters chosen for DAPC analysis. (**B**) Density of isolates on discriminant function 1, colored according to the geographical origin of each genetic cluster. **(C-E)** Sites that contribute most to population subdivision when recombinants blocks are excluded (**C**) and when they are considered (**D** and **E**). The heights of the bars correspond to the contribution of the corresponding site.

Supplementary Figure S3. Statistical significance of the difference between the average pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated for genetic clusters inferred by Discriminant Analysis of Principal Components (DAPC) for the DNA-A and DNA-B 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). Ninety-five percent bootstrap confidence intervals (CIs) for the difference between π values were estimated from 1,000 nonparametric simulations. Confidence intervals that include the value "zero" denote no statistically significant difference between the means.

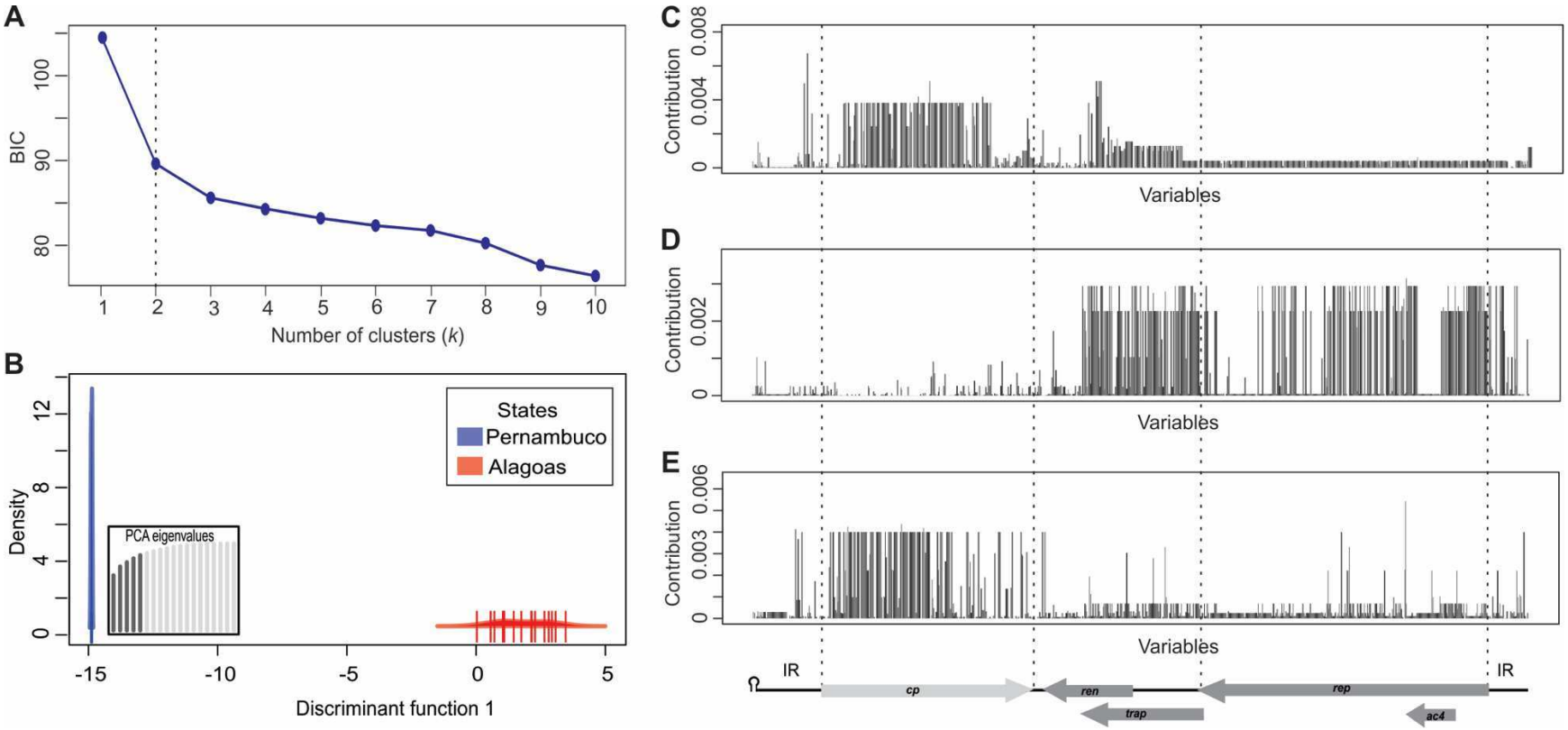
Supplementary Figure S4. Statistical significance of the difference between the average pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated for nucleotide sequences of the CP (capsid protein), Rep (replication-associated protein), REn (replication enhancer protein), TrAP (trans-activating protein), AC4, MP (movement protein) and NSP (nuclear shuttle protein) genes, the DNA-A intergenic region (IR-A), DNA-B large intergenic region (LIR-B) and DNA-B small intergenic region (SIR-B) for 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) data sets. Ninety-five percent bootstrap confidence intervals (CIs) for the difference between π values were estimated from 1,000 nonparametric simulations. Confidence intervals which include the value "zero" denote no statistically significant difference between the means. Comparisons where DNA-A genes are more variable than DNA-B genes are highlighted in red. Comparisons where there is no difference or DNA-B genes are more variable than DNA-A genes are highlighted in green.

Supplementary Figure S5. Statistical significance of the difference between the average pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated for the 5'-terminal, central and 3'-terminal regions of the CP (capsid protein), Rep (replication-associated protein), MP (movement protein) and NSP (nuclear shuttle protein) genes, the DNA-A intergenic region (IR-A) and the DNA-B large intergenic region (LIR-B) for 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) populations. Ninety-five percent bootstrap confidence intervals (CIs) for the difference between π values were estimated from 1,000 nonparametric simulations. Confidence intervals which include the value "zero" denote no statistically significant difference between the means.

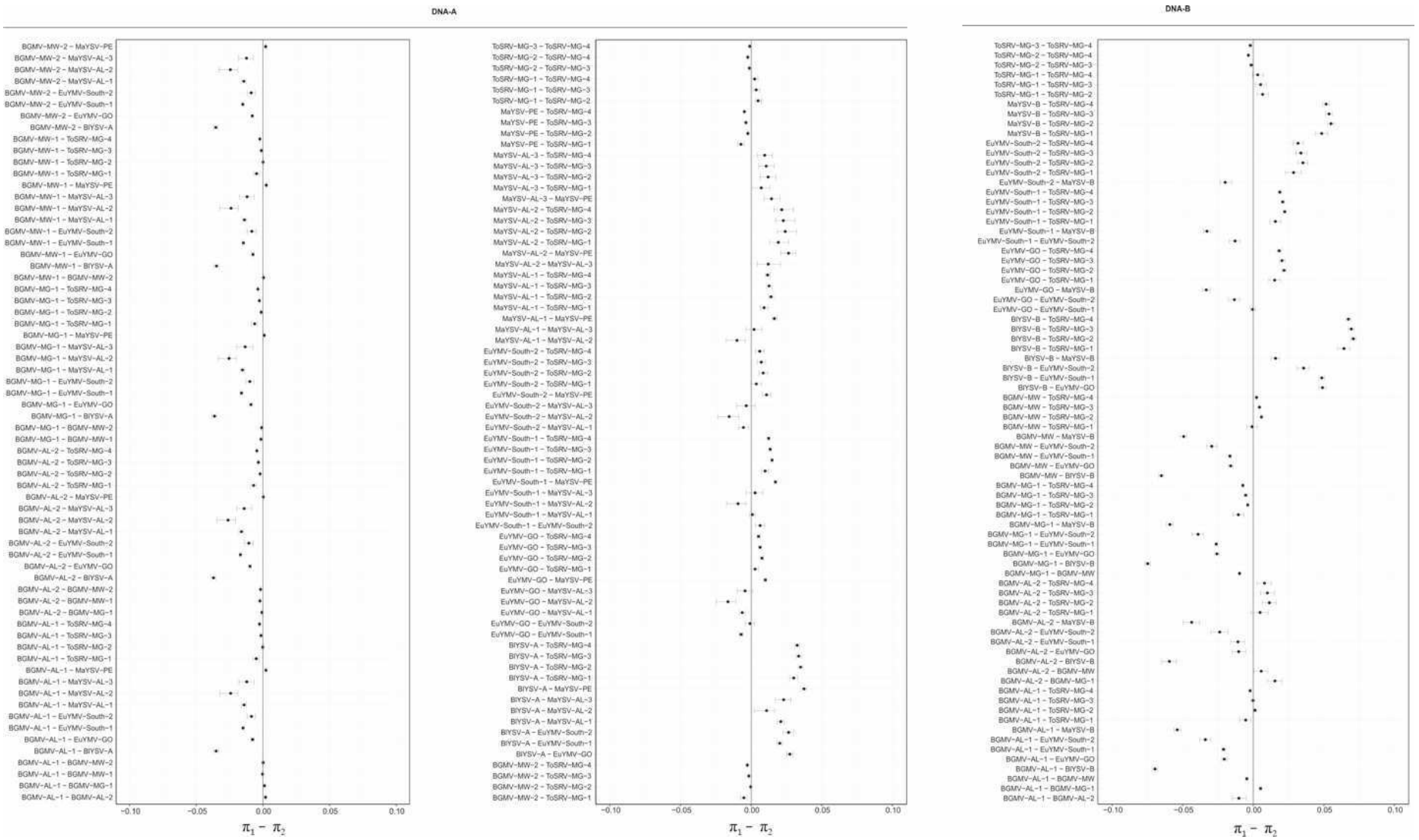
Supplementary Figure S1



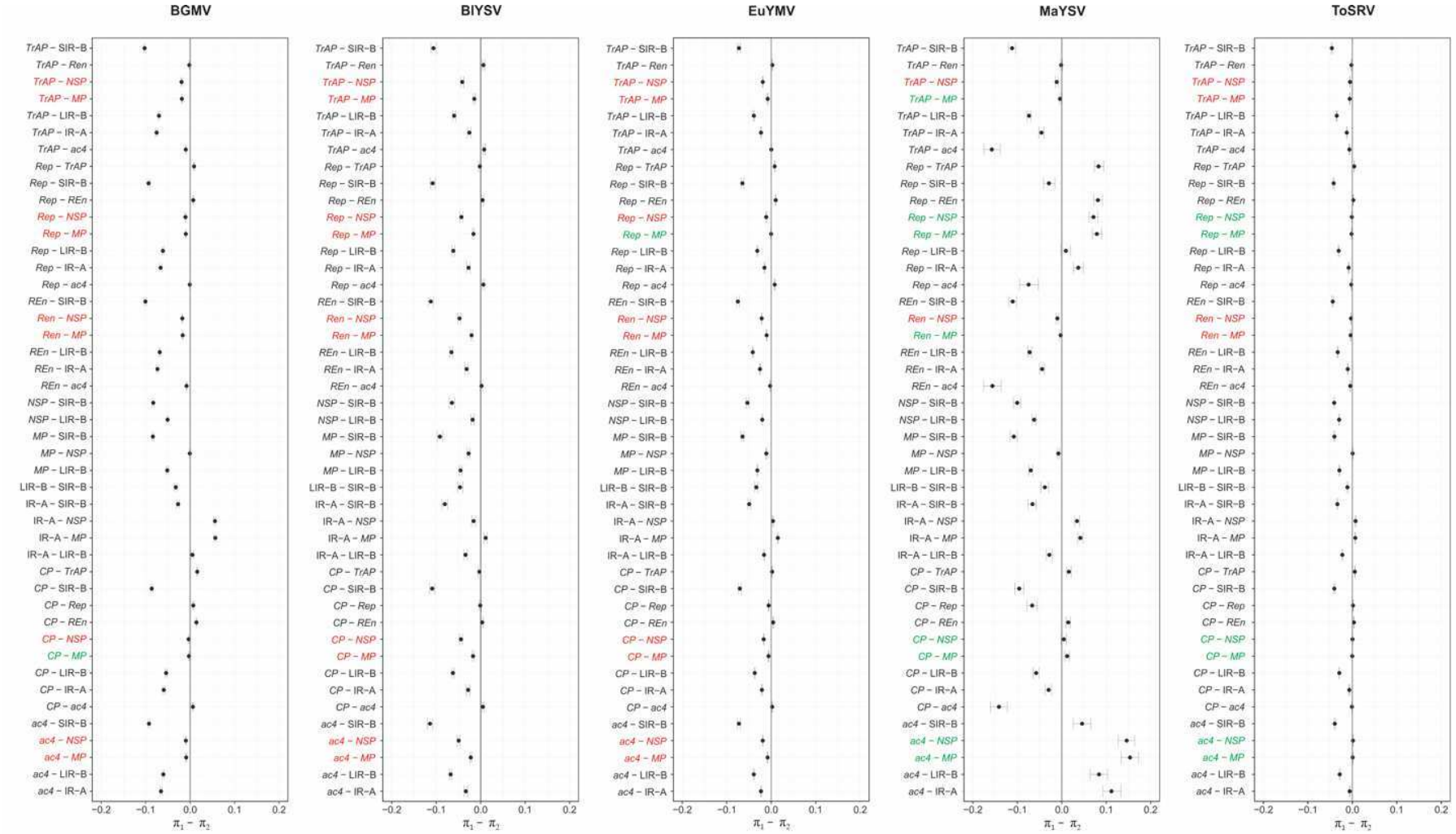
Supplementary Figure S2



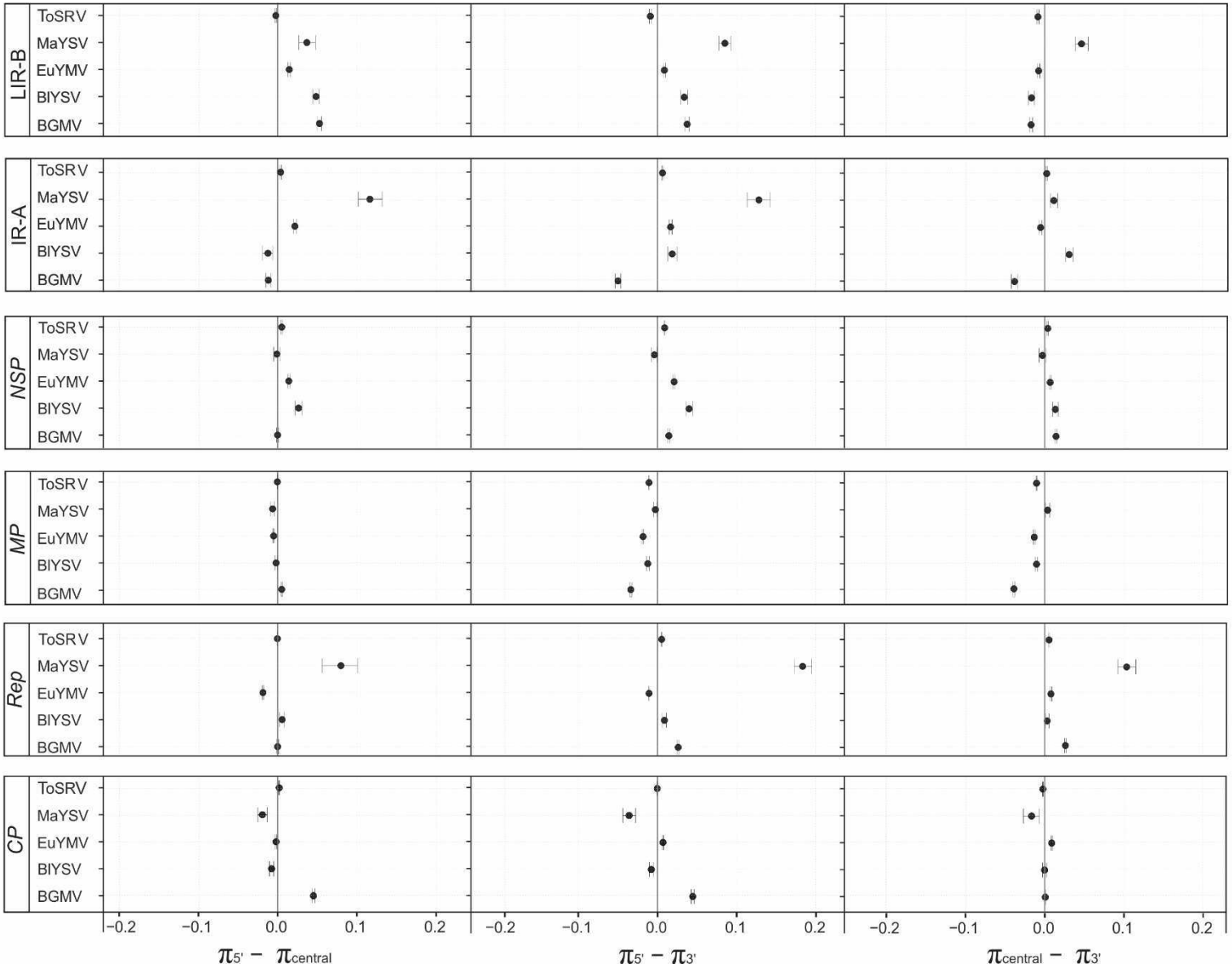
Supplementary Figure S3



Supplementary Figure S4



Supplementary Figure S5



CHAPTER 2

FUNCTIONAL ANALYSIS OF *Arabidopsis thaliana* CYTIDINE DEAMINASE (AtCDA) AND CYTIDINE DEAMINASE-like 4 (AtCDAL4) REVEAL OPPOSITE ROLES UPON BEGOMOVIRUS INFECTION

Xavier, C.A.D., Shen, W., Ascensio-Ibáñez, J.T., Hanley-Bowdoin, L., Zerbini, F.M. Functional analysis of Arabidopsis thaliana CYTIDINE DEAMINASE (AtCDA) and CYTIDINE DEAMINASE-like 4 (AtCDAL4) reveal opposite roles upon begomovirus infection. Virology (in preparation).

Functional analysis of Arabidopsis thaliana CYTIDINE DEAMINASE (AtCDA) and CYTIDINE DEAMINASE-like 4 (AtCDAL4) reveal opposite roles upon begomovirus infection

César A.D. Xavier[1], Wei Shen[2], José T. Ascencio-Ibáñez[3], Linda Hanley-Bowdoin[2*], F. Murilo Zerbini[1*]

[1] Departamento de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil.
[2] Department of Plant and Microbial Biology, North Carolina State University, Raleigh NC, 27695 USA.
[3] Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh NC 27695, USA.

*Corresponding author: Linda Hanley-Bowdoin
Phone: (+1-919) 515-6663
E-mail: lkhanley@ncsu.edu

*Corresponding author: F. Murilo Zerbini
Phone: (+55-31) 3612-2423
E-mail: zerbini@ufv.br

Abstract

Cytidine deaminases are part of an important branch of antiviral innate and adaptative immunity in animals. These proteins act mainly through the deamination of cytidine to uridine, which leads to a high accumulation of hypermutated nonfunctional viral genomes, driving the viral population to extinction. However, sub-lethal mutations may have beneficial effects by promoting viral diversification, thus increasing adaptive potential. Although plants also code for cytidine deaminases, their role during virus infection has not been studied. A mutational bias has been observed in begomovirus genomes, favoring C to T and G to A transitions, suggesting that cytidine deamination may act on these viruses. Here, we investigated the role of A. thaliana cytidine deaminases (AtCDA and AtCDAL4) as components of plant antiviral immunity. We demonstrate that 14 days after Cabage leaf curl virus (CabLCV) inoculation, AtCDA expression was highly induced, whereas AtCDAL4 was strongly repressed. While AtCDA was modulated only during CabLCV infection, AtCDAL4 was also repressed during African cassava mosaic virus infection. Overexpression of AtCDA and AtCDAL4 did not affect CabLCV accumulation during the early stages of infection. Analysis of the viral mutational spectra generated during overexpression of AtCDA and AtCDAL4 does not support a mutagenic activity on the CabLCV genome. Contrary to the tested hypothesis, knockout of AtCDA led to delayed symptom development and reduced accumulation of CabLCV in mutant plants. No such effects were observed during Beet curly top virus and Turnip mosaic virus infections, suggesting a specific proviral effect for CabLCV. Interestingly, we found evidence for a putative antiviral role of AtCDAL4, as its knockout positively affected CabLCV infection, providing an increase in viral load and symptoms. Overall, our results demonstrate that different AtCDAs, even those previously considered to be pseudogenes, may have evolved distinct functions upon virus infection. Furthemore, unlike in animals, our results suggest that AtCDAs affect viral infection independently of deaminase activity. Further work is necessary to elucidate the mechanisms by which AtCDA and AtCDAL4 affect begomovirus infection in opposite ways.

Introduction

The genus Begomovirus (family Geminiviridae) includes viral species with one or two genomic components of circular, single-strand DNA (ssDNA) packaged into geminate icosahedral particles and transmitted by whiteflies of the Bemisia tabaci cryptic species complex (Zerbini et al., 2017). It has been suggested that the high evolutionary rates of begomoviruses (as well as those of other geminiviruses) could be partially due to mutational processes acting specifically on ssDNA (Duffy and Holmes, 2008; Harkins et al., 2009). Although geminiviruses use the host replication machinery, which in theory would correct the improper incorporation of nucleotides, their genomes may not have the methylation pattern required for the host's excision repair mechanism to act (Sanz et al., 1999). It is also possible that their genomes remain for most of the infection cycle as ssDNA, thus inaccessible to the host's DNA proof-reading machinery (Sanz et al., 1999). Finally, it has been proposed that ssDNA genomes are particularly susceptible to oxidative stresses, which may occur spontaneously or mediated by host enzymes, increasing the basal mutation rate (Duffy and Holmes, 2008). Although there are several possible explanations for the high evolutionary rates of begomoviruses, these are not mutually exclusive, and the combination of different mechanisms may be a more likely explanation. In any event, studies trying to clarify the molecular mechanisms responsible for these high evolutionary rates have not yet been performed.

Interestingly, a mutational bias has been detected in begomovirus genomes, favoring C to T and G to A transitions (Duffy and Holmes, 2008; Duffy and Holmes, 2009; Ge et al., 2007). These two transitions may occur due to base deamination (cytosine to uracil and guanine to xanthine), suggesting that base deamination may contribute to the high evolutionary rates observed for begomoviruses. These biases are also observed in Human immunodeficiency virus (HIV) (Zhang et al., 2003), Hepatitis B virus (HBV) (Lei et al., 2006), and several other viruses that infect animals (Harris and Dudley, 2015). This is consistent with the presence of specifics enzymes that act on the deamination of nucleic acid (ssDNA and ssRNA), such as protein of the APOBEC family (Moris et al., 2014; Salter et al., 2016).

The APOBEC (apolipoprotein B mRNA editing enzyme, catalytic) family of proteins comprises a group of cytidine deaminases capable of editing DNA and RNA of both hosts and pathogens (Moris et al., 2014; Salter et al., 2016). These proteins are associated with several cell functions, including innate and adaptive immune responses against viruses and retroelements (Moris et al., 2014). The family is composed of 11 members: AID (activation

induced deaminase), APOBEC1, APOBEC2, APOBEC3 (A, B, C, DE, F, G and H) and APOBEC4. A role in antiviral immunity for members of the APOBEC family has been demonstrated against different groups of viruses (Fan et al., 2010; Lei et al., 2006; Sheehy et al., 2002; Suspene et al., 2011).

APOBEC proteins use different mechanisms to inhibit viral infection (Chowdhury et al., 2013; Holmes et al., 2007; Zhang et al., 2003). During reverse transcription of the HIV-1 genome, APOBEC3G induces deamination of cytidine to uridine in the minus-strand ssDNA, leading to G/A hypermutation in viral genomic DNA (Zhang et al., 2003). This hypermutation in the HIV-1 genome does not occur randomly throughout the genome, but rather at hotspots defined based on sequence and secondary structure (Holtz et al., 2013; Kim et al., 2014). APOBEC-directed hypermutation in coding regions of the HIV-1 genome introduces amino acid substitutions and premature stop codons, leading to the production of non-infectious viral genomes. In addition, the action of uracil DNA glycosylase may lead to the degradation of the viral cDNA with high levels of uridine (Chowdhury et al., 2013; Weil et al., 2013). APOBEC3G/F may also act independently of its deaminase activity through binding on genomic RNA, inhibiting replication (Belanger et al., 2013). On the other hand, APOBEC3G can be suppressed by the HIV-1 Vif protein through its degradation via the 26S proteasome (Conticello et al., 2003; Henriet et al., 2009). It is now accepted that APOBEC3G and other members of the APOBEC3 subgroup, such as APOBEC3F, constitute a branch of innate antiviral immunity in animals (Harris and Dudley, 2015; Liddament et al., 2004; Moris et al., 2014; Stavrou and Ross, 2015).

Nonetheless, hypermutation mediated by residual APOBEC activity may have a beneficial role by promoting viral population diversification, thus increasing adaptive potential (Conticello et al., 2003; Cuevas et al., 2015). HIV-1 has high levels of genetic variability, with an extraordinary capacity to develop genetically complex populations from a limited number of infectious particles (Cuevas et al., 2015; Smyth et al., 2012). A study quantifying the rate of spontaneous in vivo mutations in the HIV-1 genome showed that HIV-1 reverse transcriptase contributes to only 2% of the mutations, with 98% resulting from editing by host cytidine deaminases of the APOBEC3 family (Cuevas et al., 2015). Thus, it seems that HIV-1 has evolved ways to escape from APOBEC3-mediated immunity, turning the mechanism in its favor (Conticello et al., 2003). The degradation of APOBEC3G by Vif is not complete, allowing a few APOBEC3G subunits to be incorporated into the virion. These subunits can promote genome editing at low rates and insert sublethal mutations, which in turn may promote the diversification of HIV-1 populations. Such mutations have the potential to contribute

significantly to HIV-1 evolution, pathogenesis, immune response evasism and drug resistance (Borzooee et al., 2019; Kim et al., 2014; Sadler et al., 2010).

The APOBEC family belongs to a larger family of cytidine deaminases (CDAs) shared by a diverse group of organisms (Conticello et al., 2005; Liu et al., 2018). Plants such as Arabidopsis thaliana and Oryza sativa have been shown to encode several CDAs (Chen et al., 2016; Faivre-Nitschke et al., 1999; Kafer and Thornburg, 2000; Liu et al., 2019). Sequencing of the A. thaliana genome revealed the presence of a nine-member protein family. Bioinformatic analysis of plant CDA sequences combined with biochemical tests indicated that eight A. thaliana CDA family members are pseudogenes (named AtCDA-like 1 to 8), whereas only one, named AtCDA, is completely functional and required to maintain pyrimidine homeostasis in vivo (Chen et al., 2016). In addition, it was shown that AtCDAL4 retains a very low deaminase activity, but is not involved in pyrimidine metabolism (Chen et al., 2016). Molecular characterization showed that AtCDA encodes a typical cytidine deaminase more closely related to prokaryotic CDAs than to eukaryotic enzymes (Faivre-Nitschke et al., 1999; Kafer and Thornburg, 2000). In addition, AtCDA has similar enzymatic and substrate specificities as conventional cytidine deaminases and is competitively inhibited by cytosine-containing compounds (Chen et al., 2016; Faivre-Nitschke et al., 1999). It has been proposed that AtCDA is not involved in the editing of RNA because it does not show affinity for RNA (Faivre-Nitschke et al., 1999).

The possible involvement of plant CDAs in immune responses against different pathogens has been suggested (Carviel et al., 2009; Gowda et al., 2007; Martin et al., 2016). Gowda et al. (2007) demonstrated that O. sativa CDA expression (locus Os07g14150) was highly induced after inoculation with Magnaporthe grisea, resulting in a high accumulation of A to G and U to C mutations in mRNAs encoding defense-related genes. In addition, AtCDA was identified and required as a component of age-related resistance (ARR) response of A. thaliana to Hyaloperonospora parasitica (Carviel et al., 2009). AtCDA knockout plants were more susceptible to this pathogen, suggesting its involvement in immune responses (Carviel et al., 2009). Finally, it was demonstrated that transient overexpression of AtCDA significantly increased the number of G to A mutations in the Cauliflower mosaic virus (CaMV) genome, suggesting that deamination of viral genomes may also take place in plants (Martin et al., 2016). These results suggest that plant CDAs may have a broader activity spectrum compared to animal APOBECs. Nevertheless, the involvement of plant CDAs as a general component of antiviral defense responses is unknown (Moris et al., 2014). No studies have been carried out to explore the molecular mechanisms involved in base deamination and its possible

involvement as a branch of antiviral innate immunity, or alternatively as an accelerator of the evolution of begomoviruses.

Here, using reverse genetics, we demonstrate distinct functional roles for A. thaliana CDAs upon begomovirus infection. While AtCDA has a proviral effect during Cabbage leaf curl virus (CabLCV) infection, AtCDAL4 has antiviral activity, constituting a putative branch of antiviral immune responses in A. thaliana.

Methods

Plant material and growth conditions

Wild type (WT) A. thaliana ecotypes Columbia (Col-0) and Seis am Schlern (Sei-0) were used in the experiments. T-DNA insertion lines cda1 (GK645H07), cda2 (SALK036597) and the complementation line cda1::CDA-YFP have been described (Chen et al., 2016). The T-DNA insertion line cdal4 (SALK201914) was obtained from the Arabidopsis Resource Center (ABRC, Ohio State University). All mutants described above are in the Col-0 background. To test homozygosity, a small segregating population (20 to 30 plants for each line) was screened by PCR using different combinations of primers (Suppl. Table S1). The mutant Sei-0 RICIb, carrying a copy of Segs-I (sequence enhancing geminivirus symptoms - I) will be described elsewhere. A. thaliana seeds were grown in sterile soil in a growth chamber at 20^{o}C under a photoperiod of 8 hours light/16 hours dark. Seedlings were transplanted to 24-well trays aproximately 10 days after sowing and used for virus inoculation two weeks post-transplanting (10 to 12-leaf stage).

Nicotiana benthamiana plants were grown in sterile soil in 18-well trays in a growth chamber at 22^{o}C under a photoperiod of 12 hours light/12 hours dark. Plants were used for agroinfiltration experiments aproximatly 30 days after sowing. Solanum lycopersicum cv. Florida Lanai plants were cultivated in 18-well trays in a growth chamber at 24^{o}C under a photoperiod of 12 hours light/12 hours dark. Plants displaying 2 to 4 true expanded leaves were used for inoculation.

Viral isolates and inoculation procedures

Agroinfectious clones corresponding to the genomic component(s) of the begomoviruses CabLCV, Tomato mottle virus (ToMoV) and Tomato yellow leaf curl virus (TYLCV) and of the curtovirus Beet curly top virus (BCTV) have been previously described (Rajabu et al., 2018). An agroinfectious clone of the potyvirus Turnip mosaic virus (TuMV)

was described by Cotton et al. (2009). An infectious clone of the begomovirus African cassava mosaic virus (ACMV) was described by Ndunguru et al., (2016). An infectious clone of the caulimovirus Cauliflower mosaic virus isolate pW260 (CaMV-pW260) was described by Scholelz and Shepherd (1988).

Agroinoculation was performed as previously described (Rajabu et al., 2018; Reyes et al., 2013). Briefly, Agrobacterium tumefaciens cultures were growth overnight at 30ºC in Luria Bertani (LB) media containing the appropriate antibiotics and directly inoculated according to the treatments described below. For bipartite begomoviruses, equal amounts of DNA-A and DNA-B cultures were mixed immediately before inoculation. A. thaliana plants were gently punctured with a needle in the middle of the rosette, while tomato plants were punctured in the stem below the apex and a few drops of culture were injected into the wound. Trays were kept covered for two days after inoculation, and then plants were maintained as described above.

For biolistic inoculation, Escherichia coli cultures carrying the ACMV (DNA-A and DNA-B) and CaMV-pW260 infectious clones were grown overnight at 37ºC in LB media and plasmid DNA was extracted using the Qiagen Plasmid Midi kit following the manufacturer's instructions. For ACMV, 5 μg of each DNA component were used to inoculate six plants. For CaMV, the plasmid pW260 was previously digested with SalI (New England Biolabs, NEB) releasing the complete linearized genome, and the digestion reaction containing 6 μg of DNA was used for inoculation. A. thaliana plants were inoculated as previously described (Rajabu et al., 2018). Negative controls were inoculated with gold particles without DNA.

RNA extraction, cDNA synthesis and quantitative PCR (qPCR)

Total RNA from biological replications according to each treatment were extracted using the RNeasy Plant Mini Kit (Qiagen), according to the manufacturer's instructions. In addition, on-column DNase I treatment was performed using RNase-free DNase (Qiagen) following the manufacturer's instructions. RNA integrity was evaluated by electrophoresis in 1.0% agarose gels and quantified by spectrophotometry using a NanoDrop 2000 (ThermoScientific). An aliquot containing 1.5 microgram of total RNA was employed for cDNA synthesis using the SuperScript III First-Strand Synthesis System (Invitrogen) following the manufacturer's instructions. cDNA was diluted 20x in H_2O prior to relative quantification by qPCR. Reactions consisted of 5 μl of diluted cDNA, 5 μl of mixed primers containing 2 picomole/μl of each primer (Suppl. Table S1) and 10 μl of PowerUp SYBR Green Master Mix (ThermoFisher) in a final volume of 20 μl. The qPCR cycles consisted of an initial denaturing step at 95°C for 2 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. All

reactions were performed in triplicates, and a coefficient of variation of 1% was used as a cutoff to verify the reliability of the technical replications. Reference genes used to normalize the qPCR data are described below. Expression levels for each gene were quantified by the comparative cycle threshold method ($2^{-\Delta\Delta CT}$) (Livak and Schmittgen, 2001) using specific primers for each gene (Suppl. Table 1). For all reactions, amplification specificity was checked based on dissociation curves. In addition, qPCR products were also visualized in 1.5% of agarose gels. Amplification efficiency was evaluated based on the slopes of amplification curves per cycle number (Ramakers et al., 2003).

Kinetics of AtCDA expression upon virus infection

Relative expression levels of AtCDA and AtCDLA4 mRNA were analyzed by qPCR in A. thaliana Col-0 infected by BCTV, CabLCV, TuMV at 16 days post-inoculation (dpi). Mock-inoculated plants were inoculated with empty vector. Plants were sampled in three biological replications, each one comprised of pools of three to four individual plants from which the third, fourth and fifth leaves were collected. AtCDAL3 expression was analyzed during CabLCV infection at 16 dpi (Chen et al., 2016).

To verify the expression kinetics of AtCDA and AtCDAL4 mRNA over time, a time course experiment was conducted during CabLCV infection. Samples were collected at 7, 14, 21 and 28 dpi as described above, and four biological replications were analyzed. Total DNA was extracted, and virus accumulation was quantified over time by qPCR (described below). For relative quantification, AtSAND was used as a reference gene.

AtCDA and AtCDAL4 expression levels were also analyzed during infection by TYLCV and ACMV. A. thaliana Col-0 was agroinoculated with TYLCV, while ecotype Sei-0 and the transgenic line Sei-0 RICIb were used for the experiment with ACMV. In both experiments, plants inoculated with CabLCV were used as controls. Three to four biological replications comprised of pools of three individual plants were analyzed at 21 dpi. Virus accumulation was analyzed (as described below) using the same samples. For relative quantification, AtSAND was used as a reference gene.

The expression of Solanum lycopersicum CYTIDINE DEAMINASE 1 (SlCDA1) was analyzed in plants infected by TYLCV and ToMoV at 21 dpi. Three and five individual plants were analyzed for TYLCV and ToMoV, respectively. For relative quantification, SlACTN was used as a reference gene (Yang et al., 2019).

Cloning of the AtCDA and AtCDAL4 coding sequences

For transient overexpression assays, full-length AtCDA and AtCDAL4 coding sequences were amplified from A. thaliana Col-0 cDNA with specific primers for each gene (Suppl. Table S1) using Q5 High-Fidelity 2X Master Mix DNA Polymerase (NEB), following the manufacturer's instructions. The AtCDA coding sequence was first inserted into pDONR201 using the Gateway BP Clonase II Enzyme mix (Invitrogen), originating pDONR201-AtCDA. The AtCDAL4 coding sequence was inserted into pENTR/TEV/D-TOPO (Invitrogen), resulting in pENTR/TEV/D-AtCDAL4. These clones were sequenced bidirectionally to check sequence integrity. The sequences for both genes were transferred by homologous recombination using LR Clonase II Enzyme mix (Invitrogen) to the destination binary vector pEarlyGate201 (Earley et al., 2006), resulting in the constructs pEarlyGate201-AtCDA and pEarlyGate201-AtCDAL4. Similarly, pENTR-gus, carrying the A. thaliana β-glucuronidase (GUS) gene, provided with the LR kit, was used to transfer the GUS coding sequence to pEarlyGate201 resulting in the pEarlyGate201-gus. Recombinant plasmids were initially transformed into 5-alpha Competent E. coli (NEB) and subsequently into A. tumefaciens strain GV3101. All cloning procedures were performed according to the manufacturer's instructions.

Transient expression of AtCDA and AtCDAL4

A. tumefaciens transformed with the constructs pEarlyGate201-AtCDA, pEarlyGate201-AtCDAL4, pEarlyGate201-gus, and the agroinfections clones corresponding to CabLCV (DNA-A and DNA-B) and TuMV, were grown overnight at 30^{o}C in LB media containing appropriate antibiotics for each construct and 40 μM acetosyringone to an OD_{600} of 0.6. The cultures were centrifuged for 10 min at 4,000 rpm, 4^{o}C and resuspended in 10 mM MES pH 5.5, 10 mM $MgCl_2$, and 200 mM acetosyringone. A. tumefaciens suspensions carrying different constructs were mixed according to each treatment to a final OD_{600} of 1 for AtCDA, AtCDLA4 and GUS constructs, and 0.1 for viruses. The cultures were incubated for 2 hours at room temperature before agroinfiltration. N. benthamiana leaves (approximately four weeks post-germination) were infiltrated using sterile syringes without needles. From each plant, three leaves were agroinfiltrated at four to five spots each one. Each agroinfiltrated leaf was collected at 48, 72 and 96 hours post-agroinfiltration. Three to four agroinfiltrated spots from each leaf were pooled and used for DNA extraction. Absolute CabLCV DNA accumulation was quantified by qPCR using virus-specific primers as described below. For TuMV, samples were submitted to RNA extraction and cDNA synthesis (described above), and virus accumulation

was quantified by relative qPCR (described above) using NbH2B as a reference gene (Bruckner et al., 2017).

Viral accumulation in A. thaliana knockout plants

A. thaliana knockout lines for AtCDA (cda-1 and cda-2), the complementation line cda-1+ AtCDA-YFP and wt Col-0 were agroinoculated with CabLCV as described above, and viral absolute quantification was performed at 12 dpi. The experiment was performed twice. The effect of the AtCDA knockout was also tested upon infection by BCTV and TuMV at 14 dpi, and upon CaMV infection at 8 and 14 dpi. All experiments were sampled in biological replications comprised by pools of three individuals plants and the third, fourth and fifth leaves were collected.

For CabLCV, BCTV and CaMV, total DNA was extracted as described by Doyle and Doyle (1987) and used to determine absolute viral accumulation by qPCR using virus-specific primers (Suppl. Table S1). DNA samples were quantified using the Qubit dsDNA BR Assay kit in a Qubit 3.0 fluorometer (Life Technologies) and diluted to 2 ng/µl. Reactions consisted of 5 µl of diluted DNA, 5 µl of mixed primers containing 2 picomole/µl of each primer (Suppl. Table S1) and 10 µl of PowerUp SYBR Green Master Mix (ThermoFisher) in a final volume of 20 µl. The qPCR cycles consisted of an initial denaturing step at 95°C for 2 min followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. All reactions were performed in triplicates and a standard curve was ran in each plate. Standard curves were prepared using serial dilutions of plasmid DNA containing the complete DNA-A of CabLCV and the complete genomes of BCTV and CaMV ($3x10^2$ to $3x10^8$ copies per reaction). Standard curves were obtained by regression analysis of cycle threshold (Ct) values of three replications of a given dilution in relation to the log of the amount of DNA in each dilution. For all reactions, amplification specificity was checked based in dissociation curves.

For TuMV, samples were submitted to RNA extraction and cDNA synthesis, and virus accumulation was quantified by relative qPCR using AtSAND as a reference gene. All procedures were performed as described above.

The effect of the knockout of AtCDAL4 was also tested upon CabLCV infection at 14 dpi. A. thaliana Col-0 and the knockout line cdal4 were agroinoculated and absolute viral accumulation were performed at 14 dpi. All procedures for viral quantification were performed as described above. Symptom development was evaluated according to the scale proposed by Shen et al. (2014): 1, no symptoms; 2, mild leaf curling; 3, chlorosis in 1 to 2 leaves; 4, chlorosis spread to 3 or 4 curly leaves; 5, severe symptoms throughout new growth in the rosette.

Differential DNA denaturing PCR (3D-PCR)

To verify a possible mutagenic effect of AtCDA and AtCDAL4 on the begomovirus genome, a differential DNA denaturing PCR (3D-PCR) assay was performed. To avoid amplifying non-replicated DNA, a divergent primer pair (Suppl. Table S1) that amplifies only replicated DNA was used. A fragment of approximately 900 nt comprising the 5' region of the CP gene, the entire intergenic region and the 5' region of the Rep gene was amplified. DNA samples from the transient overexpression experiment were used as templates. Samples collected 96 hpa from two biological replications were used for each treatment. 3D-PCR was performed using Hot Start Taq DNA Polymerase (NEB). The qPCR cycles consisted of a denaturing gradient step at 85.8^{o}C, 85.7^{o}C, 85.5^{o}C, 85.2^{o}C, 84.9^{o}C, 84.7^{o}C, 84.5^{o}C and 84.3^{o}C for 5 min, followed by 30 cycles at the same gradient temperature for 30 sec, 55°C for 1 min and 65°C for 1 min. PCR products were visualized in 1.5% agarose gels. PCR products obtained with the lowest denaturing temperature (84.9^{o}C) were purified from the agarose gel using the QIAquick PCR Purification kit (Qiagen), inserted into pMiniT2.0 plasmid using the PCR Cloning kit and transformed into 5-alpha competent E. coli (NEB) following the manufacturer's instructions. Plasmid DNA from colonies previously confirmed to carry the insert were extracted using the QIAprep Spin Miniprep kit (Qiagen) and sequenced in both directions at Eton Bioscience (Research Triangle Park, North Carolina). The number of mutations was counted and compared with those from plants overexpressing GUS.

Ploidy analyses

To verify whether there was any difference in DNA ploidy levels among A. thaliana knockout lines (cda1 and cda2), the complementation line (cda1::CDA-YFP) and wt Col-0, leaf samples were collected from healthy plants. Three biological replications, each one composed of three plants, were analyzed at 21 dpi. The procedures were performed as described by Rajabu et al. (2018).

Statistical analyses

To verify the significance of gene expression levels among treatments, the data were initially verified for homogeneity of variance and normality using Bartlett and Shapiro-Wilk tests, respectively. Differences between means were tested by Student's t-test implemented in R software (R Development Core Team, 2007). Welch's correction was applied for data that presented unequal variance homogeneity. Viral loads and ploidy levels were compared using ANOVA followed by post hoc multiple comparison Tukey's test implemented in the easyanova

package in R software. For symptom scores, comparison among treatments were performed using a non-parametric Wilcoxon rank sum test implemented in R software.

Results

Expression of AtCDA and AtCDAL4 mRNAs is differentially modulated during CabLCV infection

The expression of both AtCDA and AtCDAL4 was strongly modulated at 16 days after CabCLV inoculation (Figure 1A,B). Interestingly, whereas AtCDA was highly induced, AtCDAL4 was strongly repressed. Furthermore, while AtCDA expression was not altered in BCTV- and TuMV-infected plants (Figure 1A), AtCDAL4 was induced during TuMV infection, although at a much lower level than in CabLCV-infected plants (Figure 1B). These results suggest a specific role of AtCDA during begomovirus infection and partially rule out the possibility of an indirect effect in its expression due to any disturbance in pyrimidine nucleotide pool, as expression was not affected in BCTV and TuMV-infected plants. In addition, they suggest a putative role for AtCDAL4 (considered to be a pseudogene) during virus infection.

To verify whether the effects observed in CabLCV-infected plants were specific or a general effect on A. thaliana CDAs, we also analyzed the expression of AtCDAL3, a truly non-functional CDA, at least regarding its activity as a deaminase (Chen et al., 2016). No differences were observed in AtCDAL3 expression between treatments (Figure 1C), indicating a specific effect on AtCDA and AtCDAL4.

We next carried out a time course experiment to verify AtCDA and AtCDAL4 expression patterns upon CabLCV-infected plants at 7, 14, 21 and 28 dpi. AtCDA and AtCDAL4 modulation was observed from 14 dpi and strongly increased up to 28 dpi, the last time point analyzed (Figure 1D, E). While AtCDA was strongly induced, increasing over time (Figure 1D), AtCDAL4 was strongly repressed (Figure 1E). We also quantified virus accumulation in the same samples (Figure 1F). Viral titer greatly increased from 7 to 14 dpi and then was maintained constant up to 28 dpi (Figure 1F). These results suggest a possible correlation between virus accumulation and AtCDAs modulation, as these genes begun to be induced or repressed at 14 dpi, coinciding with the increase in viral load.

AtCDAL4 but not AtCDA seems to be a conserved target modulated during bipartite begomovirus infection

We next examined whether other begomoviruses could also modulate AtCDA and AtCDAL4 expression. First, we analyzed AtCDA and AtCDAL4 expression upon infection by TYLCV, a monopartite Old World (OW) begomovirus. TYLCV-infected plants did not show any symptoms up to the end of the experiment at 40 dpi, whereas CabLCV-infected plants showed symptoms at 21 dpi (Figure 2A). PCR-based detection of TYLCV confirmed the infection at 21 dpi (Figure 2B). However, and consistent with the absence of symptoms, TYLCV accumulated between two and three orders of magnitude lower than CabLCV at 21 dpi (Figure 2C). Regardless, no differential expression was observed for either AtCDA or AtCDAL4 in TYLCV-infected plants (Figure 2D, E). These results suggest that TYLCV may be poorly adaptated to A. thaliana, and thus CDA expression would not be modulated during its infection. Another possibility would be that AtCDA and AtCDAL4 expression may not be responsive to infection by a monopartite begomovirus (in other words, modulation of CDA expression could be mediated by proteins encoded by the DNA-B).

To test this last hypothesis, we analyzed AtCDA and AtCDAL4 expression upon infection by ACMV, another bipartite begomovirus. Due to the low susceptibility of A. thaliana Col-0 to ACMV, we performed all experiments using ecotype Sei-0, which is more susceptible to this virus. We also used Sei-0 RICIb, a transgenic line carrying a complete copy of Segs-I (sequence ehancing geminivirus symptoms - I), which shows an increase in symptom severity induced by ACMV (Figure 3A). Although ACMV-infected Sei-0 RICIb plants displayed much more severe symptoms compared to ACMV-infected wt Sei-0 plants, this difference was not due to a higher viral load, as no significant difference in ACMV accumulation was observed (Figure 3B). In contrast, in both lines, ACMV accumulated significantly less than CabLCV (Figure 3B), although the diference was not as high as that observed for TYLCV. CabLCV accumulated between one and two orders of magnitude higher than ACMV in both wt Sei-0 and Sei-0 RICIb (Figure 3B). Interestingly, while AtCDA was induced only in CabLCV-infected plants, AtCDAL4 was also repressed during ACMV infection in both wt Sei-0 and Sei-0 RICIb (Figure 3C-H). Thus, although we cannot rule out the influence of virus accumulation in AtCDA expression, our results suggest that AtCDAL4 but not AtCDA seems to be a conserved target modulated during bipartite begomovirus infection.

To expand our understanding of the role of CDAs during begomovirus infection, we analyzed the expression of SlCDA, a tomato homolog of AtCDA, during infection by the bipartite ToMoV and the monopartite TYLCV, both well adapted to tomato. The results indicate that SlCDA expression was not changed 21 days after inoculation with either begomovirus compared with mock-inoculated controls (Suppl. Figure S1).

Transient overexpression of AtCDA and AtCDAL4 does not affect CabLCV accumulation

Given that AtCDA and AtCDAL4 expression were highly modulated in CabLCV-infected plants, it would be reasonable to assume that their overexpression could affect viral accumulation. To this purpose, a transient overexpression assay was conducted in N. benthamiana. AtCDA and AtCDAL4 were cloned in binary vectors and overexpressed, and viral load was monitored at 48, 72 and 96 hpa. Overexpression of either AtCDA or AtCDAL4 did not affect CabLCV accumulation, as no significant differences in viral load were observed in the analyzed time points (Figure 4A, Suppl. Figure S2). We also analyzed TuMV accumulation during CDA overexpression, as AtCDAL4 was also induced in TuMV-infected plants. However, there was no difference in TuMV accumulation upon either AtCDA or AtCDAL4 overexpression at 96 hpa (Figure 4B). The accumulation of CDA mRNAs was confirmed through RT-PCR (Figure 4C). The absence of an effect could be due to these proteins acting at later stages of the infection, since their expression is modulated after 14 dpi (Figure 1D, E). Our experiment set up, where the latest time sampled was four days after infiltration, may not allow enough time for these proteins to act.

No evidence for mutagenic activity of AtCDA or AtCDAL4 on the CabLCV genome

A 3D-PCR assay was performed to verify whether CDAs have mutagenic activity on the viral genome. The assay differentiates the presence of genomes rich in A and T mutations through the use of a gradient of denaturing temperature, where hypermutated genomes will be amplified even under the lowest temperatures. A fragment of approximately 900 bp comprising the 5' region of the CP gene, the entire intergenic region and the 5' region of the Rep gene was amplified using a divergent primer pair that amplifies only replicated genomes (Figure 4D). We could not detect any differences in amplification across different treatments (overexpression of AtCDA, AtCDAL4 and GUS) and denaturing temperatures (data not shown), suggesting that mutated genomes would be either absent or present at a very low frequency. Nevertheless, the products obtained at the lowest denaturation temperature were cloned and sequenced. In addition to the very low number of mutations detected (six), no mutated sequences indicative of cytidine deamination (G to A and C to T) were detected (Figure 4D), suggesting that the proteins do not have deaminase activity on the CabLCV genome.

AtCDA positively affects CabLCV infection

To assess a putative role of AtCDA during virus infection we conducted infectivity assays using knockout lines for the AtCDA locus (cda1 and cda2), the complementation line

cda1::AtCDA-YFP and wt Col-0. Plants were agroinoculated with CabLCV and evaluated for symptom development during the course of the experiment. Viral load was verified by qPCR at 12 dpi. To assure that the mutant lines were homozigous, a small segregating population was screened by PCR using different combinations of primers (Figure 5A-C). Interestingly, cda1 and cda2 plants displayed a significant delay in symptom development compared to wt Col-0, whereas symptom progression in cda1::CDA-YFP plants was similar to wt Col-0 (Figure 5D). In addition, the mean number of days required to reach a given symptom score was significantly higher for cda1 and cda2 plants compared to wt Col-0, while no difference was observed between the complementation line and wt Col-0 (Figure 5E). Consistent with the delayed symptom development, viral load in cda1 and cda2 plants was significantly lower compared with wt Col-0 and cda1::CDA-YFP at 12 dpi (Figure 5F). Together, these results indicate that AtCDA is a host factor enhancing the CabLCV accumulation during infection, as the loss of AtCDA function leads to an increase in resistance observed as a reduced viral load. These results are supported by the strong and specific induction of AtCDA expression in CabLCV-infected plants (Figure 1A,D).

As AtCDA is involved in the pyrimidine salvage pathway and its loss of function affects the cell's nucleotide poll (Chen et al., 2016), CabLCV accumulation could be indirectly affected. If this is true, other viruses should also be affected. To test this possibility, the effect of loss of AtCDA function was assessed during infection by BCTV and TuMV. Mutant lines (cda1 and cda2), the complementation line cda1::CDA-YFP and wt Col-0 were agroinoculated and virus accumulation was analyzed 14 dpi by qPCR. While significant lower accumulation was again observed for CabLCV infecting cda1 and cda2 plants at 14 dpi, no differences in BCTV and TuMV accumulation were detected (Suppl. Figure S3), suggesting a direct and specific effect upon CabLCV infection.

AtCDAL4 exhibits antiviral activity against CabLCV

AtCDAL4 has been considered a pseudogene due to its low deaminase activity. Nevertheless, it is strong downregulating in plants infected by two different begomoviruses, raising the question of a possible function during the begomovirus infection cycle. To assess a putative effect of AtCDAL4 on CabLCV infection we obtained a homozygous T-DNA insertion line (cdal4) for AtCDAL4 (Figure 6A, B) and an infectivity assay was conducted. The cdal4 line and wt Col-0 plants were agroinoculated and symptoms and viral load were evaluated at 14 dpi. Interestingly, cdal4 plants displayed more severe symptoms compared to wt Col-0 at 14 dpi (Figure 6C, D). In addition, CabLCV accumulated at a higher titer in cdal4 compared to

wt Col-0 plants. Together, these results suggest a role for AtCDAL4 in antiviral immunity against CabLCV. Added to the fact that AtCDAL4 was downregulated also during ACMV infection (Figure 3G), the possibility of a broader activity against other begomoviruses merits further studies.

CaMV infection induces AtCDA expression but its knockout does not affect viral accumulation

An antiviral role of AtCDA against CaMV has been previously suggested (Martin et al., 2016). To check for this possibility under our experimental set up and to verify whether AtCDA may have evolved distinct functions during infection by non-related viruses such as CabLCV and CaMV, we conducted an infectivity assay using knockout lines for AtCDA (cda1 and cda2), the complementation line cda1::AtCDA-YFP and wt Col-0. Plants were inoculated with CaMV and CDA expression, symptom development and viral load were evaluated. AtCDA expression was positively regulated during CaMV infection (Figure 7A), similar to what was observed for CabLCV (Figure 1A). In contrast, CaMV infection did not affect AtCDAL4 expression (Figure 7B). Virus-inoculated mutant lines cda1 and cda2 displayed similar symptoms to those observad in wt Col-0 and the complementation line (Figure 7C). Furthemore, no differences in virus accumulation were detected at 8 and 14 dpi in mutants lines compared to wt Col-0 and the complementation line (Figure 7D,E). Thus, under our experimental conditions, loss of AtCDA function did not contribute to increased CaMV accumulation, which would be expected should this protein have antiviral activity against this virus.

Discussion

Cytidine deamination is a conserved antiviral mechanism in mammals (Moris et al., 2014; Salter et al., 2016). The presence of proteins containing a catalytic deaminase domain, shared by highly divergent groups of eukaryotes such as Chordata, Conosa and algae, suggests an ancestral relationship among mutagenesis-based defense systems (Krishnan et al., 2018). Although plants also code for CDAs (Chen et al., 2016; Liu et al., 2019), their role as a branch of the immune system is poorly understood. A possible involvement in defense response of CDAs encoded by rice and A. thaliana has been suggested during infection by oomycetes, fungi and viruses (Carviel et al., 2009; Gowda et al., 2007; Martin et al., 2016). Here, we verified a possible role of A. thaliana cytidine deaminases (AtCDA and AtCDAL4) upon begomovirus

infection and obtained strong evidence that these genes code for proteins with distinct functions in the viral life cycle.

Several key cell components, either mediating defense or as a proviral host factor, have their expression modulated during viral infection (Alfenas-Zerbini et al., 2009; Ascencio-Ibanez et al., 2008). Thus, it would be reasonable to assume that, should AtCDAs have any effect upon virus infection, their expression would be altered. To test that, AtCDA and AtCDAL4 mRNA expression were analyzed in virus-infected plants. These two genes were selected because it has been shown that AtCDA is the only completely functional CDA and is indispensable for pyrimidine metabolism, whereas AtCDAL4 retains a very low catalytic activity and is not involved in the pyrimidine pathway (Chen et al., 2016). Our results demonstrate that both genes were modulated by CabLCV infection. However, AtCDA was upregulated, whereas AtCDAL4 was downregulated. Although the expression of one gene could be regulated in response to the other through a negative feedback loop, the exclusive downregulation of AtCDAL4, but not AtCDA, by ACMV, rules out this possibility.

Chen et al. (2016) demonstrated that AtCDA transcripts were induced in senescent A. thaliana. It was suggested that induction of AtCDA expression in Hyaloperonospora parasitica-infected plants was due to early induction of senescence, rather than a defense response (Carviel et al., 2009). CabLCV induces severe symptoms in A. thaliana as early as 12 days after inoculation, leading to a premature induction of senescence (Ascencio-Ibanez et al., 2008). Thus, specific upregulation of AtCDA during infection could be related to the induction of senescence instead of being a direct effect of the viral infection. However, induction of AtCDA did not occur in TuMV-infected plants, which induces less severe symptoms but nevertheless also triggers early senescence in A. thaliana (Manacorda et al., 2013). In addition, AtCDA was not induced in ACMV-infected Sei-0 RICIb, a particularly susceptible A. thaliana ecotype which shows very severe symptoms. Thus, AtCDA seems to be especifically modulated by CabLCV and its induction is likely directed by the virus, rather than being an indirect effect of senescence. On the other hand, AtCDAL4 seems to be a general target modulated upon bipartite begomovirus infection, as this gene was induced by both the NW virus CabLCV and the OW virus ACMV.

Knockout of AtCDA led to delayed symptom development and reduced accumulation of CabLCV, indicating a positive role of this gene during CabLCV infection. To date, the only functional role ascribed to AtCDA is its involvement in the pyrimidine salvage pathway, converting cytidine to uridine (Chen et al., 2016; Zrenner et al., 2006). Thus, the loss of AtCDA function affects the cell's nucleotide poll (Chen et al., 2016), and CabLCV accumulation could

be indirectly affected. In addition, loss of AtCDA function leads to a high accumulation of intermediate metabolites such as cytidine and cytosine and to reduced plant growth, likely due to the cytotoxic effect of these compounds (Chen et al., 2016). This metabolic stress might negatively affect the virus through an overall perturbation in the plant metabolism or even through impairment of specific proviral factors. If this was true other viruses should also be affected. However, no differences in BCTV and TuMV accumulation were observed, ruling out this possibility.

PARPs [poly(ADP-ribosyl)ation] are members of a conserved protein family encoded by eukaryotes which is involved in the regulation of important cellular process through post-translational modifications (Perina et al., 2014). In animals and plants these proteins were demonstrated to be involved in DNA repair, programmed cell death, chromatin function and maintenance of genomic stability (Rissel and Peiter, 2019). In animals, it has been demonstrated that a high concentration of intracellular cytidine caused a reduction of basal PARP1 activity and led to decreased replication of the genome (Gemble et al., 2015; Gemble et al., 2017). Begomoviruses replicate using the host replication machinery, therefore any factor affecting the cellular replication machinery could also affect them (Hanley-Bowdoin et al., 2013). Although there are no studies showing that cytidine accumulation may also inhibit PARPs in plants, the accumulation of cytidine as observed in A. thaliana mutants could compromise the cell cycle and consequently affect CabLCV accumulation. To verify this possibility, we analyzed the ploidy level in healthy mutants and wt A. thaliana (Suppl. Table S2). Surprisingly, the DNA level in mutant plants was almost twice that of wt plants. Thus, inhibition of PARP activity does not seem to be the cause of reduced begomovirus accumulation in AtCDA knockout plants.

Altogether, several lines of evidence suggest a direct and specific effect of AtCDA on CabLCV accumulation, instead of an indirect effect due to general perturbation on the plant's metabolism. The fact that little information is available on cellular functions played by AtCDA makes it difficult to infer about their mode of action as a proviral factor. Further work will be necessary to reveal how AtCDA specifically affects CabLCV.

In animals, lethal mutagenesis is the main mechanism used by APOBEC/AID proteins to restrict a range of viruses with different genomic architectures (reviewed by (Willems and Gillet, 2015). It has been demonstrated that AtCDA has a mutagenic effect on the CaMV genome (Martin et al., 2016). Transient AtCDA overexpression during CaMV infection led to a significant enrichment in G to A transitions at three days post-agroinfiltration, whereas no differences were observed between the control and eight other AtCDALs tested (Martin et al.,

2016). However, transient overexpression of either AtCDA or AtCDAL4 did not affect the mutational spectrum of CabLCV and no differences in virus accumulation were observed, at least during the early stages of infection. Since AtCDA and AtCDAL4 expression was modulated at 14 days after virus inoculation, the absence of an effect in the early stages of infection could be due these proteins acting later in the infection. Thus, our experiment set up, where the latest time sampled was four days after infiltration, may not allow enough time for this protein to act. To test this possibility, transgenic A. thaliana plants stably overexpressing AtCDA and AtCDAL4 are being produced.

Martin et al. (2016) suggested that AtCDA may act in CaMV in a similar manner as APOBEC3 restricts HIV-1 infection. CaMV replicates by reverse transcription in the cytoplasm, using the positive-sense 35S RNA as a template. During replication, the complementary-strand DNA remains transitorily single-stranded, and thus could serve as a substrate for cytidine deamination mediated by AtCDA. In contrast, begomoviruses replicate by a combination of rolling-circle replication and recombination-dependent replication in the nucleus of infected cells (Hanley-Bowdoin et al., 2013; Jeske et al., 2001). Although begomoviruses spend most of their life cycle as ssDNA, which would be a suitable substrate for AtCDA and/or AtCDAL4 (considering an activity similar to APOBECs), it has been shown that AtCDA in healthy cells locates in the cytoplasm (Chen et al., 2016). Thus, AtCDA compartmentalization could prevent the activity of this enzyme on the begomovirus genome. The subcellular localization of AtCDAL4 has not been analyzed, but in silico analysis indicates that it is also a cytoplasmic protein. Still, several hosts factors have their subcellular localization altered during viral infection (Bruckner et al., 2017; Kushwaha et al., 2017). Thus, AtCDA and AtCDAL4 subcellular localization needs to be tested in begomovirus-infected plants, as they may be redirected to the nucleus. In addition, the low deaminase activity of AtCDAL4 in vitro (Chen et al., 2016) indicates that this protein may use a different, deamination-independent mechanism to restrict viral infection, as observed for APOBEC3G/F (Belanger et al., 2013). Finally, it is of course possible that these proteins may not have a mutagenic activity on CabLCV. Further investigation will be necessary to elucidate how they affect virus infection.

Although our results do not corroborate a role for AtCDA and AtCDAL4 in the deamination of the begomovirus genome, the possibility that other enzymes deaminate viral genome cannot be ruled out. Usually, genes encoding CDAs occur as a multigenic family in plants (Chen et al., 2016; Liu et al., 2019) and animals (Münk et al., 2012). Gene duplication events may support functional diversification and provide new features to the organism (Münk et al., 2012). The APOBEC3 subfamily is a example where gene duplication was followed by

functional diversification, with changes in activity and substrate specificity but always with a convergent antiviral role (Münk et al., 2012; Stavrou and Ross, 2015). A. thaliana encodes eighth putative CDAs considered to be pseudogenes (Chen et al., 2016). Our results indicate that AtCDAL4 has an antiviral role during CabLCV infection, suggesting that other genes may also be functional and perhaps act in deamination. Liu et al. (2019) demonstrated that different genes of the CDA family in rice are differentially modulated during biotic and abiotic stimuli, suggesting functional diversification. Thus, studies addressing a possible role of other CDAs in plants are necessary to verify whether they are involved in plant antiviral immunity.

If only spontaneous deamination is responsible for the mutational bias observed in begomoviruses (Duffy and Holmes, 2008; Duffy and Holmes, 2009; Ge et al., 2007), the same bias should be observed in members of other genera in the Geminiviridae. However, different biases are observed in begomoviruses and mastreviruses (van der Walt et al., 2008), suggesting the action of specific enzymes (other than CDAs).

Interestingly, a significant overrepresentation of C to U and G to A transitions was observed in an artificial non-coding sequence inserted on the TuMV genome, in both N. benthamiana (Lin et al., 2009) and A. thaliana (Lafforgue et al., 2011). These transitions are typical signatures of cytidine deaminases, and the authors suggest a possible involvement of CDAs in antiviral immune responses against RNA plant viruses. While it could be suggested that AtCDA would also be involved in antiviral defense against RNA viruses, as demonstrated for CaMV, Faivre-Nitschke et al. (1999) suggested that AtCDA is not involved in the editing of RNA as it does not show affinity for RNA. Our results demonstrate that neither transient AtCDA overexpression nor the loss of AtCDA function affect TuMV accumulation. Together, these results indicate that AtCDA is probably not involved in virus RNA edition, and that other enzymes would be involved in antiviral defense against RNA plant viruses. The subclass PLS of pentatricopeptide repeat (PPR) proteins, a widespread class of proteins in plants, carry a conserved DYW domain that contains a signature similar to the active domain of known CDAs (Ichinose and Sugita, 2016; Shikanai, 2015; Zhang et al., 2017). Interestingly, plants, different of other organisms, have over 1,000 genes encoding PPR proteins (Ichinose and Sugita, 2016; O'Toole et al., 2008). Although these proteins are recognized to act in organelle RNA editing, it would be tempting to speculate a possible involvement of PPRs in antiviral defense in plants, in a similar way as APOBECs, which were first identified as mRNA editing proteins (Salter et al., 2016).

In addition to the AtCDA mutagenic effect on the CaMV genome, it was demonstrated that partial silencing of AtCDA resulted in low CaMV accumulation (Martin et al., 2016).

However, no significant differences in mutational spectrum were observed between silenced and wt A. thaliana plants (Martin et al., 2016). We tested whether the complete knockout of AtCDA would have a stronger effect than that observed by Martin et al. (2016) in CaMV accumulation, as they obtained a 24% reduction in AtCDA mRNA expression. Although AtCDA was upregulated during CaMV infection, AtCDA knockout did not affect virus accumulation at 8 and 14 dpi. Martin et al. (2016), using an RNAi assay that targeted multiple CDA genes, simultaneously silenced AtCDA and AtCDAL1, 3, 5, 6, 8. According to the putative role of AtCDAL4 in CabLCV defense, it is possible that one or more of the other genes silenced together with AtCDA may have a antiviral effect against CaMV. However, this remains to be demonstrated.

Overall, we demonstrate here that two AtCDAs have evolved distinct functions upon CabLCV infection, independent of deamination activity. AtCDA seems to have evolved distinct functions for distinct viruses. In addition, it is demonstrated for the first time that AtCDLA4 may constitute a branch in antiviral immunity in plants. Further studies investigating the role of other proteins containing CDA domains may reveal a higher diversity of defense mechanisms against distinct viruses, as demonstrated for animal APOBEC/AID proteins.

Acknowledgements

References

Alfenas-Zerbini, P., Maia, I.G., Fávaro, R.D., Cascardo, J.C.M., Brommonschenkel, S.H., Zerbini, F.M., 2009. Genome-wide analysis of differentially expressed genes during the early stages of tomato infection by a potyvirus. Mol Plant-Mic Interac 22, 352-361.

Ascencio-Ibanez, J.T., Sozzani, R., Lee, T.J., Chu, T.M., Wolfinger, R.D., Cella, R., Hanley-Bowdoin, L., 2008. Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol 148, 436-454.

Belanger, K., Savoie, M., Rosales Gerpe, M.C., Couture, J.F., Langlois, M.A., 2013. Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses. Nucleic Acids Res 41, 7438-7452.

Borzooee, F., Joris, K.D., Grant, M.D., Larijani, M., 2019. APOBEC3G Regulation of the Evolutionary Race Between Adaptive Immunity and Viral Immune Escape Is Deeply Imprinted in the HIV Genome. Frontiers in Immunology 9.

Bruckner, F.P., Xavier, A.D.S., Cascardo, R.S., Otoni, W.C., Zerbini, F.M., Alfenas-Zerbini, P., 2017. Translationally controlled tumour protein (TCTP) from tomato and Nicotiana benthamiana is necessary for successful infection by a potyvirus. Mol Plant Pathol 18, 672-683.

Carviel, J.L., Al-Daoud, F., Neumann, M., Mohammad, A., Provart, N.J., Moeder, W., Yoshioka, K., Cameron, R.K., 2009. Forward and reverse genetics to identify genes involved in the age-related resistance response in Arabidopsis thaliana. Mol Plant Pathol 10, 621-634.

Chen, M., Herde, M., Witte, C.P., 2016. Of the nine cytidine deaminase-like genes in Arabidopsis, eight are pseudogenes and only one is required to maintain pyrimidine homeostasis in vivo. Plant Physiol 171, 799-809.

Chowdhury, S., Kitamura, K., Simadu, M., Koura, M., Muramatsu, M., 2013. Concerted action of activation-induced cytidine deaminase and uracil-DNA glycosylase reduces covalently closed circular DNA of duck hepatitis B virus. FEBS Lett 587, 3148-3152.

Conticello, S.G., Harris, R.S., Neuberger, M.S., 2003. The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. Curr Biol 13, 2009-2013.

Conticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K., Neuberger, M.S., 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol 22, 367-377.

Cotton, S., Grangeon, R., Thivierge, K., Mathieu, I., Ide, C., Wei, T.Y., Wang, A.M., Laliberte, J.F., 2009. Turnip mosaic virus RNA replication complex vesicles are mobile, align with microfilaments, and are each derived from a single viral genome. J Virol 83, 10460-10471.

Cuevas, J.M., Geller, R., Garijo, R., Lopez-Aldeguer, J., Sanjuan, R., 2015. Extremely high mutation rate of HIV-1 in vivo. PLoS Biol 13.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochemical Bulletin 19, 11-15.

Duffy, S., Holmes, E.C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus Tomato yellow leaf curl virus. J Virol 82, 957-965.

Duffy, S., Holmes, E.C., 2009. Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. J Gen Virol 90, 1539-1547.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., Pikaard, C.S., 2006. Gateway-compatible vectors for plant functional genomics and proteomics. Plant J 45, 616-629.

Faivre-Nitschke, S.E., Grienenberger, J.M., Gualberto, J.M., 1999. A prokaryotic-type cytidine deaminase from Arabidopsis thaliana: gene expression and functional characterization. European Journal of Biochemistry 263, 896-903.

Fan, J., Ma, G., Nosaka, K., Tanabe, J., Satou, Y., Koito, A., Wain-Hobson, S., Vartanian, J.P., Matsuoka, M., 2010. APOBEC3G generates nonsense mutations in human T-cell leukemia virus type 1 proviral genomes in vivo. J Virol 84, 7278-7287.

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.

Gemble, S., Ahuja, A., Buhagiar-Labarchede, G., Onclercq-Delic, R., Dairou, J., Biard, D.S., Lambert, S., Lopes, M., Amor-Gueret, M., 2015. Pyrimidine pool disequilibrium induced by a cytidine deaminase deficiency inhibits PARP-1 activity, leading to the under replication of DNA. PLoS Genet 11.

Gemble, S., Buhagiar-Labarchède, G., Onclercq-Delic, R., Jaulin, C., Amor-Guéret, M., 2017. Cytidine deaminase deficiency impairs sister chromatid disjunction by decreasing PARP-1 activity. Cell Cycle 16, 1128-1135.

Gowda, M., Venu, R.C., Li, H., Jantasuriyarat, C., Chen, S., Bellizzi, M., Pampanwar, V., Kim, H., Dean, R.A., Stahlberg, E., Wing, R., Soderlund, C., Wang, G.-L., 2007. Magnaporthe grisea infection triggers RNA variation and antisense transcript expression in rice. Plant Physiol 144, 524-533.

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.

Harkins, G.W., Delport, W., Duffy, S., Wood, N., Monjane, A.L., Owor, B.E., Donaldson, L., Saumtally, S., Triton, G., Briddon, R.W., Shepherd, D.N., Rybicki, E.P., Martin, D.P., Varsani, A., 2009. Experimental evidence indicating that mastreviruses probably did not co-diverge with their hosts. Virol J 6, 104.

Harris, R.S., Dudley, J.P., 2015. APOBECs and virus restriction. Virology 479–480, 131-145.

Henriet, S., Mercenne, G., Bernacchi, S., Paillart, J.C., Marquet, R., 2009. Tumultuous relationship between the human immunodeficiency virus type 1 viral infectivity factor (Vif) and the human APOBEC-3G and APOBEC-3F restriction factors. Microbiology and Molecular Biology Reviews 73, 211-232.

Holmes, R.K., Malim, M.H., Bishop, K.N., 2007. APOBEC-mediated viral restriction: not simply editing? Trends Biochem Sci 32, 118-128.

Holtz, C.M., Sadler, H.A., Mansky, L.M., 2013. APOBEC3G cytosine deamination hotspots are defined by both sequence context and single-stranded DNA secondary structure. Nucleic Acids Res 41, 6139-6148.

Ichinose, M., Sugita, M., 2016. RNA editing and its molecular mechanism in plant organelles. Genes 8, 5.

Jeske, H., Lutgemeier, M., Preiss, W., 2001. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. EMBO J 20, 6158-6167.

Kafer, C., Thornburg, R.W., 2000. Arabidopsis thaliana cytidine deaminase 1 shows more similarity to prokaryotic enzymes than to eukaryotic enzymes. J Plant Biol 43, 162-170.

Kim, E.Y., Lorenzo-Redondo, R., Little, S.J., Chung, Y.S., Phalora, P.K., Maljkovic Berry, I., Archer, J., Penugonda, S., Fischer, W., Richman, D.D., Bhattacharya, T., Malim, M.H., Wolinsky, S.M., 2014. Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection. PLoS Pathog 10.

Krishnan, A., Iyer, L.M., Holland, S.J., Boehm, T., Aravind, L., 2018. Diversification of AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in immunity. Proc Natl Acad Sci USA 115, E3201-E3210.

Kushwaha, N.K., Bhardwaj, M., Chakraborty, S., 2017. The replication initiator protein of a geminivirus interacts with host monoubiquitination machinery and stimulates transcription of the viral genome. PLoS Path 13, e1006587.

Lafforgue, G., Martinez, F., Sardanyes, J., de la Iglesia, F., Niu, Q.W., Lin, S.S., Sole, R.V., Chua, N.H., Daros, J.A., Elena, S.F., 2011. Tempo and mode of plant RNA virus escape from RNA interference-mediated resistance. J Virol 85, 9686-9695.

Lei, Y.-C., Hao, Y.-H., Zhang, Z.-M., Tian, Y.-J., Wang, B.-J., Yang, Y., Zhao, X.-P., Lu, M.-J., Gong, F.-L., Yang, D.-L., 2006. Inhibition of hepatitis B virus replication by APOBEC3G in vitro and in vivo. World Journal of Gastroenterology 12, 4492-4497.

Liddament, M.T., Brown, W.L., Schumacher, A.J., Harris, R.S., 2004. APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. Curr Biol 14, 1385-1391.

Lin, S.S., Wu, H.W., Elena, S.F., Chen, K.C., Niu, Q.W., Yeh, S.D., Chen, C.C., Chua, N.H., 2009. Molecular Evolution of a Viral Non-Coding Sequence under the Selective Pressure of amiRNA-Mediated Silencing. PLoS Path 5, -.

Liu, M.C., Liao, W.Y., Buckley, K.M., Yang, S.Y., Rast, J.P., Fugmann, S.D., 2018. AID/APOBEC-like cytidine deaminases are ancient innate immune mediators in invertebrates. Nat Commun 9, 018-04273.

Liu, W., Wang, P.F., Li, Z., Wang, Q.G., Wang, Y.Y., Yao, F.Y., Yang, L.Q., Pan, J.W., 2019. Genome-Wide Identification and Expression Analysis of the Cytidine Deaminase Subfamily in Rice. Russ J Plant Physl+ 66, 203-213.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 25, 402-408.

Manacorda, C.A., Mansilla, C., Debat, H.J., Zavallo, D., Sanchez, F., Ponz, F., Asurmendi, S., 2013. Salicylic acid determines differential senescence produced by two Turnip mosaic virus strains involving reactive oxygen species and early transcriptomic changes. Mol Plant-Mic Interac 26, 1486-1498.

Martin, S., Cuevas, J.M., Grande-Perez, A., Elena, S.F., 2016. A putative antiviral role of plant cytidine deaminases. F1000Res 6.

Moris, A., Murray, S., Cardinaud, S., 2014. AID and APOBECs span the gap between innate and adaptive immunity. Frontiers in Microbiology 5.

Münk, C., Willemsen, A., Bravo, I.G., 2012. An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. BMC Evol Biol 12, 71.

Ndunguru, J., De Leon, L., Doyle, C.D., Sseruwagi, P., Plata, G., Legg, J.P., Thompson, G., Tohme, J., Aveling, T., Ascencio-Ibanez, J.T., Hanley-Bowdoin, L., 2016. Two Novel DNAs That Enhance Symptoms and Overcome CMD2 Resistance to Cassava Mosaic Disease. J Virol 90, 4160-4173.

O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., Small, I., 2008. On the expansion of the pentatricopeptide repeat gene family in plants. Mol Biol Evol 25, 1120-1128.

Perina, D., Mikoc, A., Ahel, J., Cetkovic, H., Zaja, R., Ahel, I., 2014. Distribution of protein poly(ADP-ribosyl)ation systems across all domains of life. DNA Repair 23, 4-16.

R Development Core Team, 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rajabu, C.A., Kennedy, G.G., Ndunguru, J., Ateka, E.M., Tairo, F., Hanley-Bowdoin, L., Ascencio-Ibanez, J.T., 2018. Lanai: A small, fast growing tomato variety is an excellent model system for studying geminiviruses. J Virol Methods 256, 89-99.

Ramakers, C., Ruijter, J.M., Deprez, R.H., Moorman, A.F., 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339, 62-66.

Reyes, M.I., Nash, T.E., Dallas, M.M., Ascencio-Ibáñez, J.T., Hanley-Bowdoin, L., 2013. Peptide Aptamers That Bind to Geminivirus Replication Proteins Confer a Resistance Phenotype to Tomato Yellow Leaf Curl Virus and Tomato Mottle Virus Infection in Tomato. J Virol 87, 9691-9706.

Rissel, D., Peiter, E., 2019. Poly(ADP-Ribose) polymerases in plants and their human counterparts: parallels and peculiarities. International Journal of Molecular Sciences 20, 1638.

Sadler, H.A., Stenglein, M.D., Harris, R.S., Mansky, L.M., 2010. APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. J Virol 84, 7396-7404.

Salter, J.D., Bennett, R.P., Smith, H.C., 2016. The APOBEC protein family: united by structure, divergent in function. Trends Biochem Sci 41, 578-594.

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.

Scholelz, J.E., Shepherd, R.J., 1988. Host range control of cauliflower mosaic virus. Virology 162, 30-37.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., Malim, M.H., 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418, 646-650.

Shen, W., Dallas, M.B., Goshe, M.B., Hanley-Bowdoin, L., 2014. SnRK1 phosphorylation of AL2 delays Cabbage leaf curl virus infection in Arabidopsis. J Virol 88, 10598-10612.

Shikanai, T., 2015. RNA editing in plants: machinery and flexibility of site recognition. Biochimica et biophysica acta 1847, 779-785.

Smyth, R.P., Davenport, M.P., Mak, J., 2012. The origin of genetic diversity in HIV-1. Virus Res 169, 415-429.

Stavrou, S., Ross, S.R., 2015. APOBEC3 proteins in viral immunity. Journal of Immunology 195, 4565-4570.

Suspene, R., Aynaud, M.M., Koch, S., Pasdeloup, D., Labetoulle, M., Gaertner, B., Vartanian, J.P., Meyerhans, A., Wain-Hobson, S., 2011. Genetic editing of herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and in vivo. J Virol 85, 7594-7602.

van der Walt, E., Martin, D.P., Varsani, A., Polston, J.E., Rybicki, E.P., 2008. Experimental observations of rapid Maize streak virus evolution reveal a strand-specific nucleotide substitution bias. Virol J 5, 104.

Weil, A.F., Ghosh, D., Zhou, Y., Seiple, L., McMahon, M.A., Spivak, A.M., Siliciano, R.F., Stivers, J.T., 2013. Uracil DNA glycosylase initiates degradation of HIV-1 cDNA containing misincorporated dUTP and prevents viral integration. Proceedings of the National Academy of Sciences 110, E448-E457.

Willems, L., Gillet, N.A., 2015. APOBEC3 interference during replication of viral genomes. Viruses 7, 2999-3018.

Yang, Y., Liu, T., Shen, D., Wang, J., Ling, X., Hu, Z., Chen, T., Hu, J., Huang, J., Yu, W., Dou, D., Wang, M.B., Zhang, B., 2019. Tomato yellow leaf curl virus intergenic siRNAs target a host long noncoding RNA to modulate disease symptoms. PLoS Path 15.

Zerbini, F.M., Briddon, R.W., Idris, A., Martin, D.P., Moriones, E., Navas-Castillo, J., Rivera-Bustamante, R., Varsani, A., ICTV Consortium, 2017. ICTV Virus Taxonomy Profile: Geminiviridae. J Gen Virol 98, 131-133.

Zhang, B., Liu, G., Li, X., Guo, L., Zhang, X., Qi, T., Wang, H., Tang, H., Qiao, X., Zhang, J., Xing, C., Wu, J., 2017. A genome-wide identification and analysis of the DYW-deaminase genes in the pentatricopeptide repeat gene family in cotton (Gossypium spp.). PLoS ONE 12, e0174201.

Zhang, H., Yang, B., Pomerantz, R.J., Zhang, C., Arunachalam, S.C., Gao, L., 2003. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature 424, 94-98.

Zrenner, R., Stitt, M., Sonnewald, U., Boldt, R., 2006. Pyrimidine and purine biosynthesis and degradation in plants. Annu Rev Plant Biol 57, 805-836.

Figure legends

Figure 1. Expression of AtCDA and AtCDAL4 mRNAs are differentially modulated during Cabbage leaf curl virus (CabLCV) infection. Relative expression levels of AtCDA (**A**) and AtCDAL4 (**B**) mRNA were analysed in A. thaliana Col-0 infected by Beet curly top virus (BCTV), CabLCV and Turnip mosaic virus (TuMV) using quantitative PCR (qPCR) at 16 days post-inoculation (dpi). (**C**) Relative expression levels of the non-functional AtCDAL3 gene were also analysed in CabLCV infected plants as a control. Relative expression levels of AtCDA (**D**) and AtCDAL4 (**E**) mRNA were analysed at 7, 14, 21 and 28 dpi in CabLCV infected-plants. Mock inoculated plants were inoculated with empty vectors. Data are presented as means and error bars represent standard deviation of three to four biological replications, each comprised of pools of three plants. Asterisks indicate a significant difference between virus-infected plants compared to mock-inoculated plants using Student's t-test (*, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$). (**G**) Absolute quantification of CabLCV DNA-A was performed at 7, 14, 21 and 28 dpi in the same samples used in (**D**) and (**E**) using 10 ng of total DNA as template for qPCR. Means are presented as dots and error bars represent standard deviations. Different letters indicate significant differences by Tukey's test ($p \leq 0.05$).

Figure 2. Expression of AtCDA and AtCDAL4 mRNAs are not modulated during Tomato yellow leaf curl virus (TYLCV) infection. (**A**) Representative images of A. thaliana Col-0 mock-inoculated, Cabbage leaf curl virus (CabLCV)- and TYLCV-infected plants at 21 days post-inoculation (dpi). (**B**) PCR-based detection of TYLCV at 21 dpi in assymptomatic plants used in the gene expression analysis. PCR reactions were performed using a set of divergent primers that amplify only replicated viral DNA. C+, PCR positive control; C-, PCR negative control; Mock, PCR perfomed using DNA from mock-inoculated plants. (**C**) Absolute quantification of viral DNA performed at 21 dpi using 10 ng of total DNA extracted from systemically infected leaves with specific primers for each virus. Relative expression levels of AtCDA (**D**) and AtCDAL4 (**E**) mRNA were analysed at 21dpi in TYLCV- and CabLCV-infected plants. Data are presented as means and error bars represent standard deviation of four biological replications, each comprise of pools of three plants. Asterisks indicate a significant difference between treatments by Student's t-test (*, $p \leq 0.01$).

Figure 3. AtCDAL4 but not AtCDA seems to be a conserved target modulated during bipartite begomovirus infection. (**A**) Symptoms induced by African cassava mosaic virus (ACMV) in

wild type A. thaliana Sei-0 (wt Sei-0) and in A. thaliana Sei-0 RICIb (Sei-0 RICIb), a transgenic line harboring a sequence enhancing geminivirus symptoms (SEGS-I). Images were obtained at 21 days post-inoculation (dpi). (**B**) Absolute quantification of viral DNA performed at 21 dpi using specific DNA-A primers for each virus. Data are presented as means and error bars represent standard deviation. Different letters indicate significant differences by Tukey's test ($p \leq 0.05$). Relative expression levels of AtCDA (**C, D** and **E**) and AtCDAL4 (**F, G** and **H**) mRNA were analysed at 21 dpi in wt Sei-0 (**C**, **D**, **F** and **G**) during ACMV and Cabbage leaf curl virus (CabLCV) infection and in Sei-0 RICIb (**E** and **H**) infected by ACMV. Data are presented as means and error bars represent standard deviation of three to four biological replications, each comprise of pools of three plants. Asterisks indicate a significant difference between virus-infected plants compared to mock-inoculated plants using Student's t-test (*, $p \leq 0.05$; **, $p \leq 0.01$).

Figure 4. Transient overexpression of AtCDA and AtCDAL4 does not affect Cabbage leaf curl virus (CabLCV) accumulation during the early stages of infection and does not have any mutagenic effect on the viral genome. (**A**) Absolute quantification of viral DNA performed at 48, 72 and 96 hours post-agroinfiltration (hpa). Expression cassettes harboring AtCDA, AtCDAL4 and GUS were co-infiltrated with CabLCV DNA-A and DNA-B infectious clones in N. benthamiana leaves. Total DNA was extracted from agroinfiltrated spots at each time and 10 ng of total DNA were used for qPCR. Data are presented as means and standard deviation from all samples of three independent experiments. (**B**) Relative quantification of Turnip mosaic virus (TuMV) at 96 hpa co-infiltrated with expression cassettes harboring AtCDA, AtCDAL4 and GUS. Means and standard deviations from six biological replications are shown. (**C**) Total RNA was extracted from representative samples for each treatment and reverse trasncribed. cDNAs were used as templates for PCR amplification with specific primers for each construct to confirm gene expression. Non-RT control corresponds to amplification from non-reverse transcribed RNA. H2B amplification was used as an internal control. (**D**) Mutational spectra from CabLCV sequences isolated from N. benthamiana tissues transiently overexpressing AtCDA, AtCDAL4 and GUS. For each treatment approximately 15,000 nucleotides were sequenced from two biological replications from two independent experiment each.

Figure 5. AtCDA knockout delays symptom development and negatively affects Cabbage leaf curl virus (CabLCV) accumulation. (**A**) Schematic representation showing T-DNA insertions

(triangles) and primer annealing sites (arrowheads) used for genotyping cda1 and cda2 knockout lines. The black continuos line represents the translated region and dashed lines indicate non-translated regions. (**B**) Schematic representation showing the construct inserted in the cda1 line used to generate the complementation line. The yellow line corresponds to the yellow fluorescent protein (YFP) coding region fused to the AtCDA coding region. (**C**) PCR-based genotyping of the cda1, cda2 and cda1-CDA::YFP lines. T-DNA detection and homozygosity confirmation was performed for a small population of progeny plants after one self-fertilization cycle. AtCDA knockout plants show delayed symptom development during CabLCV infection. The mean symptom scores from a representative experiment containing 21 plants from each plant genotype (**E**) and the average time required to reach a given symptom score (**F**) are shown. Asterisks indicate significant differences between treatments by Wilcoxon rank sum test (*, $p \leq 0.05$; **, $p \leq 0.001$). (**D**) Absolute quantification of viral DNA performed at 12 days post-inoculation using specific primers for CabLCV DNA-A. Data are presented as means and error bars represent standard deviation from 10 biological replications, each comprised of pools of three plants from two independent experiments. Different letters indicate significant differences by Tukey's test ($p \leq 0.05$).

Figure 6. Symptom severity and accumulation of Cabbage leaf curl virus (CabLCV) increase in a cdal4 knockout line. (**A**) Schematic representation showing T-DNA insertion (triangle) and primer annealing sites (arrowheads) used for genotyping the cdal4 knockout line. The black continuos line represents the translated region. (**B**) PCR-based genotyping of the cdal4 line. T-DNA detection and homozygosity confirmation was performed for a small population of progeny plants after one self-fertilization cycle. (**C**) Symptoms induced by CabLCV in A. thaliana wild type (Col-0) and in the cdal4 knockout line at 14 days post-inoculation (dpi). (**D**) The mean symptom scores were obtained at 14 dpi. The asterisk indicates a significant difference between treatments by Wilcoxon rank sum test ($p \leq 0.05$). (**E**) Absolute quantification of viral DNA performed at 14 dpi using specific primers for CabLCV DNA-A and 10 ng of total DNA. Data are presented as means and error bars represent standard deviation from 11 and 15 biological replications for the Col-0 and cdal4 lines, respectively. The asterisk indicates a significant difference between treatments by Student's t-test ($p \leq 0.05$).

Figure 7. Cauliflower mosaic virus (CaMV) infection induces AtCDA expression but its knockout does not affect CaMV accumulation. Relative expression levels of AtCDA (**A**) and AtCDAL4 (**B**) mRNAs were analysed in A. thaliana Col-0 infected by CaMV using quantitative

PCR (qPCR) at 21 days post-inoculation (dpi). Data are presented as means and error bars represent standard deviation from five biological replicates, each comprise of pools of three plants. The asterisk indicates a significant difference between virus-infected plants compared to mock-inoculated plants using Student's t-test (*, $p \leq 0.001$). (**C**) Symptoms induced by CaMV in *A. thaliana* wild type (Col-0), cda1-CDA::YFP (a complementation line), cda1 and cda2 knockout lines at 14 dpi. Absolute quantification of viral DNA was performed at 8 dpi (**D**) and 14 dpi (**E**) using 10 ng of total DNA. Data are presented as means and error bars represent standard deviation from five (**D**) and 10 (**E**) biological replications ,each comprised of pools of three plants. Data in (**E**) refer to all biological replications from two independent experiments. Different letters indicate significant differences by Tukey's test ($p \leq 0.05$).

Figure 1

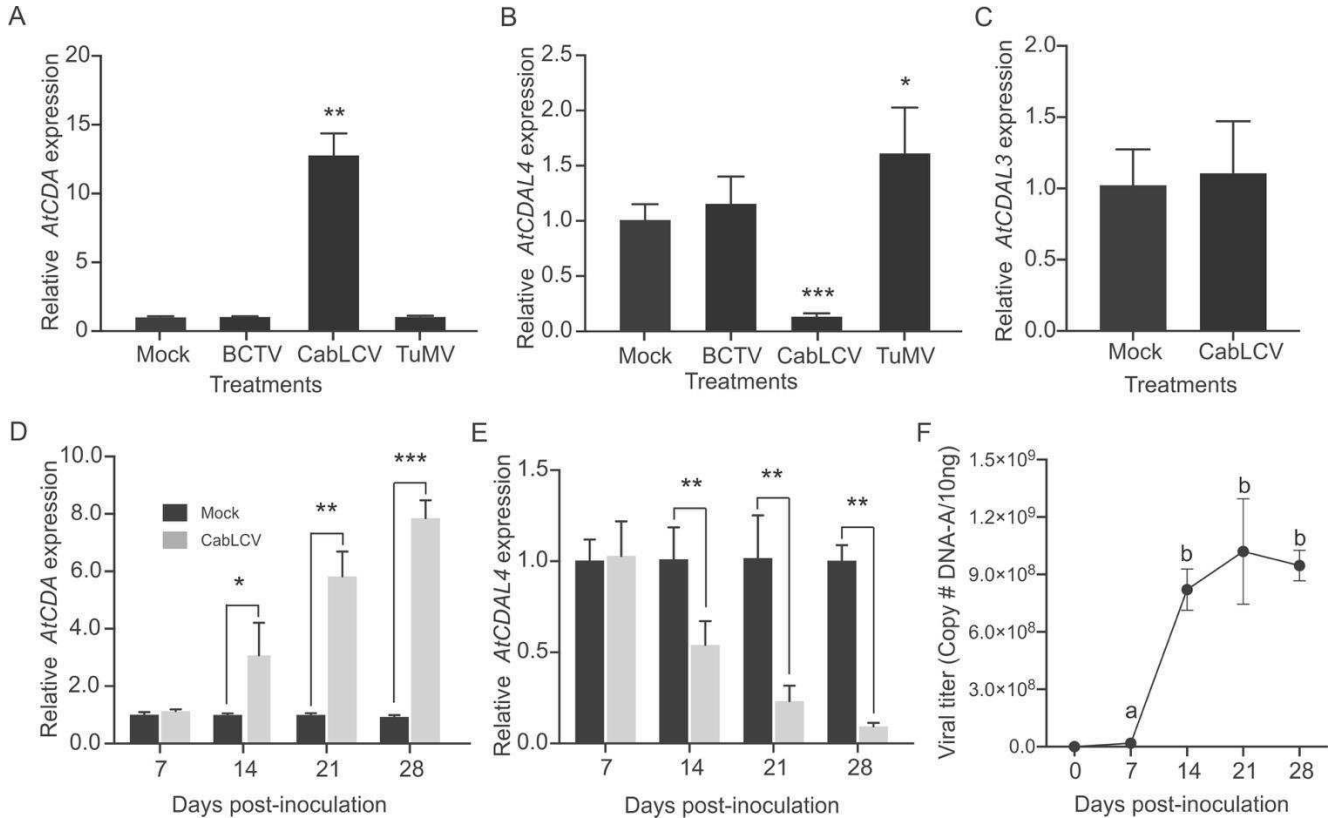

A
Relative AtCDA expression
20
15
10
5
0
**
Mock
BCTV
CabLCV
TuMV
Treatments
B
Relative AtCDAL4 expression
2.5
2.0
1.5
1.0
0.5
0.0
*
***
Mock
BCTV
CabLCV
TuMV
Treatments
C
Relative AtCDAL3 expression
2.0
1.5
1.0
0.5
0.0
Mock
CabLCV
Treatments
D
Relative AtCDA expression
10.0
8.0
6.0
4.0
2.0
0.0
Mock
CabLCV
*
**
***
7
14
21
28
Days post-inoculation
E
Relative AtCDAL4 expression
1.5
1.0
0.5
0.0
**
**
**
7
14
21
28
Days post-inoculation
F
Viral titer (Copy # DNA-A/10ng)
1.5×10^9
1.2×10^9
9.0×10^8
6.0×10^8
3.0×10^8
0.0
a
b
b
b
0
7
14
21
28
Days post-inoculation

Figure 2

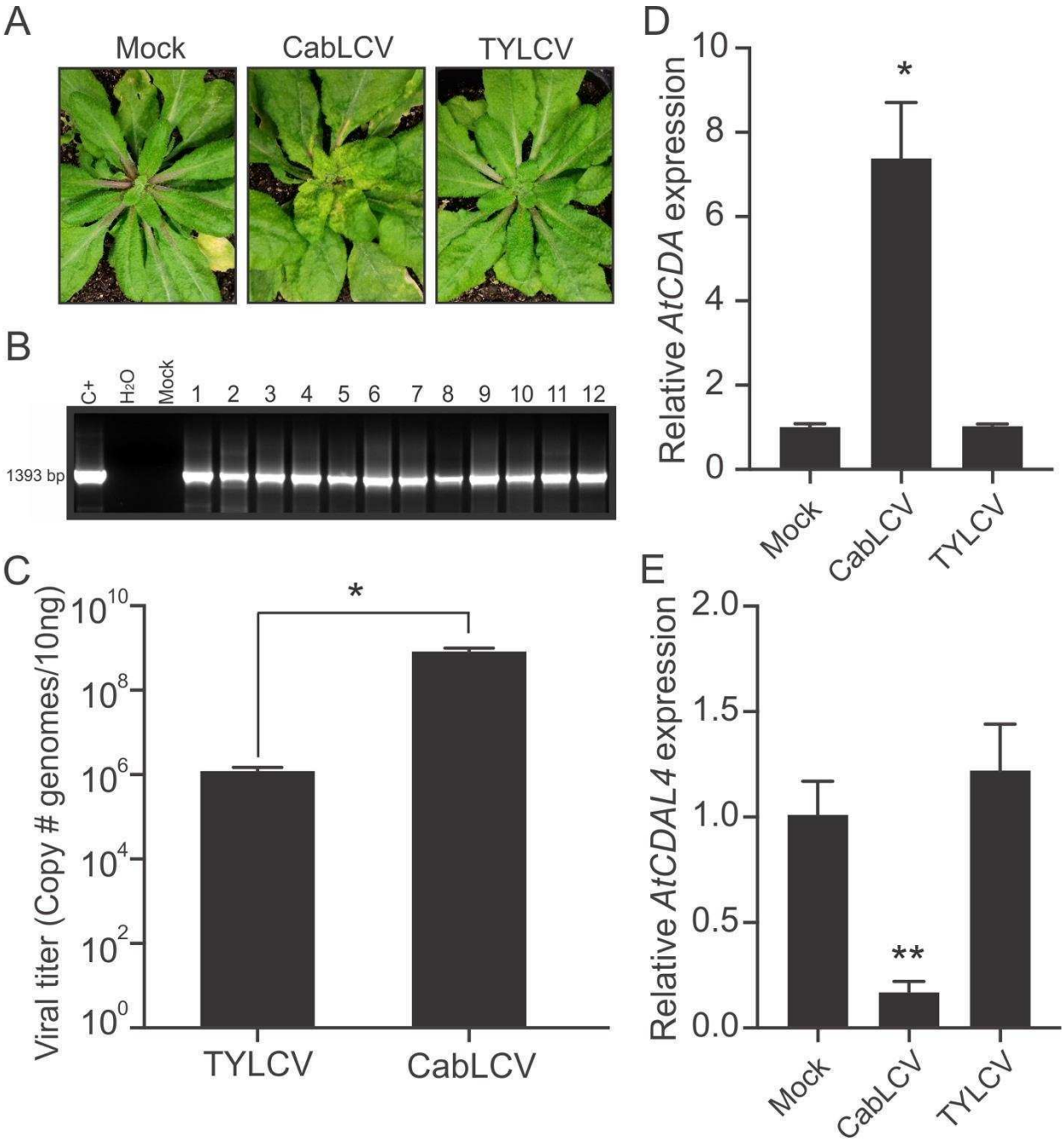

A
Mock
CabLCV
TYLCV
B
C+
H2O
Mock
1 2 3 4 5 6 7 8 9 10 11 12
1393 bp
C
Viral titer (Copy # genomes/10ng)
10^10
10^8
10^6
10^4
10^2
10^0
*
TYLCV
CabLCV
D
Relative AtCDA expression
10
8
6
4
2
0
*
Mock
CabLCV
TYLCV
E
Relative AtCDAL4 expression
2.0
1.5
1.0
0.5
0.0
**
Mock
CabLCV
TYLCV

Figure 3

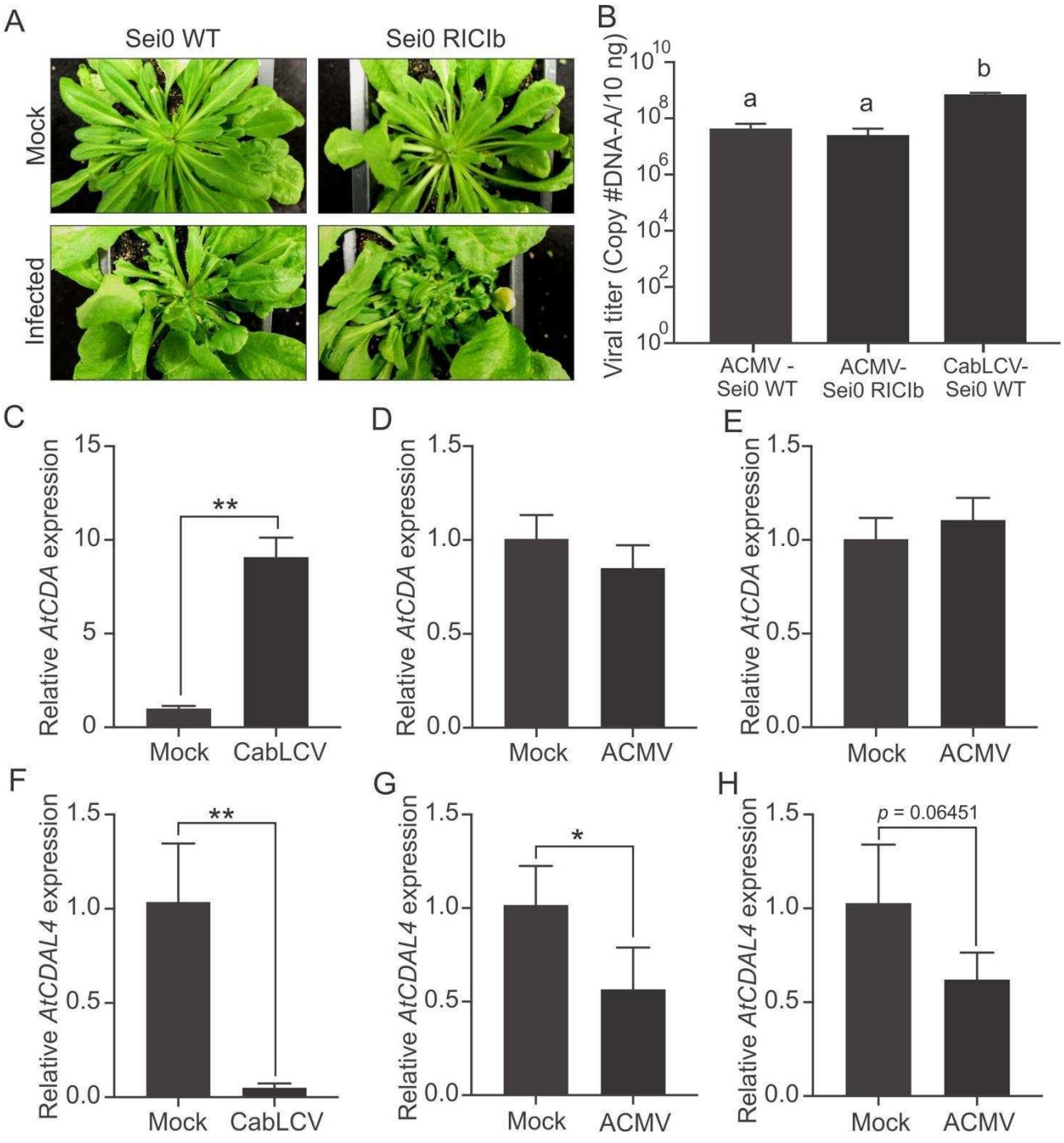

A
Sei0 WT
Sei0 RICIb
Mock
Infected
B
Viral titer (Copy #DNA-A/10 ng)
a
a
b
ACMV - Sei0 WT
ACMV- Sei0 RICIb
CabLCV- Sei0 WT
C
Relative AtCDA expression
**
Mock
CabLCV
D
Relative AtCDA expression
Mock
ACMV
E
Relative AtCDA expression
Mock
ACMV
F
Relative AtCDAL4 expression
**
Mock
CabLCV
G
Relative AtCDAL4 expression
*
Mock
ACMV
H
Relative AtCDAL4 expression
p = 0.06451
Mock
ACMV

Figure 4

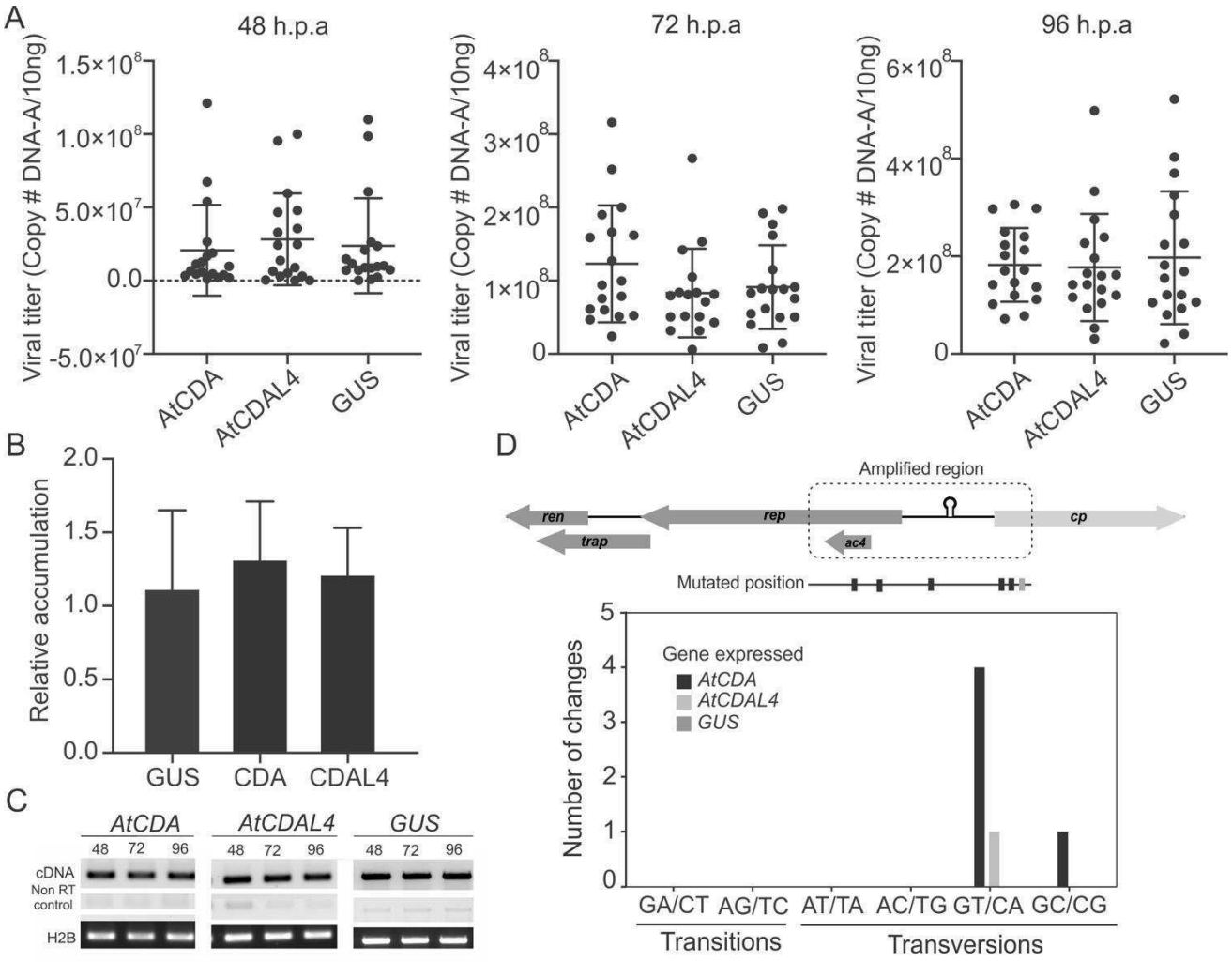

A
48 h.p.a
72 h.p.a
96 h.p.a
Viral titer (Copy # DNA-A/10ng)
AtCDA
AtCDAL4
GUS
B
Relative accumulation
GUS
CDA
CDAL4
C
AtCDA
AtCDAL4
GUS
48 72 96
cDNA
Non RT control
H2B
D
Amplified region
ren
rep
cp
trap
ac4
Mutated position
Gene expressed
AtCDA
AtCDAL4
GUS
Number of changes
GA/CT AG/TC AT/TA AC/TG GT/CA GC/CG
Transitions
Transversions

Figure 5

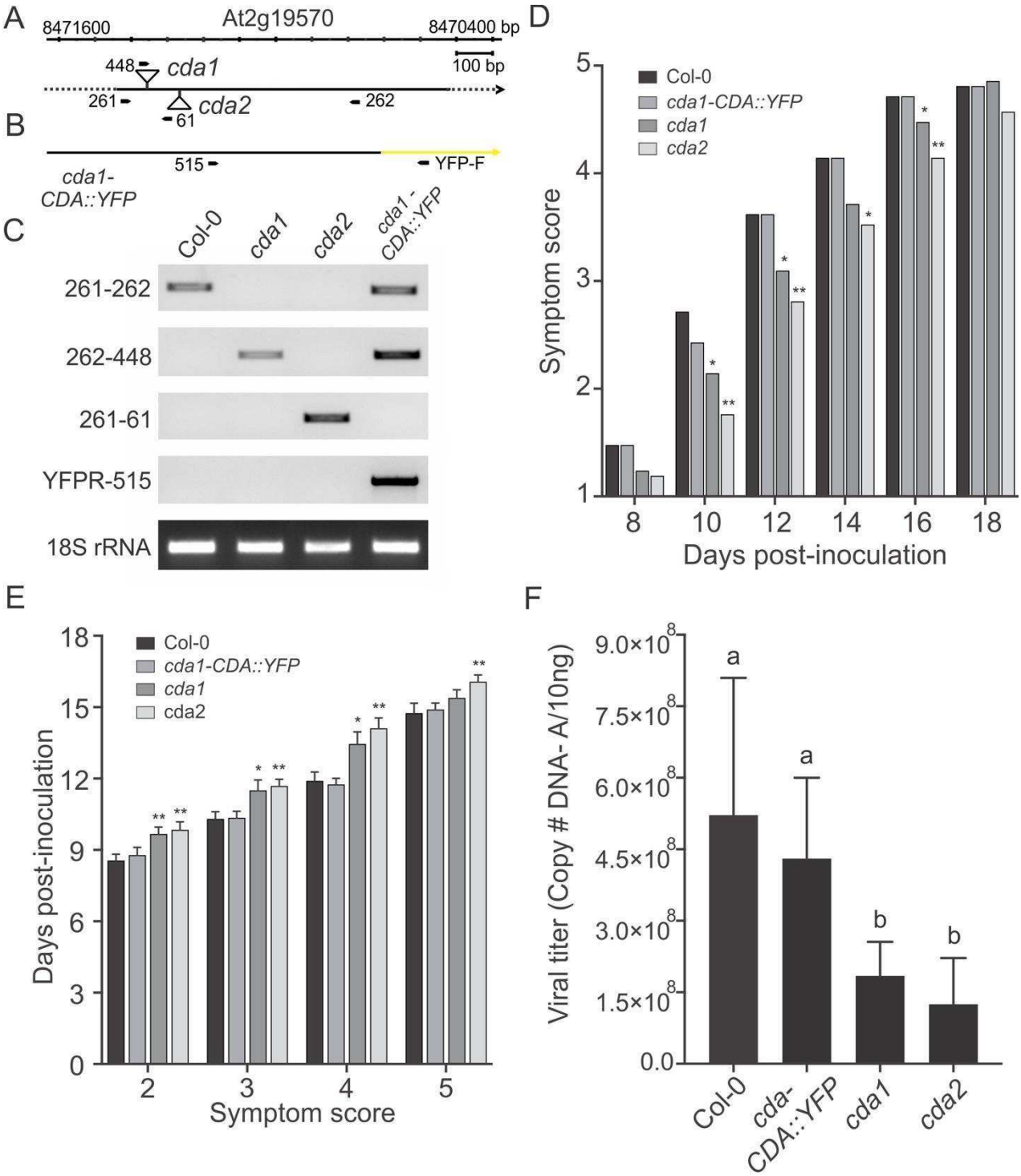

A
8471600
At2g19570
8470400 bp
448
cda1
100 bp
261
cda2
262
61
B
515
YFP-F
cda1-CDA::YFP
C
Col-0
cda1
cda2
cda1-CDA::YFP
261-262
262-448
261-61
YFPR-515
18S rRNA
D
Col-0
cda1-CDA::YFP
cda1
cda2
Symptom score
Days post-inoculation
E
Col-0
cda1-CDA::YFP
cda1
cda2
Days post-inoculation
Symptom score
F
Viral titer (Copy # DNA- A/10ng)
a
a
b
b
Col-0
cda-CDA::YFP
cda1
cda2

Figure 6

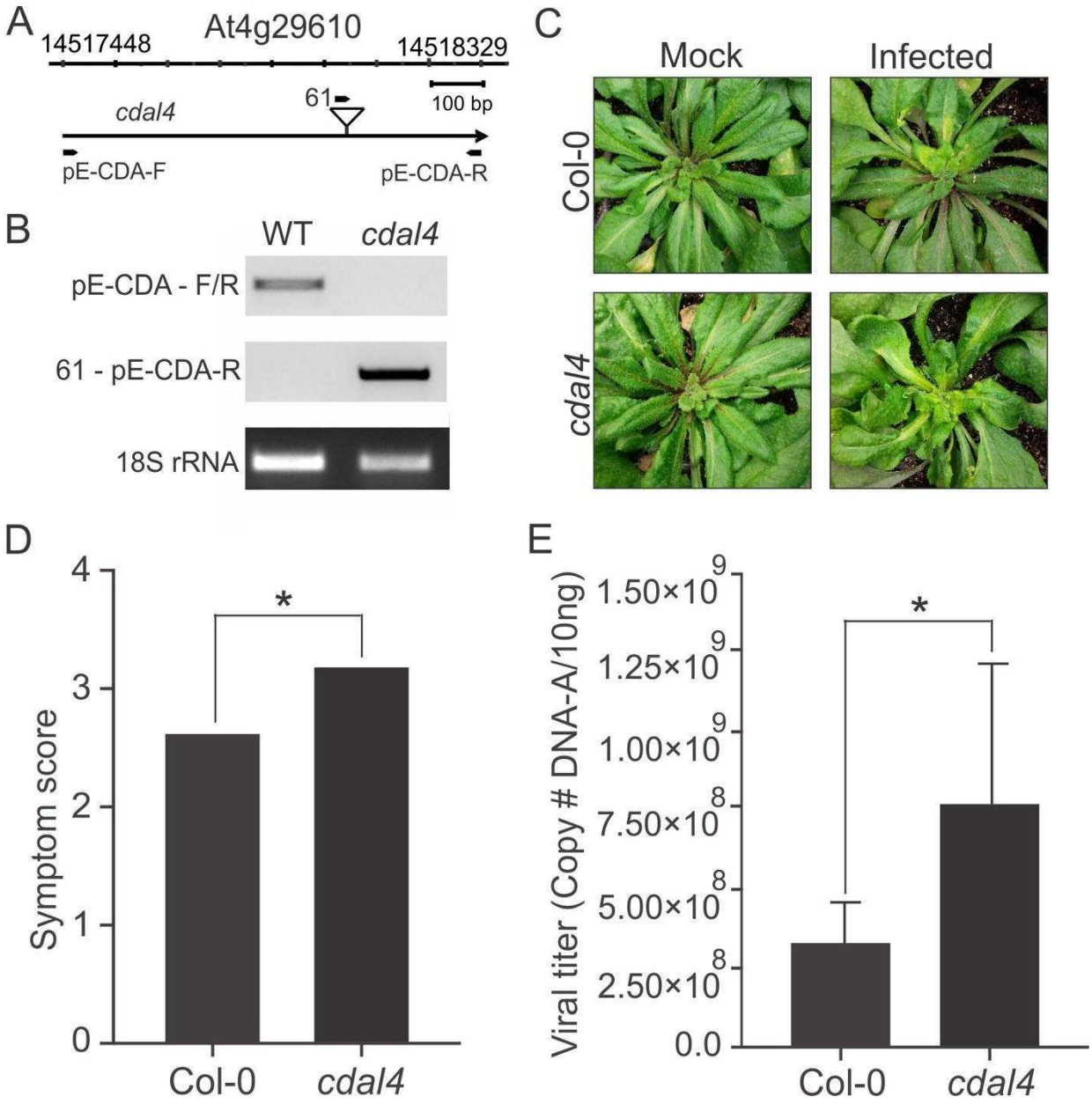
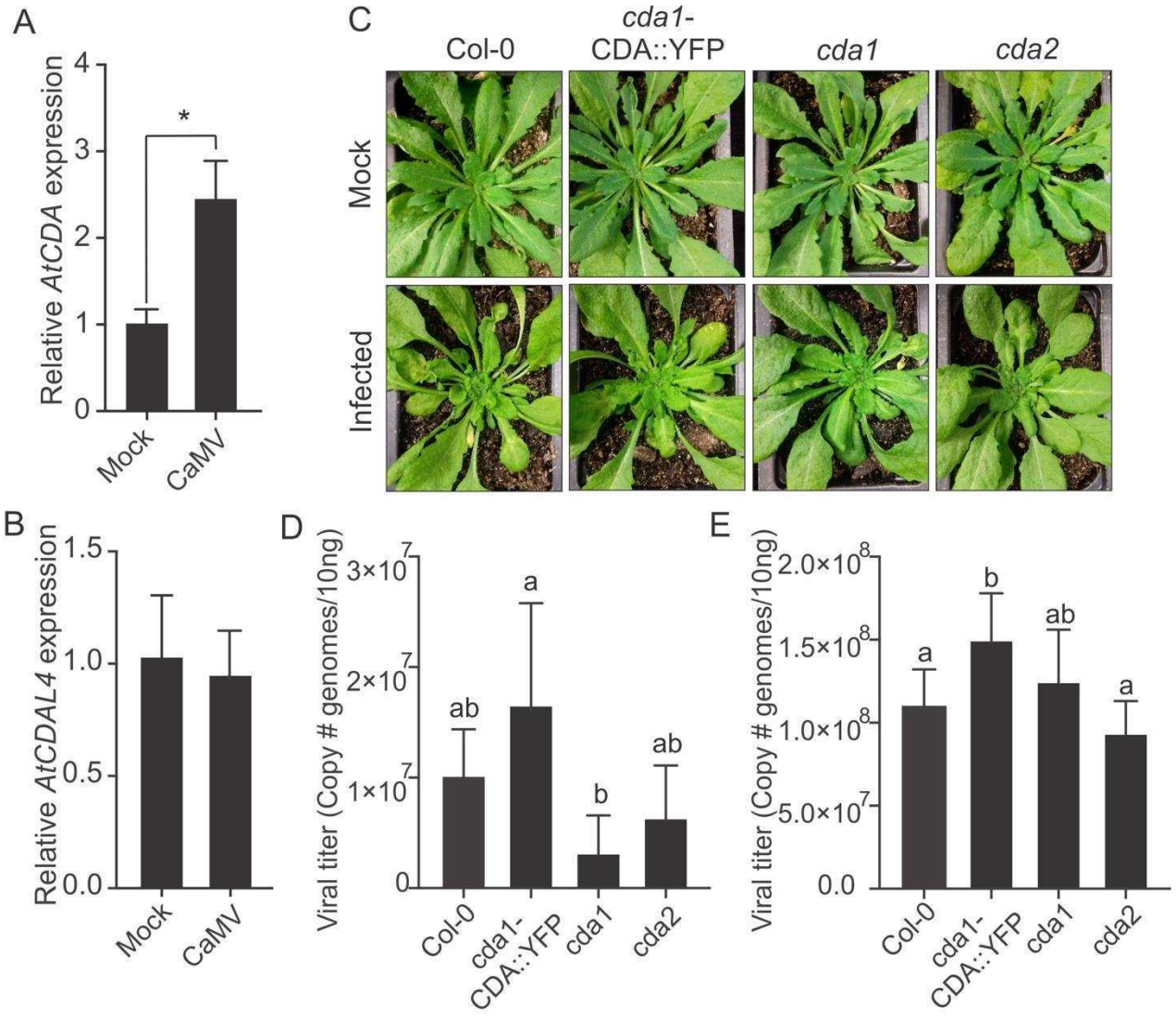

A
14517448
At4g29610
14518329
61
100 bp
cdal4
pE-CDA-F
pE-CDA-R
B
WT
cdal4
pE-CDA - F/R
61 - pE-CDA-R
18S rRNA
C
Mock
Infected
Col-0
cdal4
D
Symptom score
0
1
2
3
4
*
Col-0
cdal4
E
Viral titer (Copy # DNA-A/10ng)
0.0
2.50×10^8
5.00×10^8
7.50×10^8
1.00×10^9
1.25×10^9
1.50×10^9
*
Col-0
cdal4

Figure 7


A
Relative AtCDA expression
4
3
2
1
0
*
Mock
CaMV
C
Col-0
cda1-
CDA::YFP
cda1
cda2
Mock
Infected
B
Relative AtCDAL4 expression
1.5
1.0
0.5
0.0
Mock
CaMV
D
Viral titer (Copy # genomes/10ng)
3×10^7
2×10^7
1×10^7
0
ab
a
b
ab
Col-0
cda1-
CDA::YFP
cda1
cda2
E
Viral titer (Copy # genomes/10ng)
2.0×10^8
1.5×10^8
1.0×10^8
5.0×10^7
0.0
a
b
ab
a
Col-0
cda1-
CDA::YFP
cda1
cda2

Supplementary Table S1. List of primers used in this study.

Primer	Sequence (5'-3')	Purpose	Reference
qN515	CAAGAAATTCGCGACGCACCTG	AtCDA quantification - qPCR	Chen et al., 2016
qN516	TCGGAATCGGCGGAGTTGTTTG		
qAtCDAL4-1F	ACCTTCCAAAGCTGATCGAC	AtCDAL4 quantification - qPCR	This study
qAtCDAL4-1R	CAACTGCTCCGACTTTGAATG		
qCDAL3-1F	GTCAGTTAGCAGTCGCCAT	AtCDAL3 quantification- qPCR	This study
qCDAL3-1R	TCTGCTTCATGCTTTGATTTGAG		
qCabLCV-A-DpnI-1F	GTTCAACATGTTCGACAACGAG	CabLCV quantification - qPCR	This study
qCabLCV-A-DpnI-1R	CATACTGCCCACCGGTTAC		
qTuMV-F	CCGAACATAAACGGAATGTGG GTGA	TuMV quantification - qPCR	Bruckner et al., 2016
qTuMV-R	TGGCGTGGTCAATGAGCGGTT		
H2B-qPCR-F	GAGAAGCTGGCTCAGGAATC	Reference gene for N. benthamiana	Bruckner et al., 2016
H2B-qPCR-R	CACAAGCCTCACAGCAGTCT		
GUS-F	GTACCTCGCATTACCCTTACG	To verify GUS expression	This study
GUS-R	CTTCTCTGCCGTTTCCAAATC		
YFP-F	ACAAGTTCAGCGTGTCCG	To confirm complementation lines	This study
YFP-R	TCCTGGACGTAGCCTTCG		
CDAL4-pENTR-F	CACCATGAAGTTCGTTTACACTCCCAGC	Complete AtCDAL4 CDS amplification	This study
CDAL4-pENTR-R	TTAATTTTTATTGGTCTTATAGC		
CDAGW-Fwd	AAAAAGCAGGCTTCACAATGGATAAGCCAAGCTTCG	Complete AtCDA CDS amplification	This study
CDAGW-Rev_1	AGAAAGCTGGGTCCTAAGCTTCATAGCAATGAAACAC		
qCaMV-1F	GAGATACTTTACTGCGTCTCGAC	CaMV quantification - qPCR	This study
qCaMV-1R	TGGACATGGTTCTTCTGACTTC		
qPCR-SlCDA-1F	GTGGGTATGAGCGAATAGTGG	SlCDA quantification - qPCR	This study
qPCR-SlCDA-1R	CAGAAAGATCCTAGCAGTGTCC		
UpBCTV-qRT (Fwd)	TGACGTCGGAGCTGGATTTAG	BCTV quantification - qPCR	Luna et al., 2017
LowBCTV-qRT (Rev)	CTACACGAAGATGGGCAACCT		
Actin-F	TGGTCGGAATGGGACAGAAG	Reference gene for tomato - TYLCV and ToMoV	Yang et al., 2019
Actin-R	CTCAGTCAGGAGAACAGGGT		

N261	TCGTAATCCAATCCAAAGAAGC	Knockout line confirmation	
N262	TAAAGCCGTTTGTTTCAAATCG	Knockout line confirmation	Chen et al., 2016
N61	TGGTTCACGTAGTGGGCCATC	Knockout line confirmation	
448	ATATTGACCATCATACTCATTGC	Knockout line confirmation	
qRT_18SRNA-Fwd	TTTGCGCGCCTGCTGCC	RT-PCR control	Calil et al., 2018
qRT_18SRNA-Rvs	TGTGCTGGCGACGCATCATT		
SAND-F	GTTGGGTCACACCAGATTTTG	Reference gene for A. thaliana	Lilly et al., 2011
SNAD-R	GCTCCTTGCAAGAACACTTCA		
TYLCV div For	AAATAATGCGGGTCTAGCTCA	TYLCV detection - divergent primer pair	This study
TYLCV div Rev	TGTATGGGCTGTCGAAGTTC		
P3P-AA2F	TCTGCAATCCAGGACCTACC	ACMV quantification - qPCR	This study
P3P-AA2R+4R	GGCTCGCTTCTTGAATTGTC		
CabLCV A dirPCR Fwd	CTCTAGGAACATCTGGGCTTCTA	CabLCV detection - divergent primer pair	This study
CabLCV A dirPCR Rev	CCTTATAATTGCGAGACGCTCT		

Supplementary Table 2. DNA ploidy levels in healthy plants of wt A. thaliana (Col-0), mutants (cda1 and cda2) and the complementation line cda1::CDA-YFP.

Genotype	Ploidy level*	Mean (SD)#
Col-0	2.28	
	2.01	2.08 (0.178) a
	1.95	
cda1::CDA-YFP	2.43	
	2.83	2.81 (0.372) b
	3.17	
cda1	5.52	
	5.74	5.60 (0.120) c
	5.55	
cda2	4.95	
	5.39	5.18 (0.217) c
	5.21	

* Each row represents a biological replication, each one composed by pools of three plants.

Means followed by different letters indicate significant differences by Tukey's test ($p \leq 0.05$).

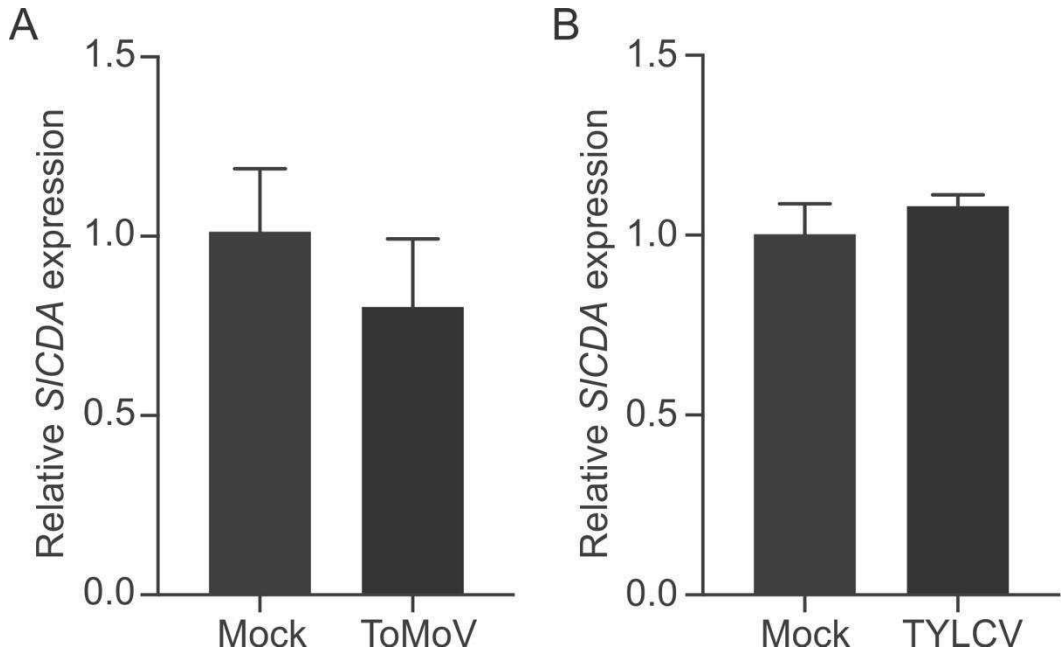
Supplementary Figure legends

Supplementary Figure S1. Expression levels of tomato CYTIDINE DEAMINASE (SlCDA), a homolog of AtCDA, does not change during infection by (**A**) Tomato mottle virus (ToMoV) and (**B**) Tomato yellow leaf curl virus (TYLCV) in tomato plants. Relative expression levels were quantified at 21 days post-inoculation. Data are presented as means and error bars represent standard deviation from five (**A**) and three (**B**) biological replications.

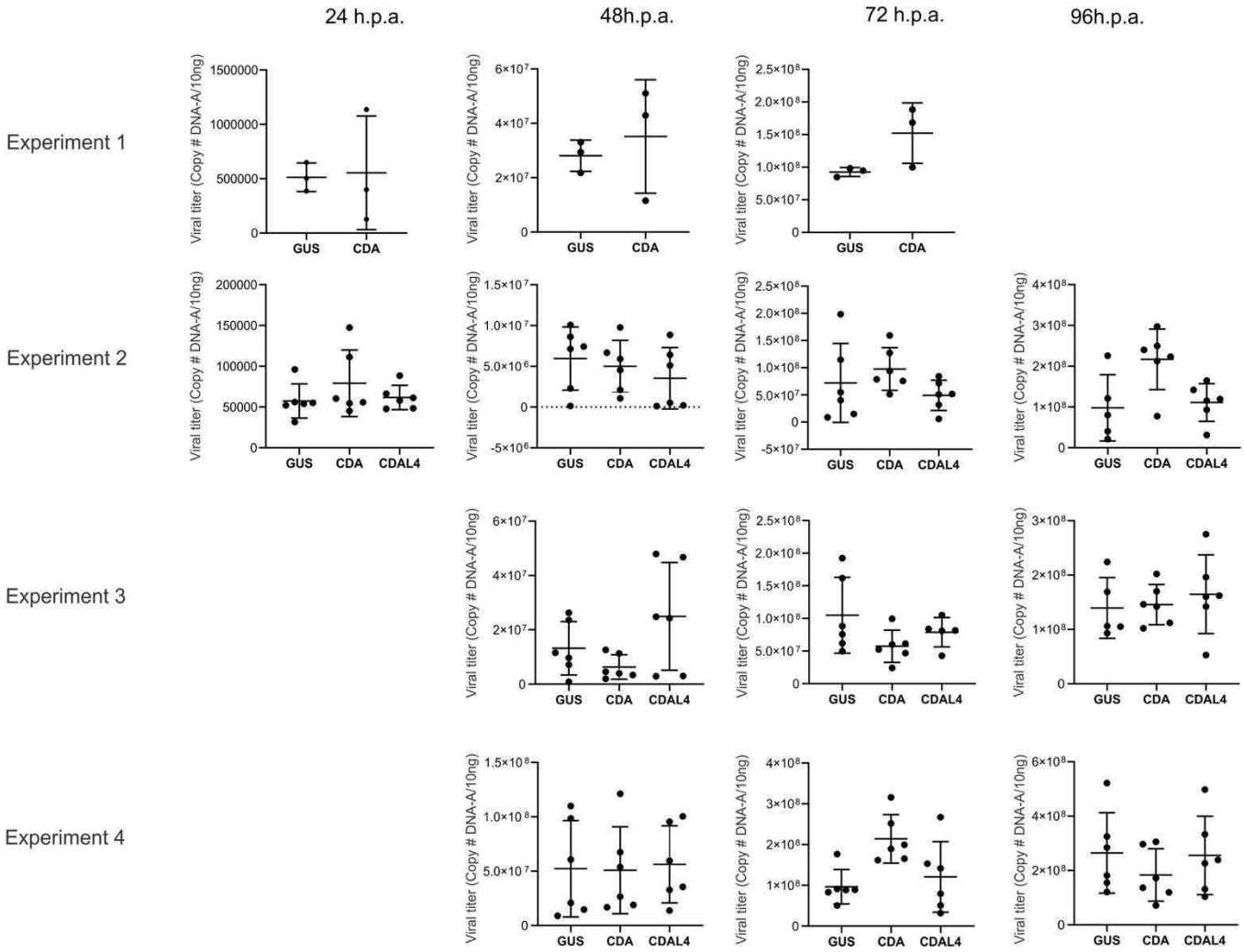
Supplementary Figure S2. Absolute quantification of viral DNA performed at 24, 48, 72 and 96 hours post-agroinfiltration. Expression cassettes harboring AtCDA, AtCDAL4 and GUS were co-infiltrated with CabLCV DNA-A and DNA-B in N. benthamiana leaves. Total DNA was extracted from agroinfiltrated spots at each time and 10 ng of DNA were used for qPCR. Data are presented as means and error bars represent standard deviation shown for each individual experiment. The experiments 2, 3 and 4 were pooled and the results are presented in Figure 4A.

Supplementary Figure S3. AtCDA knockout negatively affects Cabbage leaf curl virus (CabLCV) accumulation but there is no effect in Beet curly top virus (BCTV) and Turnip mosaic virus (TuMV) accumulation. Absolute quantification for CabLCV (**A**) and BCTV (**B**) and relative quantification for TuMV (**C**) were perfomed at 14 days post-inoculation. Data are presented as means and error bars represent standard deviation from five biological replications for CabLCV and TuMV and four biological replications for BCTV, each compose by pools of three plants. Different letters indicate significant differences by Tukey's test ($p \leq 0.05$).

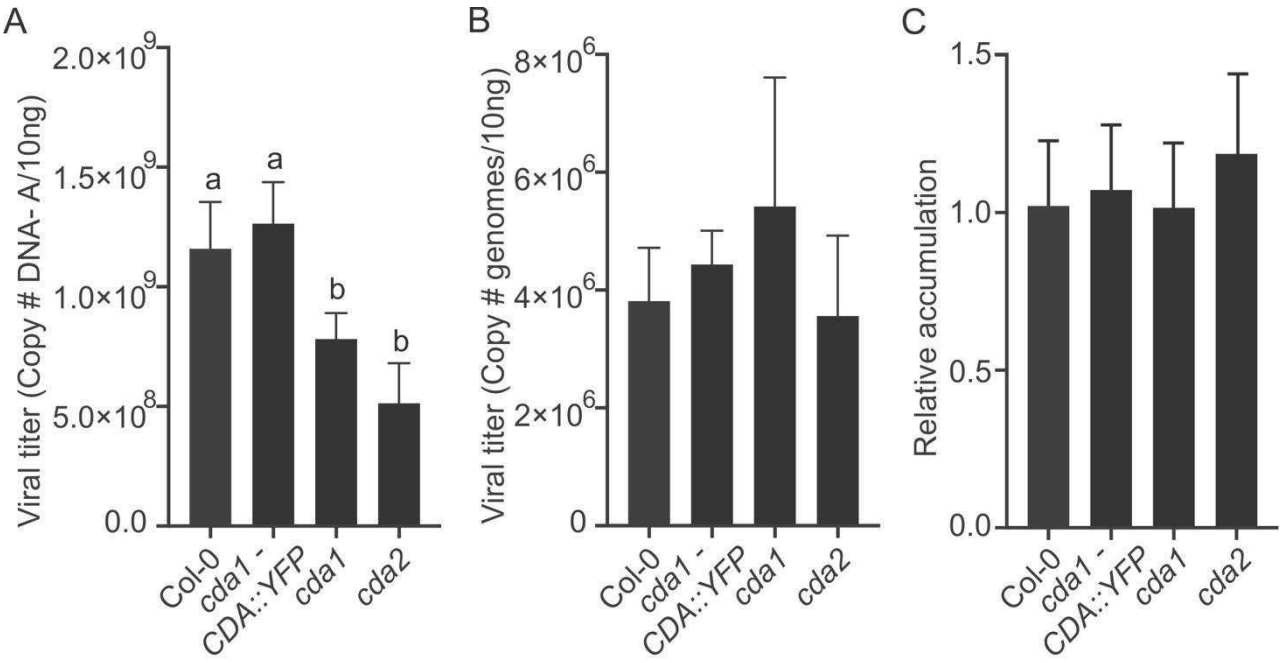
Supplementary Figure S1



Supplementary Figure S2



Supplementary Figure S3



CHAPTER 3

EMERGENCE AND ADAPTATION OF TOMATO BEGOMOVIRUSES IN BRAZIL: ASSESSING REPLICATIVE AND TRANSMISSION FITNESS

Xavier, C.A.D., Nogueira, A.M., Orílio, A.F., Bello, V.H., Krause-Sakate, R., Zerbini, F.M. Emergence and adaptation of tomato begomoviruses in Brazil: assessing replicative and transmission fitness. Virology Journal, in preparation.

Emergence and adaptation of tomato begomoviruses in Brazil: assessing replicative and transmission fitness

César A. Diniz Xavier[1], Angélica M. Nogueira[1], Anelise F. Orílio[1], Vinícius H. Bello[2], Renate Krause-Sakate[2], F. Murilo Zerbini[1*]

[1] Departamento de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil.
[2] Departamento de Proteção Vegetal, Universidade Estadual Paulista, Botucatu, SP 18610-034, Brazil.

*Corresponding author: F. Murilo Zerbini
Phone: (+55-31) 3612-2423
E-mail: zerbini@ufv.br

Abstract

The prevalence of only a few begomovirus species infecting tomatoes in Brazil is an intriguing fact, on light of the great species diversity that has been reported naturally infecting this crop. Most studies that have been performed in the attempt to understand the begomovirus emergence process and consequent epidemics in Brazil have focused on the genetic structure and dispersion patterns of begomovirus populations. Studies addressing the underlying mechanisms leading to emergence and the current patterns of prevalence have not been conducted. Here, we quantified the replicative fitness of two begomoviruses infecting tomato in Brazil, Tomato severe rugose virus (ToSRV) and Tomato yellow spot virus (ToYSV), in tomato plants and in non-cultivated hosts to which each virus has often been associated, and quantified the transmission efficiency by B. tabaci Middle East-Asia Minor 1 (MEAM1) and B. tabaci Mediterranean (MED). Interestingly, ToYSV presented similar adaptation level to the ToSRV in tomato. In addition, no fitness trade-off across host was observed for ToYSV when viral accumulation was evaluated in tomato and wild host Leonurus sibiricus. In contrast, ToSRV performed better in tomato than in Nicandra physalloides. These results reinforce that ToSRV is adaptated tomato virus, that occasionally is found in wild hosts, while ToYSV came from the wild plant L. sibiricus and is adaptated to both hosts. Moreover, we compared the replicative fitness between ToSRV and ToYSV during single or mixed infection. Interestingly, while there were no difference in fitness to the 14 days after inoculation (dpi) between two viruses, from 21 to 28 dpi ToYSV had a gain in fitness and it was higher compared to ToSRV. Furthermore, we got evidence for possible negative interference of ToSRV over ToYSV, as 28 dpi, ToYSV in single infection accumulated more than ToYSV in mixed infection. Thus, these results suggest that adaptation to the host does not explain the prevalence of ToSRV over ToYSV, as both viruses presented similar adaptation levels as evaluated here. However, while ToSRV was transmitted by B. tabaci MEAM1 and MED, ToYSV was not transmitted by them, reason by which this virus probably is not spread in the field.

Introduction

Structural simplicity, large population size and high evolutionary rates confer to viruses a high adaptive potential to an ever-changing environment. Over the last decades, many emerging and re-emerging viral diseases affecting humans, animals and plants have been reported [1-5]. These emergence events are usually associated with increased pathogenicity and disease severity. In the case of plant viruses, the main ecological factors driving virus dispersal and the emergence of viral diseases are changes in agricultural practices and in insect vector populations, with the global spreading of polyphagous, invasive vectors [6, 7]. The spread of so-called supervectors, along with agricultural practices that facilitate the introduction of new hosts and viruses in areas where they do not naturally occur, have greatly contributed to an increase in the potential for emergence of new plant viral diseases [8-10].

Although there are many reports of disease-causing viruses in crops, only a few ones are widely spread and associated with severe epidemics [11], whereas most represent recurrent events of unsuccessful dead-end infection [12]. Adaptation to a new host constitutes a complex process that involves a cascade of genetic and ecological events [13]. Vector-mediated exposure of the virus to a new host, followed by recurrent spillover events, allows enough exposure which, coupled with the high evolutionary rates associated with recurrent selection, may promote adaptation. Moreover, changes that optimize transmission efficiency within the new host population can trigger off a successful emergence event and consequent epidemics [10, 13].

However, fitness trade-offs associated with infection of new hosts may limit the emergence process [14]. Mutations that provide an increase in fitness in a given host may be deleterious in another one [15]. Several studies addressing this adaptive cost during host range expansion of a plant virus have been performed [15-17]. The nature of the adaptive cost and its effect during adaptation to new hosts, although important, are still poorly understood during the process of viral disease emergence [18].

For viruses, fitness is defined as "the capacity of a virus to produce infectious progeny in a given environment" [19]. This concept, also called replicative fitness, has been widely explored as the ability to replicate and/or accumulate is quantified and compared between two or more viral variants [20, 21]. Although replicative fitness may predict the adaptation level of a virus to a host, it may not reflect epidemiological fitness, which is defined as the efficiency of the spreading pattern and dominance (or prevalence) in the field [21]. Due to the obligate relationship between virus and host and the host's finite life, a virus needs to be efficiently

transmitted to become prevalent over time and space. Therefore, transmission efficiency is an important factor driving epidemiological fitness.

Diseases caused by viruses in the genus Begomovirus represent some of the best examples of emerging viral diseases around the world [6, 22-25]. These viruses have a genome composed of one or two circular, single stranded DNA (ssDNA) molecules which are encapsidated in twinned icosahedral particles, and are transmitted by different whitefly species of the Bemisia tabaci complex [26]. The main factors that have contributed to the emergence and successful spread of diseases caused by begomoviruses are the increased dissemination and populational increase of polyphagous vector species [6, 27, 28] and the emergence of aggressive viral variants, mainly due to the ability of these viruses to evolve quickly [25, 29].

In Brazil, the most important diseases caused by begomoviruses are the golden mosaic of common bean caused by Bean golden mosaic virus (BGMV) and rugose mosaic and leaf curl in tomato caused by a begomovirus species complex [30]. To date, eighteen begomoviruses have been described infecting tomatoes in Brazil [31-39]. Interestingly, only a few species are widely disseminated in production fields causing disease, whereas most are restricted to certain regions [25, 35]. This asymmetrical dispersal pattern is believed to be the result of different levels of host adaptation and/or differences in transmission efficiency by the vector [25]. However, studies aiming to quantify begomovirus replicative and transmission fitness in different hosts, which would help to understand these patterns, have not yet been performed.

The emergence of begomoviruses infecting tomato in Brazil occurred after the introduction of B. tabaci Middle East-Asia Minor 1 (B. tabaci MEAM1) in the mid-1990s [40]. B. tabaci MEAM1, unlike the indigenous B. tabaci New World (B. tabaci NW), has a wider host range and greater ability to colonize tomato [41]. The introduction of B. tabaci MEAM1 may have facilitated the transfer of begomoviruses, previously restricted to non-cultivated hosts, to crops of agronomic importance such as the tomato [25, 34, 36, 42, 43]. In addition, Brazilian begomoviruses populations have the capacity to evolve rapidly [25, 44, 45]. Thus, the introduction of a new vector which allowed exposure between viruses and a new host, followed by rapid adaptation due to high evolutionary rates, were the primary factors that allowed the emergence of begomoviruses in tomato in Brazil [25, 46].

Tomato severe rugose virus (ToSRV) is the predominant begomovirus infecting tomatoes in central Brazil [35, 47-50]. Other species such as Tomato mild mosaic virus (ToMiMV), Tomato leaf distortion virus (ToLDV) and Tomato yellow spot virus (ToYSV) have been reported sporadically in the field [25, 34, 51]. Interestingly, ToYSV, ToMiMV and ToLDV are phylogenetically related to begomoviruses that infect non-cultivated hosts [25, 51].

In fact, ToYSV and ToMiMV have often been reported infecting the non-cultivated plants Leonurus sibiricus and Sida urens, respectively [25, 42, 50, 52, 53]. These results strongly suggest that these viruses have non-cultivated plants as their primary, natural hosts and are sporadically found in tomato plants due to recurrent, unsuccessful spillover events, probably due to a failure in host adaptation.

ToYSV causes severe symptoms in tomato including dwarfism, yellow mosaic and leaf curl [54]. Fernandes et al. [50] reported the presence of ToYSV in L. sibiricus plants within tomato fields. However, none of the thirty tomato samples collected in the same field were infected with ToYSV. The efficient way in which ToYSV infects tomato under experimental conditions (artificial inoculation) is consistent with the hypothesis raised by Rocha et al. [25], of vector transmission playing a significant role in terms of begomovirus prevalence in the field. Nevertheless, a high adaptive cost due to host switching could be an additional component that prevents ToYSV prevalence in tomatoes.

Currently, B. tabaci MEAM1 is the predominant whitefly species in Brazil, colonizing crops of economic importance such as soybeans and tomatoes [55, 56]. However, B. tabaci MED was first detected in the country in 2015 [57], and is quickly spreading [56, 58]. The dissemination of this B. tabaci MED is of great concern as it is more polyphagous compared to B. tabaci MEAM1 and has a greater competitive ability, often replacing other species [59-62]. Coupled with a high begomovirus species diversity, the dispersal of B. tabaci MED throughout the country could lead to drastic changes in begomovirus epidemiology not only in tomatoes, but also in other economically important crops in which these viruses have not been a problem so far, such as cotton and soybean.

Here In this work we assessed the replicative fitness of two begomoviruses, ToSRV and ToYSV, in tomato and in non-cultivated hosts with which they have been associated, and analyzed the transmission efficiency by B. tabaci MEAM1 and B. tabaci MED.

Methods

Plant material and viral isolates

Tomato (Solanum lycopersicum cv. Santa Clara), the experimental host N. benthamiana, and the non-cultivated plants Leonurus sibiricus and Nicandra physaloides were used in the experiments. Seeds for non-cultivated plants were obtained from Cosmos Agrícola Produção e Serviços Rurais (Artur Nogueira, SP, Brazil). All plants were grown and kept in a

whitefly-free walk-in growth chamber at 22ºC and a 12 hours light/12 hours dark photoperiod throughout the experiments.

Plants presenting two to four pairs of fully expanded leaves were inoculated using biolistics [63] with infectious clones corresponding to the DNA-A and DNA-B of the isolates ToSRV-[BR:Pir1:05] (infectious clone pVIR096 and pVIR97 for the DNA-A and DNA-B, respectively; [64]) and ToYSV-[Bi2] (infectious clone pToYSV-A1.2 and pToYSV-B1.2 for the DNA-A and DNA- B, respectively; [51]). Ten micrograms of each viral component, previously mixed, were used to inoculate six plants of each species. For mixed infection, infectious clones corresponding to the two viruses were mixed and delivered in the same shoot. Negative controls were inoculated with tungsten particles without DNA. After bombardment, plants were kept in growth chamber and visually observed for symptoms until 28 days post-inoculation (dpi). Viral infection was verified at 28 dpi by PCR using GoTaq Colorless Master Mix (Promega) and primers ToYSV-A-Forward (5'-CAC TTC GTT CGA GAA AAC CTT TGA G-3') and ToYSV-A-Reverse (5'-CAG TAG TTG CCC TCA AAT TGA AG-3') for ToYSV, and ToSRV-A-Forward (5'-CAG TAG TTG CCC TCA AAT TGA AG-3') and ToSRV-A-Reverse (5'-CAC GTG TAG CAA TCT CCT TAA AGA G-3') for ToSRV. The number of infected plants and symptoms were registered.

Viral quantification

Plants were inoculated as described above, and viral quantification was performed at 14, 21 and 28 dpi. For each time point, total DNA was extracted from the second youngest leaf of infected plants as described by Doyle and Doyle [65] and used to determine the absolute viral DNA accumulation by quantitative-real time PCR (qPCR) using virus-specific primers (Suppl. Table 1). DNA samples were quantified by spectrophotometry using a NanoDrop 2000 (Thermo Scientific). Reactions were performed using FastSYBR Green Master Mix (ThermoFisher) in a final volume of 10 µl containing 10 ng total DNA and 200mM of each primer. All reactions were performed in triplicates. Standard curves were prepared using serial dilutions of plasmid DNA containing a genome-length copy of the DNA-A for each virus ($3x10^0$ to $3x10^6$ copies per reaction). Standard curves were obtained by regression analysis of cycle threshold (Ct) values of three replications of a given dilution in relation to the log of the amount of DNA in each dilution. For all reactions, amplification specificity was checked based in dissociation curves. Eight biological replications per treatment from two independent experiment were analyzed. The primers used were: qToYSV4-F (5'-CCC TCA TGC TTG ATT TTC TCC-3') and qToYSV4-R (5'-ACA ATA AAA CTT AGG GAC CAC G-3') for ToYSV

and qToSRVA-R (5'-GCC GTT CAA CAA ATT GGG-3') and qToSRVA-F (5'AAA GTA AAG TGA TTG TCT GTG G-3') for ToSRV. The qPCR cycles consisted of an initial denaturation step at 95°C for 2 min, followed by 40 cycles for 15 sec at 95°C and 30 sec at 60°C for qToSRVA and 64ºC for qToYSV4.

To assess the statistical significance in virus accumulation across treatments, the non-parametric Kruskal-Wallis test was used, followed by post hoc multiple comparison test using Fisher's least significant difference implemented in the Agricolae package in R software [66].

Transmission assays

As a measure of transmission fitness, the efficiency of ToSRV and ToYSV transmission by Bemisia tabaci MEAM1 was compared. All transmission assays were performed in tomato plants and in single infection. The whiteflies used in this study were kept in cabbage plants (Brassica oleracea var. capitata) inside whitefly-proof cages in a room with controlled temperature of 25°C and a 14h light/10h dark photoperiod. To confirm that the colony was aviruliferous, total DNA from fifteen samples composed by pools of five individuals was extracted and subjected to PCR amplification using universal begomovirus primers [67].

In the first assay, eight acquisition-access periods (AAP) were tested: 0 min, 30 min, 1 h, 2 h, 3 h, 6 h, 12 h and 24 h. About 2,000 aviruliferous whiteflies were collected and kept in Falcon tubes for a starvation period of 3 h. Next, whiteflies were transferred to a cage containing ToSRV- or ToYSV-infected tomato plants at 35 dpi. At the end of each AAP, 15 whiteflies were randomly collected using a hand-held aspirator and transferred to tomato healthy plants, on which they were kept for 24 h for virus transmission. As a negative control, aviruliferous whiteflies were transferred to healthy tomato plants. Whiteflies were eliminated mechanically and through application of acetamiprid (80 mg a.i./L). Plants were kept in a greenhouse and observed for symptoms until 35 dpi. At the end of this period, the second youngest leaf of each plant was collected and used for total DNA extraction [65]. The DNA obtained was used as a template for PCR using virus-specific primers (Suppl. Table 1). Results were expressed as the percentage of plants positive for the presence of viral DNA. In this experiment seven plants per treatment were used and two independent experiments were conducted.

In the second assay, eight inoculation-access periods (IAP) were tested: 0 min, 30 min, 1 h, 2 h, 3 h, 6 h, 12 h and 24 h. Aviruliferous whiteflies were placed to feed on tomato plants infected with ToSRV or ToYSV (35 dpi) in a 24 h AAP. Next, 15 viruliferous whiteflies were transferred to healthy tomato plants, remaining in contact with the plants according to each IAP. As a negative control, aviruliferous whiteflies were transferred to healthy tomato plants. After

each IAP, the whiteflies were eliminated as described above, and the plants were kept in a greenhouse until 35 dpi. At the end of this period, the second youngest leaf of each plant was collected and used for viral detection as described above. Seven plants were used per treatment, and two independent experiments were conducted.

In addition, we also carried out a transmission assay using whitefly colonies identified as B. tabaci MEAM1 and B. tabaci MED at UNESP. All experimental procedures were performed as previously described by Nogueira, 2019.

Results and Discussion

The prevalence of only a few begomoviruses infecting tomatoes in Brazil is an intriguing fact, considering the high number of species richness that has been reported naturally in the field. This pattern may be due to two main factors. First, intimate host-virus co-evolution follow by recurrent selection may result in great viral adaptation, causing a virus to efficiently infect a host plant and prevail over other less adapted viruses [25]. Second, differences in transmission efficiency by the vector may result in the prevalence of viruses that are transmitted more efficiently [25]. Although these factors may act in isolation, one possibility does not necessarily rule out the other ones and their relevance may differ for each virus. To obtain an insight on which factors drive the differential predominance of begomoviruses infecting tomato in Brazil, we analyzed replicative and transmission fitness for two begomoviruses that present distinct dispersion patterns. While ToSRV has been demonstrated to be widespread [35, 47-49], ToYSV is rarely detected naturally infecting tomato [25, 54].

First, we analyzed the infectivity rate and latent period in four distinct hosts, N. benthamiana and tomato (common hosts for both viruses), plus two non-cultivated hosts with which these viruses have been frequently associated in the field, L. sibiricus for ToYSV and N. physalloides for ToSRV (Table 1). In N. benthamiana, no differences in infectivity were observed in single or mixed infection, but ToYSV-infected plants presented symptoms much earlier and more severe than ToSRV-infected plants (Table 1, Figure 1A-C). Plants infected with both viruses showed a similar dynamic of infection to that observed for ToYSV alone with regards to symptoms and latent period (Table 1, Figure 1D). In tomato, while no differences in symptoms appearance were observed between ToSRV and ToYSV, again, symptoms induced by ToYSV were more severe (Table 1, Figure 1E-G). In addition, in mixed infection, the symptoms were more severe than in single infections (Figure 1H). Interestingly, a lower infectivity was observed in mixed infections (Table 1). A number of plants that were inoculated with both viruses became infected only with ToSRV (Table 1), indicating that one virus may

be interfering with the other, as previously demonstrated during co-infection between two begomoviruses [68, 69]. Interestingly, Alves-Junior et al. [69] reported a negative interference of ToRMV over ToYSV, even though both viruses seemed to be equally adapted to tomato, with no differences in infectivity and latent period. Silva et al. [68], on the other hand, reported a negative interference of ToRMV over ToSRV. Here, ToSRV seems to be negatively interfering over ToYSV. Since no differences in infectivity and latent period were observed among these three viruses in single infections, differences in transmission efficiency may be the reason why ToSRV is the prevalent virus in the field, while both ToRMV and ToYSV are rarely found in tomato in natural infections.

In non-cultivated hosts, ToYSV and ToSRV performed as well as during tomato infection, showing high infectivity (Table 1). In addition, the latent period of ToYSV in L. sibiricus was the same as observed in tomato, whereas in ToSRV-infected N. physalloides the symptoms appeared later compared to ToSRV-infected tomato (Table 1). Symptoms induced by ToYSV were less severe compared to tomato, with only yellow mosaic without leaf distortion (Figure 1). On the other hand, symptoms induced by ToSRV in tomato were less severe compared with N. physalloides. While ToYSV was able to infect N. physalloides (albeit at a very low rate), ToSRV did not infect L. sibiricus (Table 1). Together these results suggest a differential level of adaptation of ToSRV and ToYSV across hosts.

To better explore this putative differential level of adaptation, we conducted replicative fitness assays for each virus across different hosts. In these assays, viral accumulation was used as a parameter to quantify replicative fitness [21]. No significant differences in ToSRV accumulation in tomato and N. benthamiana was observed at 14, 21 and 28 dpi, whereas ToSRV accumulated significantly lower in N. physalloides in all time points (Figure 2A). These results are consistent with the greater latent period observed in N. physalloides, indicating that ToSRV is not very well adaptated to this non-cultivated host. ToYSV accumulated significantly higher in N. benthamiana than in all other hosts at all time points analyzed (Figure 2B). Interestingly, no differences in ToYSV accumulation were observed in tomato and L. sibiricus (Figure 2B). Consistent with the low infectivity and mild symptoms induced by ToYSV in N. physaloides, viral accumulation was significantly lower in this host compared to all other ones at 21 and 28 dpi (Figure 2B). The replicative fitness assays were repeated several times, with a remarkable consistency in absolute accumulation for both viruses across all hosts. Therefore, it is possible to compare absolute accumulation of the two viruses. Such comparison indicates that ToYSV reaches a titer that is one to three orders of magnitude higher than ToSRV in all

hosts. Together, these results indicate that replicative fitness does not explain the prevalence of ToSRV over ToYSV in tomatoes in the field.

Although the emergence of begomoviruses infecting tomato in Brazil has been attributed to the transfer of indigenous viruses from wild hosts by B. tabaci MEAM1, no experimental studies have addressed this question. Here we obtained strong evidence corroborating a non-cultivated host origin for ToYSV. The phylogenetic relationship of ToYSV with viruses from non-cultivated hosts [51], its high incidence in L. sibiricus under natural conditions [50, 53], and its high infectivity and replicative fitness in this host, reported here, strongly suggest that L. sibiricus is its natural host. Interestingly, this virus performed well in both L. sibiricus and tomato, with no apparent fitness cost due to the switch between the natural and the spill-over host. Bera et al. [15] demonstrated that acquisition of a new host by Tobacco mild green mosaic virus (TMGMV), as ssRNA virus (gen. Tobamovirus, family Virgaviridae), was followed by adaptation by mutational dynamics and selection, but also led to a dramatic loss in fitness in the original host. The ToYSV isolate used here was obtained from tomato [70] and has been maintained in this host (and also in N. benthamiana) for a long time. In contrast with TMGMV, our results indicate that ToYSV seems to have optimized its fitness to both hosts. It will be interesting to test whether an isolate obtained from L. sibiricus would behave in the same way.

Conversely, N. physalloides does not seem to be the natural host of ToSRV, although it may act as an alternate/reservoir host. Mapping mutations across the ToSRV genome for several isolates demonstrated that the oldest mutations are shared by isolates from tomato and pepper, but not from non-cultivated hosts [46], suggesting that this virus probably has been associated to cultivated plants for a long time. Indeed, the results obtained here, based on infectivity and replicative fitness, corroborate this hypothesis, as they suggest a higher adaptation of ToSRV to tomato than to N. physalloides.

To verify whether replicative fitness differs between ToSRV and ToYSV when considering the same host, we compared virus accumulation in tomato and N. benthamiana during single and mixed infections. No significant differences in accumulation between ToSRV and ToYSV were observed at 14 dpi in single or mixed infection of tomato plants (Figure 3A). Interestingly, at 21 dpi the titer of ToYSV was higher than that of ToSRV when either single or mixed infections were compared (Figure 3A). Moreover, the small difference in ToYSV accumulation in single and mixed infection was not statistically significant (Figure 3A). However, this difference was significant at 28 dpi (Figure 3A), again suggesting a negative interference of ToSRV over ToYSV. Nevertheless, ToYSV accumulated to a higher titer than

ToSRV when either single or mixed infections were compared at 28 dpi (Figure 3A). In N. benthamiana, while no differences were observed between single and mixed infection for either virus, ToYSV always accumulated to higher levels than ToSRV at 14, 21, and 28 dpi (Figure 3B). These results demonstrate that the non-prevalent ToYSV has an increased replicative fitness over time compared to the prevalent ToSRV, and reinforce the conclusion from the previous experiment that replicative fitness does not explain the prevalence of ToSRV over ToYSV in tomatoes in the field.

To assess the transmission fitness of ToYSV and ToSRV by B. tabaci MEAM1, the access-acquisition period (AAP) and inoculation-access period (IAP) were evaluated. The results of the transmission experiment demonstrate that B. tabaci MEAM1 was able to acquire ToSRV in a minimum AAP of 30 minutes, with a transmission percentage of 23.1% (Table 2). The percentage of transmission increased with the increase in the AAP, reaching a maximum of 84.6% when the AAP was 12 h (Table 2). In contrast, ToYSV was not transmitted under our experimental conditions, even when B. tabaci MEAM1 was subjected to a 24h AAP (Table 2).

Similar results were obtained for both viruses when the inoculation-access period was evaluated. The minimum IAP for ToSRV was 30 minutes, with a transmission percentage of 23.1% (Table 3). There was an increase in the percentage of transmission with an increase in the IAP, reaching a maximum of 50% when the IAP was 24 h (Table 3). Again, ToYSV was not transmitted even when B. tabaci MEAM1 was subjected to a 24h IAP (Table 3).

To confirm that the non-transmissibility of ToYSV is not specific for the B. tabaci MEAM1 colony used here, a second transmission assay was carried out using B. tabaci MEAM1 and B tabaci MED colonies at UNESP. While ToSRV was transmitted to 80% and 10% of the plants by B. tabaci MEAM1 and B. tabaci MED, respectively, ToYSV was not transmitted by either species (Figure 4A). These results indicate that the ToYSV isolate used here, obtained from tomato, is non-transmissible by two different species of the B. tabaci complex. They also suggest that the recently introduced B. tabaci MED it is not a very efficient vector for ToSRV compared to B. tabaci MEAM1. This should be confirmed using a higher number of plants.

Fernandes et al. [50] reported the presence of ToYSV infecting L. sibiricus within tomato fields. Intriguingly, none of the 33 tomato samples analyzed were infected with ToYSV. Instead, they were all infected with ToSRV. However, they showed that in non-choice experiment, ToYSV was transmitted to both hosts by B. tabaci MEAM1. An ecological barrier could prevent transmission under natural conditions, for instance a low suitability of L. sibiricus to B. tabaci MEAM1. If this was true, a distinct whitefly species, for which tomato would not

be a host, could be transmitting the virus among L. sibiricus plants. The characterization of whitefly populations from L. sibiricus would be required to test this hypothesis.

To verify possible differences in the capsid protein (CP) between the ToYSV isolate used here and isolates from L. sibiricus, which could explain the difference in transmissibility, we performed an amino acid alignment of 20 ToYSV sequences: two isolated from tomato, including the isolate used here (DQ336350), and 18 isolates from L. sibiricus, including the isolate used by Fernandes et al. [50] (JQ429791; Figure 4B). Interestingly, only one difference was detected that differentiates between tomato and non-tomato isolates (Figure 4B). While aspartic acid in the position 5 of the N-terminal region was conserved in all non-tomato isolates, tomato isolates have a glycine at this position. In addition, we also compared all ToYSV CPs against 144 CPs from New World begomoviruses (data not shown). Interestingly, this aspartic acid residue was present in 91% of the sequences. An additional 7% of the sequences have a glutamic acid at this position (same group as aspartic acid), and only 2% of the analyzed CPs have other amino acids). These results suggest that this site may be a determinant of begomovirus transmissibility by the whitefly vector. Additional studies are needed to confirm this hypothesis.

Based on the results presented, it can be concluded that the prevalence of ToSRV, at least in relation to ToYSV, in tomato crops in the Southeast and Midwest regions of Brazil can be attributed to a higher transmission efficiency by the insect vector, as the well-adaptated ToYSV was not transmissible under our experimental conditions. The factors determining the transmission of ToYSV isolates infecting L. sibiricus to tomatoes in natural conditions remain to be determined. Understanding these factors is important to prevent epidemics, allowing the development of strategies aimed at minimizing the damage and reducing the risks associated with the emergence of new diseases.

References

1. Chitti SV, Prasad AK, Saxena SK: **Emerging Zika virus disease: a public health emergency of global concern.** Virus disease 2016, **27:**211-214.
2. Zhou J, Tzanetakis IE: **Soybean vein necrosis virus: an emerging virus in North America.** Virus Genes 2019, **55:**12-21.
3. Gray SM, Power AG: **Anthropogenic influences on emergence of vector-borne plant viruses: the persistent problem of Potato virus Y.** Curr Opin Virol 2018, **33:**177-183.
4. Jacobson AL, Duffy S, Sseruwagi P: **Whitefly-transmitted viruses threatening cassava production in Africa.** Current Opinion in Virology 2018, **33:**167-176.

5. Ketkar H, Herman D, Wang P: **Genetic Determinants of the Re-Emergence of Arboviral Diseases.** 2019, **11**.

6. Gilbertson RL, Batuman O, Webster CG, Adkins S: **Role of the insect supervectors Bemisia tabaci and Frankliniella occidentalis in the emergence and global spread of plant viruses.** Annual Review of Virology 2015, **2:**67-93.

7. Roossinck MJ, Garcia-Arenal F: **Ecosystem simplification, biodiversity loss and plant virus emergence.** Current Opinion in Virology 2015, **10:**56-62.

8. Gessain A, Garcia-Arenal F: **Editorial overview: Emerging viruses: interspecies transmission.** Curr Opin Virol 2015, **10:**v-viii.

9. Pagán I, González-Jara P, Moreno-Letelier A, Rodelo-Urrego M, Fraile A, Piñero D, García-Arenal F: **Effect of biodiversity changes in disease risk: Exploring disease emergence in a plant-virus system.** PLoS Pathogens 2012, **8:**e1002796.

10. Elena SF, Fraile A, Garcia-Arenal F: **Evolution and emergence of plant viruses.** Advances in Virus Research 2014, **88:**161-191.

11. Scholthof KB, Adkins S, Czosnek H, Palukaitis P, Jacquot E, Hohn T, Hohn B, Saunders K, Candresse T, Ahlquist P, et al: **Top 10 plant viruses in molecular plant pathology.** Mol Plant Pathol 2011, **12:**938-954.

12. Power AG, Mitchell CE: **Pathogen spillover in disease epidemics.** American Naturalist 2004, **164:**S79-S89.

13. Elena SF, Bedhomme S, Carrasco P, Cuevas JM, Iglesia Fdl, Lafforgue G, Lalić J, Pròsper À, Tromas N, Zwart MP: **The evolutionary genetics of emerging plant RNA viruses.** Molecular Plant Microbe Interactions 2011, **24:**287-293.

14. Lalic J, Cuevas JM, Elena SF: **Effect of host species on the distribution of mutational fitness effects for an RNA virus.** PLoS Genetics 2011, **7**.

15. Bera S, Fraile A, García-Arenal F: **Analysis of Fitness Trade-Offs in the Host Range Expansion of an RNA Virus, Tobacco Mild Green Mosaic Virus.** Journal of Virology 2018, **92:**e01268-01218.

16. García-Arenal F, Fraile A: **Trade-offs in host range evolution of plant viruses.** Plant Pathology 2013, **62:**2-9.

17. Cervera H, Lalić J, Elena SF: **Effect of Host Species on Topography of the Fitness Landscape for a Plant RNA Virus.** Journal of Virology 2016, **90:**10160-10169.

18. Elena SF, Agudelo-Romero P, Lalić J: **The evolution of viruses in multi-host fitness landscapes.** The open virology journal 2009, **3:**1-6.

19. Domingo E, Holland JJ: **RNA virus mutations and fitness for survival.** Annual Review of Microbiology 1997, **51:**151-178.

20. Quinones-Mateu ME, Arts EJ: **Virus fitness: concept, quantification, and application to HIV population dynamics.** Current Topics in Microbiology Immunology 2006, **299:**83-140.

21. Wargo AR, Kurath G: **Viral fitness: definitions, measurement, and current insights.** Curr Opin Virol 2012, **2:**538-545.

22. Moriones E, Navas-Castillo J, Diaz-Pendon JA: **Emergence of Begomovirus Diseases.** In Recent Advances in Plant Virology. Edited by Caranta C, Aranda MA, Tepfer M, Lopez-Moya JJ. Norfolk, UK: Caister Academic Press; 2011: 301-320

23. Moriones E, Praveen S, Chakraborty S: **Tomato leaf curl New Delhi virus: An emerging virus complex threatening vegetable and fiber crops.** Viruses 2017, **9:**264.

24. Mabvakure B, Martin DP, Kraberger S, Cloete L, van Brunschot S, Geering AD, Thomas JE, Bananej K, Lett JM, Lefeuvre P, et al: **Ongoing geographical spread of Tomato yellow leaf curl virus.** Virology 2016, **498:**257-264.

25. Rocha CS, Castillo-Urquiza GP, Lima ATM, Silva FN, Xavier CAD, Hora-Junior BT, Beserra-Junior JEA, Malta AWO, Martin DP, Varsani A, et al: **Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location.** Journal of Virology 2013, **87:**5784-5799.

26. Zerbini FM, Briddon RW, Idris A, Martin DP, Moriones E, Navas-Castillo J, Rivera-Bustamante R, Varsani A, ICTV Consortium: **ICTV Virus Taxonomy Profile: Geminiviridae.** Journal of General Virology 2017, **98:**131-133.

27. Hadjistylli M, Roderick GK, Brown JK: **Global population structure of a worldwide pest and virus vector: Genetic diversity and population history of the Bemisia tabaci sibling species group.** PLoS ONE 2016, **11:**e0165105.

28. Navas-Castillo J, Fiallo-Olivé E, Sánchez-Campos S: **Emerging virus diseases transmitted by whiteflies.** Annual Review of Phytopathology 2011, **49:**219-248.

29. Duffy S, Holmes EC: **Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus Tomato yellow leaf curl virus.** Journal of Virology 2008, **82:**957-965.

30. Inoue-Nagata AK, Lima MF, Gilbertson RL: **A review of geminivirus diseases in vegetables and other crops in Brazil: current status and approaches for management.** Horticultura Brasileira 2016, **34:**8-18.

31. Ribeiro SG, Ambrozevicius LP, Ávila AC, Bezerra IC, Calegario RF, Fernandes JJ, Lima MF, Mello RN, Rocha H, Zerbini FM: **Distribution and genetic diversity of tomato-infecting begomoviruses in Brazil.** Archives of Virology 2003, **148:**281-295.

32. Fernandes JJ, Carvalho MG, Andrade EC, Brommonschenkel SH, Fontes EPB, Zerbini FM: **Biological and molecular properties of Tomato rugose mosaic virus (ToRMV), a new tomato-infecting begomovirus from Brazil.** Plant Pathology 2006, **55:**513-522.

33. Ribeiro SG, Lohuis H, Goldbach R, Prins M: **Tomato chlorotic mottle virus is a target of RNA silencing but the presence of specific short interfering RNAs does not guarantee resistance in transgenic plants.** Journal of Virology 2007, **81:**1563-1573.

34. Castillo-Urquiza GP, Beserra Jr. JEA, Bruckner FP, Lima ATM, Varsani A, Alfenas-Zerbini P, Zerbini FM: **Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil.** Archives of Virology 2008, **153:**1985-1989.

35. Fernandes FR, Albuquerque LC, Giordano LB, Boiteux LS, Ávila AC, Inoue-Nagata AK: **Diversity and prevalence of Brazilian bipartite begomovirus species associated to tomatoes.** Virus Genes 2008, **36:**251-258.

36. Albuquerque LC, Varsani A, Fernandes FR, Pinheiro B, Martin DP, Ferreira PTO, Lemos TO, Inoue-Nagata AK: **Further characterization of tomato-infecting begomoviruses in Brazil.** Archives of Virology 2012, **157:**747-752.

37. Macedo MA, Albuquerque LC, Maliano MR, Souza JO, Rojas MR, Inoue-Nagata AK, Gilbertson RL: **Characterization of tomato leaf curl purple vein virus, a new monopartite New World begomovirus infecting tomato in Northeast Brazil.** Arch Virol 2017, **163:**737-743.

38. Rego-Machado CM, Nakasu EYT, Blawid R, Nagata T, Inoue-Nagata AK: **Complete genome sequence of a new bipartite begomovirus infecting tomato in Brazil.** Archives of Virology 2019, **20:**019-04380.

39. Quadros AFF, Silva JP, Xavier CAD, Zerbini FM, Boari AJ: **Two new begomoviruses infecting tomato and Hibiscus sp. in the Amazon region of Brazil.** Archives of virology 2019**:**1-5.

40. Lourenção AL, Nagai H: **Surtos populacionais de Bemisia tabaci no Estado de São Paulo.** Bragantia 1994, **53:**53-59.

41. Bedford ID, Briddon RW, Brown JK, Rosell RC, Markham PG: **Geminivirus transmission and biological characterization of Bemisia tabaci (Gennadius) biotypes from different geographical regions.** Annals of Applied Biology 1994, **125:**311-325.

42. Tavares SS, Ramos-Sobrinho R, Gonzalez-Aguilera J, Lima GSA, Assunção IP, Zerbini FM: **Further molecular characterization of weed-associated begomoviruses in Brazil with an emphasis on Sida spp.** Planta Daninha 2012, **30:**305-315.

43. Barreto SS, Hallwass M, Aquino OM, Inoue-Nagata AK: **A study of weeds as potential inoculum sources for a tomato-infecting begomovirus in central Brazil.** Phytopathology 2013, **103:**436-444.

44. Lima ATM, Sobrinho RR, Gonzalez-Aguilera J, Rocha CS, Silva SJC, Xavier CAD, Silva FN, Duffy S, Zerbini FM: **Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts.** Journal of General Virology 2013, **94:**418-431.

45. Ramos-Sobrinho R, Xavier CAD, Pereira HMB, Lima GSA, Assunção IP, Mizubuti ESG, Duffy S, Zerbini FM: **Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts.** Journal of General Virology 2014, **95:**2540-2552.

46. García-Arenal F, Zerbini FM: **Life on the edge: geminiviruses at the interface between crops and wild plant hosts.** Annual Review of Virology 2019, **6**.

47. Cotrim MA, Krause-Sakate R, Narita N, Zerbini FM, Pavan MA: **Genetic diversity ot tomato-infecting begomoviruses in Central São Paulo state (in Portuguese).** Summa Phytopathologica 2007, **33:**300-303.

48. Lima ATM, Pereira CO, Alfenas PF, Paula MB, Mello RN, Zerbini FM: **Primeiro relato de infecção pelo geminivírus Tomato severe rugose virus (ToSRV) em tomateiro no estado de Santa Catarina.** Fitopatologia Brasileira 2006, **31(Suplemento):**S224.

49. González-Aguilera J, Tavares SS, Sobrinho RR, Xavier CAD, Dueñas-Hurtado F, Lara-Rodrigues RM, Silva DJH, Zerbini FM: **Genetic structure of a Brazilian population of the begomovirus Tomato severe rugose virus (ToSRV).** Tropical Plant Pathology 2012, **37:**346-353.

50. Fernandes NAN, Boiteux LS, Fonseca MEN, Gonzales-Segnana L, Kitajima EW: **Report of Tomato yellow spot virus infecting Leonurus sibiricus in Paraguay and within tomato fields in Brazil.** Plant Disease 2014, **98:**1445.

51. Andrade EC, Manhani GG, Alfenas PF, Calegario RF, Fontes EPB, Zerbini FM: **Tomato yellow spot virus, a tomato-infecting begomovirus from Brazil with a closer relationship to viruses from Sida sp., forms pseudorecombinants with begomoviruses from tomato but not from Sida.** Journal of General Virology 2006, **87:**3687-3696.

52. Barbosa JC, Eckstein B, Bergamin-Filho A, Rezende JAM, Dallagnol LJ: **First report of Tomato yellow spot virus infecting Leonurus sibiricus in Brazil.** Plant Disease 2012, **97:**289-289.

53. Ferro CG, Silva JP, Xavier CAD, Godinho MT, Lima ATM, Mar TB, Lau D, Zerbini FM: **The ever increasing diversity of begomoviruses infecting non-cultivated hosts: new species from Sida spp. and Leonurus sibiricus, plus two New World alphasatellites.** Annals of Applied Biology 2017, **170:**204-218.

54. Calegario RF, Ferreira SS, Andrade EC, Zerbini FM: **Characterization of Tomato yellow spot virus (ToYSV), a novel tomato-infecting begomovirus from Brazil.** Pesquisa Agropecuária Brasileira 2007, **42:**1335-1343.

55. Marubayashi JM, Kliot A, Yuki VA, Rezende JA, Krause-Sakate R, Pavan MA, Ghanim M: **Diversity and localization of bacterial endosymbionts from whitefly species collected in Brazil.** PLoS One 2014, **9:**e108363.

56. Moraes LA, Muller C, Bueno R, Santos A, Bello VH, De Marchi BR, Watanabe LFM, Marubayashi JM, Santos BR, Yuki VA, et al: **Distribution and phylogenetics of whiteflies and their endosymbiont relationships after the Mediterranean species invasion in Brazil.** Scientific Reports 2018, **8:**14589.

57. Barbosa LF, Yuki VA, Marubayashi JM, De Marchi BR, Perini FL, Pavan MA, Barros DR, Ghanim M, Moriones E, Navas-Castillo J, Krause-Sakate R: **First report of Bemisia tabaci Mediterranean (Q biotype) species in Brazil.** Pest Management Science 2015, **71:**501-504.

58. Moraes LA, Marubayashi JM, Yuki VA, Ghanim M, Bello VH, De Marchi BR, Barbosa LF, Boykin LM, Krause-Sakate R, Pavan MA: **New invasion of Bemisia tabaci Mediterranean species in Brazil associated to ornamental plants.** Phytoparasitica 2017, **45:**517-525.

59. Teng X, Wan FH, Chu D: **Bemisia tabaci biotype Q dominates other biotypes across China.** Florida Entomologist 2010, **93:**363-368.

60. Chu D, Wan FH, Zhang YJ, Brown JK: **Change in the biotype composition of Bemisia tabaci in Shandong province of China from 2005 to 2008.** Environmental Entomology 2010, **39:**1028-1036.

61. Watanabe LFM, Bello VH, De Marchi BR, Silva FBd, Fusco LM, Sartori MM, Pavan MA, Krause-Sakate R: **Performance and competitive displacement of Bemisia tabaci MEAM1 and MED cryptic species on different host plants.** Crop Protection 2019, **124:**104860.

62. Vyskočilová S, Seal S, Colvin J: **Relative polyphagy of "Mediterranean" cryptic Bemisia tabaci whitefly species and global pest status implications.** Journal of Pest Science 2019, **92:**1071-1088.

63. Aragão FJL, Barros LMG, Brasileiro ACM, Ribeiro SG, Smith FD, Sanford JC, Faria JC, Rech EL: **Inheritance of foreign genes in transgenic bean (Phaseolus vulgaris L.) co-transformed via particle bombardment.** Theoretical and Applied Genetics 1996, **93:**142-150.

64. Lima ATM: **Caracterização de dois begomovírus (Tomato severe rugose virus e Tomato yellow vein streak virus) que infectam tomateiro e obtenção de clones infecciosos.** Tese M.S. Universidade Federal de Viçosa, Dep de Fitopatologia; 2007.

65. Doyle JJ, Doyle JL: **A rapid DNA isolation procedure for small amounts of fresh leaf tissue.** Phytochemical Bulletin 1987, **19:**11-15.

66. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2007.

67. Rojas MR, Gilbertson RL, Russell DR, Maxwell DP: **Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses.** Plant Disease 1993, **77:**340-347.

68. Silva FN, Lima ATM, Rocha CS, Castillo-Urquiza GP, Alves-Júnior M, Zerbini FM: **Recombination and pseudorecombination driving the evolution of the begomoviruses Tomato severe rugose virus (ToSRV) and Tomato rugose mosaic virus (ToRMV): Two recombinant DNA-A components sharing the same DNA-B.** Virology Journal 2014, **11:**66.

69. Alves-Junior M, Alfenas-Zerbini P, Andrade EC, Esposito DA, Silva FN, Cruz ACF, Ventrella MC, Otoni WC, Zerbini FM: **Synergism and negative interference during co-**

infection of tomato and Nicotiana benthamiana with two bipartite begomoviruses. Virology 2009, **387:**257-266.

70. Ambrozevicius LP, Calegario RF, Fontes EPB, Carvalho MG, Zerbini FM: **Genetic diversity of begomoviruses infecting tomato and associated weeds in Southeastern Brazil.** Fitopatologia Brasileira 2002, **27:**372-377.

Table 1. Infectivity and latent period of the begomoviruses Tomato severe rugose virus (ToSRV) and Tomato yellow spot virus (ToYSV) in different hosts inoculated by biolistics.

Virus	Nicotiana benthamiana		Solanum lycopersicum		Nicandra physalloides		Leonurus sibiricus	
	I (%)[*]	LP[#]	I (%)	LP	I (%)	LP	I (%)	LP
ToSRV	16/16 (100)	10	15/16 (93.75)	10	14/16 (87.50)	15	0/16 (0)	-
ToYSV	15/16 (93.75)	5	14/16 (87.50)	10	4/16 (25)	12	16/16 (100)	10
ToSRV+ToYSV[€]	16/16 (100)	5	10/16 (68.75)	10	0/16 (0)	-	0/16 (0)	-
ToSRV (ToSRV+ToYSV)[&]	0/16 (0)	-	4/16 (25.00)	10	12/16 (75)	15	0/16 (0)	-
ToYSV (ToSRV+ToYSV)[£]	0/16 (0)	-	2/16 (12.50)	10	0/16 (0)	-	14/16 (87.5)	10

[*] Infectivity, expressed as the number of infected plants/number inoculated plants detected by PCR at 28 days post-inoculation.
[#] Latent period, expressed as the number of days between inoculation and onset of symptoms.
[€] Plants inoculated with both viruses in which both viruses were detected.
[&] Plants inoculated with both viruses in which only ToSRV was detected.
[£] Plants inoculated with both viruses in which only ToYSV was detected.

Table 2. Transmission efficiency of the begomoviruses Tomato severe rugose virus (ToSRV) and Tomato yellow spot virus (ToYSV) by Bemisia tabaci MEAM1 to tomato (S. lycopersicum cv. Santa Clara) after different acquisition-access periods.

Virus	**Experiment**	I**nfected plants/inoculated plants (%)***							
		0[#]	0.5	1	2	3	6	12	24
ToSRV	1	0/7 (0)	2/7 (28.6)	5/7 (71.4)	4/7 (57.1)	5/7 (71.4)	6/7 (85.7)	7/7 (100.0)	6/7 (85.7)
	2	0/6 (0)	1/6 (16.7)	2/5 (40)	4/5 (80.0)	2/6 (33.3)	4/5 (80.0)	4/6 (66.7)	4/6 (66.7)
	Avg	0/13 (0)	3/13 (23.1)	7/12 (58.3)	8/12 (66.7)	7/13 (53.8)	10/12 (83.3)	11/13 (84.6)	10/13 (76.9)
ToYSV	1	0/7 (0)	0/7 (0)	0/7 (0)	0/7 (0)	0/7 (0)	0/7 (0)	0/7 (0)	0/7 (0)
	2	0/5 (0)	0/4 (0)	0/5 (0)	0/5 (0)	0/5 (0)	0/6 (0)	0/5 (0)	0/4 (0)
	Avg	0/12 (0)	0/11 (0)	0/12 (0)	0/12 (0)	0/12 (0)	0/13 (0)	0/12 (0)	0/11 (0)

* Number of infected plants determined by PCR using virus-specific primers, at 35 days post-inoculation.
[#]Length of the acquisition-access period, in hours.

Table 3. Transmission efficiency of the begomoviruses Tomato severe rugose virus (ToSRV) and Tomato yellow spot virus (ToYSV) by Bemisia tabaci MEAM1 to tomato (S. lycopersicum cv. Santa Clara) after different inoculation-access periods.

Virus	**Experiment**	I**nfected plants/inoculated plants (%)***							
		0[#]	0.5	1	2	3	6	12	24
ToSRV	1	0/7 (0)	2/7 (28.6)	0/7 (0)	0/7 (0)	3/7 (42.9)	3/7 (42.9)	3/7 (42.9)	4/6 (66.7)
	2	0/6 (0)	1/6 (16.7)	1/6 (16.7)	2/6 (33.3)	1/6 (16.7)	2/6 (33.3)	3/6 (50.0)	2/6 (33.33)
	Avg	0/13 (0)	3/13 (23.1)	1/13 (7.7)	2/13 (15.4)	4/13 (30.8)	5/13 (38.5)	6/13 (46.2)	6/12 (50.0)
ToYSV	1	0/7 (0)	0/7 (0)	0/7 (0)	0/7 (0)	0/7 (0)	0/7 (0)	0/7 (0)	0/7 (0)
	2	0/5 (0)	0/6 (0)	0/6 (0)	0/6 (0)	0/5 (0)	0/6 (0)	0/6 (0)	0/6 (0)
	Avg	0/12 (0)	0/13 (0)	0/13 (0)	0/13 (0)	0/12 (0)	0/13 (0)	0/13 (0)	0/13 (0)

* Number of infected plants determined by PCR using virus-specific primers, at 35 days post-inoculation.
[#]Length of the inoculation-access period, in hours.

Figure legends

Figure 1. Symptoms induced by Tomato severe rugose virus (ToSRV) and Tomato yellow spot virus (ToYSV) in different hosts in single or mixed infection at 28 days post-biolistic inoculation. Mock: plants biolistically inoculated with tungsten particles without DNA. **A-D,** Nicotiana benthamiana; **E-H,** Tomato (Solanum lycopersicum); **I-J,** Leonurus sibiricus; **K,** Nicandra physaloides.

Figure 2. Accumulation of (**A**) Tomato severe rugose virus (ToSRV) and (**B**) Tomato yellow spot virus (ToYSV) during infection of tomato (S. lycopersicum) and the non-cultivated hosts Nicotiana benthamiana, Leonurus sibiricus and Nicandra physaloides. Absolute quantification of viral DNA was performed at 14, 21 and 28 days post-inoculation by quantitative-real time PCR using total DNA extracted from systemically infected leaves. Specific DNA-A primers were used for each virus. Each dot on the scatter plot represents a biological replication from two independent experiments. The median is indicated as a horizontal bar. Different letters indicate significant differences between groups according to the non-parametric Kruskal-Wallis test followed by post hoc multiple comparison test using Fisher's least significant difference ($p < 0.05$).

Figure 3. Virus accumulation of Tomato severe rugose virus (ToSRV) and Tomato yellow spot virus (ToYSV) in two common hosts in single or mixed infection. Absolute quantification of ToSRV and ToYSV were performed during a time-course experiment in infected (**A**) tomato (S. lycopersicum) and (**B**) N. benthamiana at 14, 21 and 28 days post-biolistic inoculation. Viral load was determined by quantitative-real time PCR using total DNA extracted from systemically infected leaves. Specific DNA-A primers were used for each virus. Each dot on the scatter plot represents a biological replication from two independent experiments. The median is indicated as a horizontal bar. Different letters indicate significant differences between groups according to the non-parametric Kruskal-Wallis test followed by post hoc multiple comparison test using Fisher's least significant difference ($p < 0.05$).

Figure 4. Transmission efficiency of Tomato severe rugose virus (ToSRV) and Tomato yellow spot virus (ToYSV) by B. tabaci MEAM1-UNESP and B. tabaci MED. **A.** Agarose gel depicting viral DNA detection by PCR using virus-specific primers. Total DNA used as a template was extracted at 35 days post-inoculation by B. tabaci MEAM1-UNESP or B. tabaci MED. For each combination of virus-whitefly, 10 biological replications were used. **B.** Alignment of the amino acid sequence of the ToYSV capsid protein, including the isolate used in this study (DQ336350, first row) and additional isolates retrieved from GenBank. Access numbers are indicated and highlighted according to the host from which the isolates were obtained: dark gray from tomato and light gray from hosts other than tomato. ToYSV isolate used by [50] in transmission assay is underlined. Variable residues across the alignment

are in gray and sites containing amino acids which are specific for tomato isolates is indicated by an arrow and highlighted in red.

Figure 1

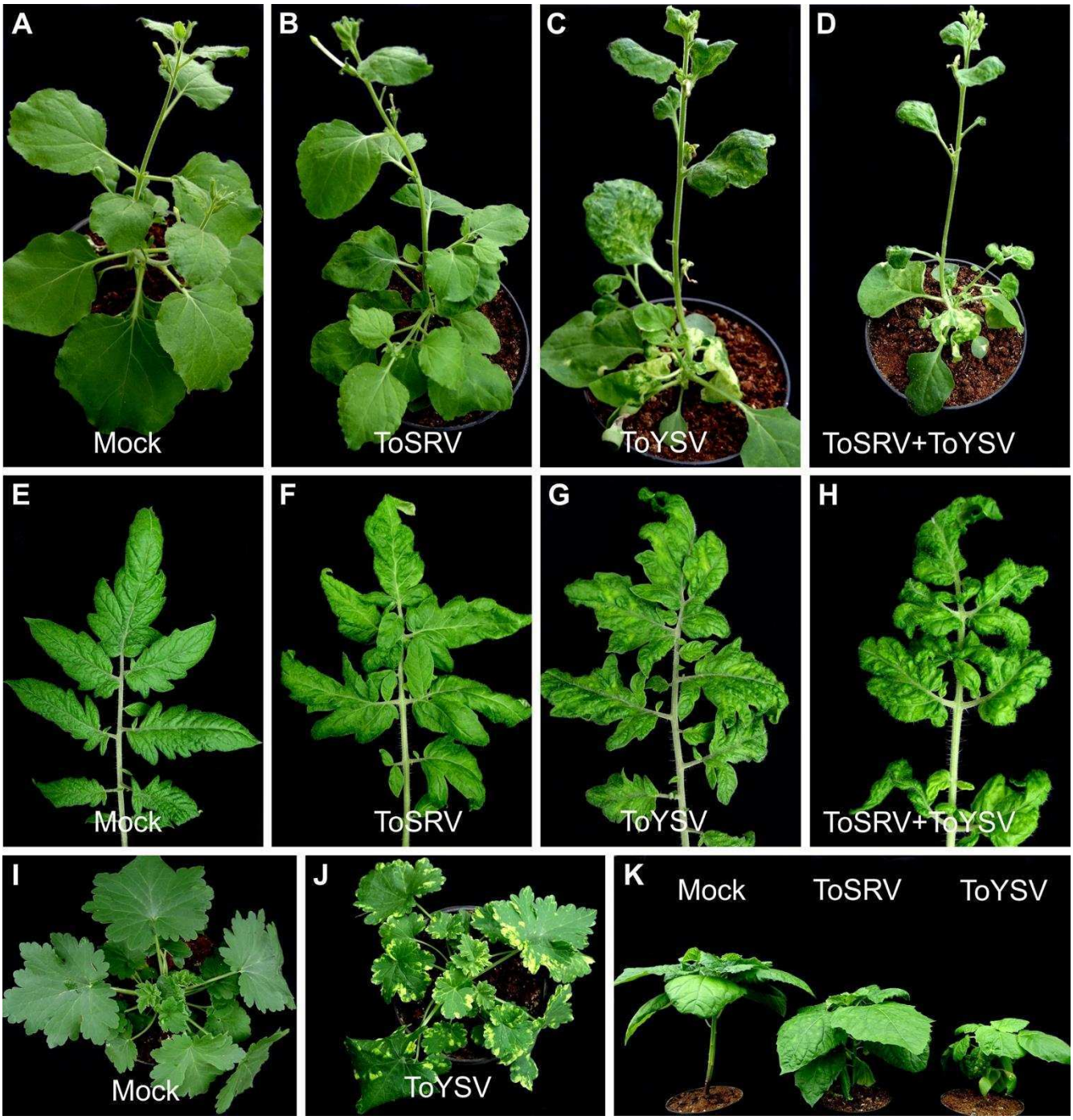

A
Mock
B
ToSRV
C
ToYSV
D
ToSRV+ToYSV
E
Mock
F
ToSRV
G
ToYSV
H
ToSRV+ToYSV
I
Mock
J
ToYSV
K
Mock
ToSRV
ToYSV

Figure 2

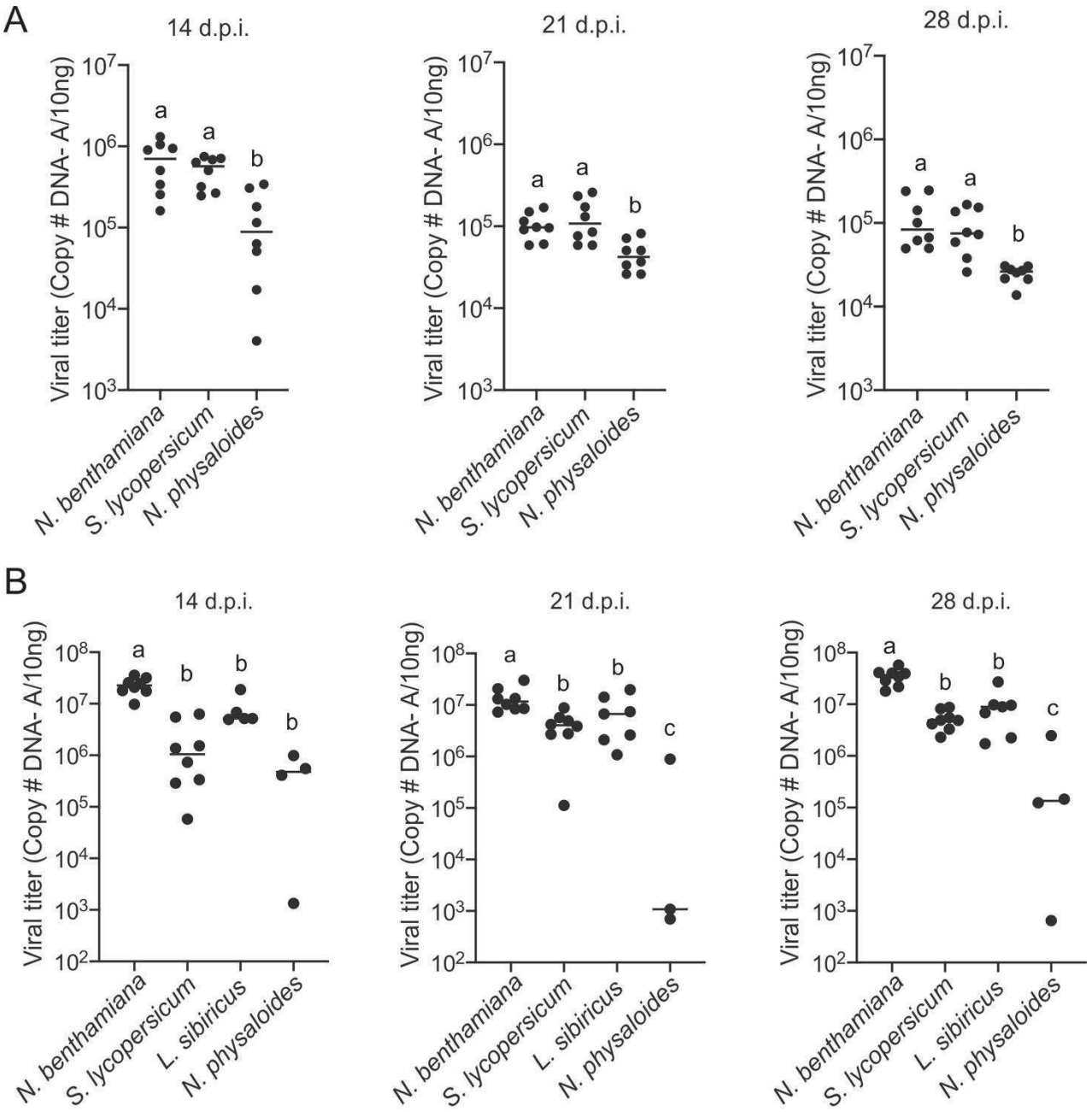

A
14 d.p.i.
21 d.p.i.
28 d.p.i.
Viral titer (Copy # DNA- A/10ng)
10^7
10^6
10^5
10^4
10^3
a
a
b
N. benthamiana
S. lycopersicum
N. physaloides
B
14 d.p.i.
21 d.p.i.
28 d.p.i.
Viral titer (Copy # DNA- A/10ng)
10^8
10^7
10^6
10^5
10^4
10^3
10^2
a
b
b
c
N. benthamiana
S. lycopersicum
L. sibiricus
N. physaloides

Figure 3

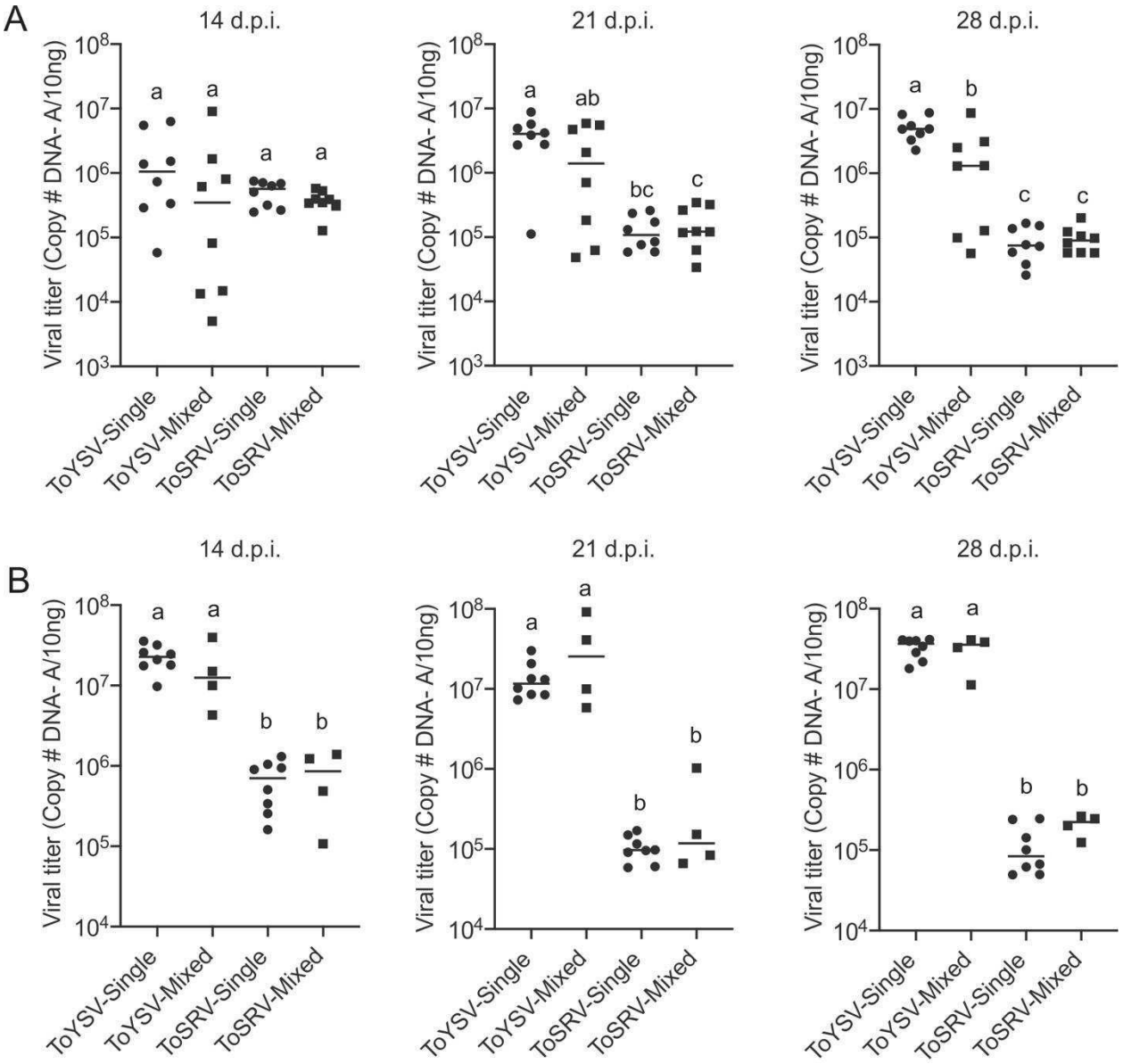
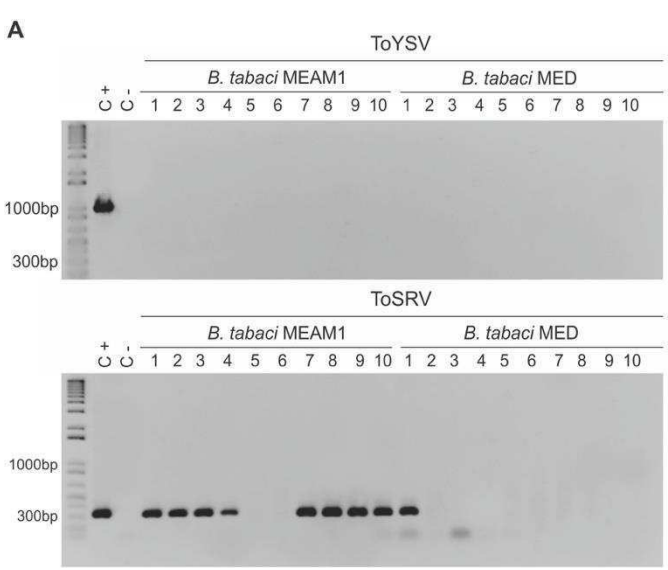

A
14 d.p.i.
21 d.p.i.
28 d.p.i.
Viral titer (Copy # DNA- A/10ng)
ToYSV-Single
ToYSV-Mixed
ToSRV-Single
ToSRV-Mixed
B
14 d.p.i.
21 d.p.i.
28 d.p.i.
Viral titer (Copy # DNA- A/10ng)
ToYSV-Single
ToYSV-Mixed
ToSRV-Single
ToSRV-Mixed

Figure 4



B

		% Identity (aa)
DQ336350	MPKRGPSWRPMVGTSKVSRTSNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	-
KC706628	MPKRGPSWRPMVGTSKVSRTSNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	100
KX348172	MPKRDPSWRPMAGTTKVSRTSNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.81
KX348168	MPKRDPSWRPMVGTTKVSRISNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.81
JQ429791	MPKRDPSWRPMVGTTKVSRTSNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	99.20
KX348167	MPKRDPSWRPMVGTTKVSRTSNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	99.20
KJ742419	MPKRDPSWRPMVGTSKVSRTSNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	99.60
FJ538207	MPKRDPSWRPMVGTSKVSRTSNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	99.60
KX348175	MPKRDPSWRPMVGTSKISRTTNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.81
KX348176	MPKRDPSWRPMVGTSKISRTTNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.81
KX348178	MPKRDPSWRPMVGTSKISRTTNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.81
KX348179	MPKRDPSWRPMVGTSKISRTSNF-8-FNKASEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	99.20
KX348169	MPKRDPSWRPMVGTTKVSRTSNF-8-FNKAHEWVNRPM-80-LVRDRRRYGTPMD-147-DSITN*	98.40
KX348166	MPKRDPSWRPMVGTTKVSRTTNF-8-FNKAHEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.01
KX348173	MPKRDPSWRPMVGTTKVSRTTNF-8-FNKAHEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.01
KX348165	MPKRDPSWRPMVGTTKVSRTTNF-8-FNKAHEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.41
KX348177	MPKRDPSWRPMVGTTKVSRTTNF-8-FNKAHEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.40
KX348171	MPKRDPSWRPMVGTTKVSRTTNF-8-FNKAHEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.41
KX348170	MPKRDPSWRPMVGTTKVSRTTNF-8-FNKAHEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.41
KX348174	MPKRDPSWRPMVGTTKVSRTTNF-8-FNKAHEWVNRPM-80-LVRDRRPYGTPMD-147-DSITN*	98.41

CHAPTER 4

ASSESSING VECTOR DIVERSITY TO INFER ABOUT VIRAL DYNAMICS IN A HOST POPULATION: THE ABSENCE OF BEGOMOVIRUS EPIDEMICS IN CASSAVA IN BRAZIL AS A CASE STUDY

Xavier, C.A.D., Nogueira, A.M., Bello, V.H., Alves-Júnior, M., Barbosa, L.F., Beserra-Junior, J.E.A., Boari, A.J., Calegario, R.F., Gorayeb, E.S., Honorato-Júnior, J., Koch, G., Lima, G.S.A., Lopes, C.A., Mello, R.N., Pantoja, K.F.C., Silva, F.N., Ramos-Sobrinho, R., Santana, E.N., Silva, J.W.P., Krause-Sakate, R., F.M. Zerbini. Assessing vector diversity to infer about viral dynamics in a host population: the absence of begomovirus epidemics in cassava in Brazil as a case study. Annals of Applied Biology, in preparation.

Assessing vector diversity to infer about viral dynamics in a host population: the absence of begomovirus epidemics in cassava in Brazil as a case study

César A.D. Xavier[1], Angélica M. Nogueira[1], Vinícius H. Bello[2], Miguel Alves-Júnior[3], Leonardo F. Barbosa[4], José E.A. Beserra-Junior[5], Alessandra J. Boari[6], Renata F. Calegario[7], Eduardo S. Gorayeb[8], Jaime Honorato-Júnior[9], Gabriel Koch[7], Gaus S.A. Lima[10], Cristian A. Lopes[4], Raquel N. Mello[11], Késsia F. C. Pantoja[6], Fabio N. Silva[8], Roberto Ramos-Sobrinho[10#], Enilton N. Santana[12], José W.P. Silva[13], Renate Krause-Sakate[2]*, F.M. Zerbini[1]*

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

[2] Dep. de Defesa Fitossanitária, Universidade Estadual Paulista, Botucatu, SP 18610-034, Brazil

[3] Lab. de Fitopatologia Agrícola e Florestal, Faculdade de Engenharia Agronômica, Universidade Federal do Pará, Altamira, PA 68372-040, Brazil

[4] Instituto Federal do Sudeste de Minas Gerais, Campus Rio Pomba, Rio Pomba, MG 36180-000, Brazil

[5] Dep. de Fitotecnia, Universidade Federal do Piauí, Teresina, PI 64049-518, Brazil

[6] Embrapa Amazônia Oriental, Belém, PA 66095-903, Brazil

[7] Dep. de Fitotecnia e Fitossanidade, Universidade Federal do Paraná, Curitiba, PR 80035-250, Brazil

[8] Centro de Ciências Agroveterinárias, Universidade do Estado de Santa Catarina, Lages, SC 88520-000, Brazil

[9] Centro Multidisciplinar do Campus de Barra, Universidade Federal do Oeste da Bahia, Barra, BA 47100-000, Brazil

[10] Centro de Ciências Agrárias/Fitossanidade, Universidade Federal de Alagoas, Rio Largo, AL 57100-000, Brazil

[11] Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO 75375-000, Brazil

[12] Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural, CRDR-Linhares, Linhares, ES 29900-970, Brasil

[13] Lab. de Entomologia Agrícola, Faculdade de Engenharia Florestal, Universidade Federal do Pará, Altamira, PA 68372-040, Brazil

[#] Present address: School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA

* Corresponding author: Renate Krause-Sakate

Phone: (55-14) 3880-7487

E-mail: renate.krause@unesp.br

* Corresponding author: F. Murilo Zerbini

Phone: (+55-31) 3612-2423

E-mail: zerbini@ufv.br

Abstract

Virus ecology is strongly dependent on that of its vector. The necessity of a competent vector for transmission is a primary ecological factor driving the host range expansion of a virus, with vectors playing an essential role in promoting disease emergence. Cassava begomoviruses severely constrain cassava production in Africa. Curiously, begomoviruses have never been reported in cassava fields in South America, its center of origin. It has been hypothesized that the absence of a competent vector that efficiently colonizes cassava is the reason why begomoviruses have not emerged in South America. To test this hypothesis, we performed a country-wide whitefly diversity study in cassava in Brazil. Adults and/or nymphs of whiteflies were collected from sixty-six cassava fields across twelve states representing the main agroecological zones of the country. A total of 1,385 individuals were genotyped based on mitochondrial cytochrome oxidase I (mtCOI) sequences. A high degree of species richness was observed, with five previously described species and two putative new ones. The most prevalent species were Tetraleurodes acaciae and Bemisia tuberculata, representing over 75% of the analyzed individuals. Although we detected, for the first time, the presence of whitefly species of the Bemisia tabaci complex colonizing cassava in Brazil, they were not prevalent. The species composition varied across regions and cropping systems, with the Northeast region showing the highest diversity. These results expand our knowledge of whitefly diversity in cassava, and support the hypothesis that begomovirus epidemics have not occurred in cassava in Brazil due to the absence of competent vector populations. However, they indicate an ongoing adaptation process of B. tabaci Middle East-Asia Minor 1 to cassava, increasing the likelihood of begomovirus emergence in this crop in the future.

Introduction

Cassava (Manihot esculenta Crantz) is a perennial shrub of the Euphorbiaceae family with great economic and social importance, especially in Africa, Asia and Latin America. Currently, cassava is the third most important source of calories after rice and corn, and is a staple food for more than one billion of people living mainly in developing countries (Montagnac et al., 2009). Although the botanical and geographical origin of M. esculenta is still debated, studies based on genetic markers and archaeological evidence suggest that domesticated cassava originated from the wild relative progenitor M. esculenta ssp. flabellifolia in the Amazon basin, with the domestication center located at the southern border of the Amazon in Brazil (Olsen and Schaal, 1999; Leotard et al., 2009; Clement et al., 2016; Watling et al., 2018). After its introduction in west Africa by Portuguese traders during the 16th century, cassava quickly disseminated throughout tropical Africa and Asia (Carter et al., 1997). Currently, the African continent is the world's biggest cassava producer, followed by Asia and South America (FAO, 2017). Due to its high resilience to adverse environmental conditions, especially drought, high yield per unit of land and low level of management and inputs required during its life cycle, cassava is a suitable crop for poor and small farmers, partially ensuring food security in many African countries (Alves and Setter, 2004; El-Sharkawy, 2004; Gleadow et al., 2016).

Cassava mosaic disease (CMD) is considered the most significant constraint to cassava production in Africa (Rey and Vanderschuren, 2017; Jacobson et al., 2018). CMD is caused by viruses of the genus Begomovirus (family Geminiviridae), which are transmitted in a circulative manner by whiteflies of the Bemisia tabaci cryptic species complex (Zerbini et al., 2017). To date, nine cassava mosaic begomoviruses (CMBs) have been reported in association with CMD, seven of them in Africa and two in the Indian subcontinent (Patil and Fauquet, 2009; Legg et al., 2015). The presence in Africa of B. tabaci species adapted to cassava has been hypothesized as the reason of the occurrence of CMD in that continent (Legg et al., 2014; Wosula et al., 2017; Legg et al., 2015).

Whiteflies comprise a diverse group of insects in the family Aleyrodidae, with more than 1,500 species assigned to 126 genera of which over 20 species have been reported colonizing cassava worldwide (Vasquez-Ordonez et al., 2015). In addition to the direct damage due feeding in the plant phloem, whiteflies cause indirect damage by deposition of honeydew, favoring the growth of sooty mold fungi on the leaf surface, and mainly by transmission of a broad range of viruses to several crops, including cassava (Navas-Castillo et al., 2011).

Currently, species included in the genera Aleurodicus, Bemisia and Trialeurodes have been shown to constitute effective vectors of plant viruses included in five families (Navas-Castillo et al., 2011, Njoroge et al., 2017, Maruthi et al., 2017). While Aleurodicus dispersus has been shown to transmit only one virus in the Ipomovirus genus, and T. vaporariorum and T. abutilonea to transmit a few viruses included in the Crinivirus and Torradovirus genera, the B. tabaci complex is one of the most important vectors of plant viruses, transmitting over 430 viruses, the majority included in the genus Begomovirus (Navas-Castillo et al., 2011; Zerbini et al., 2017). Moreover, other viruses classified in the genera Carlavirus, Crinivirus, Ipomovirus, Torradovirus and a recently reported new virus included in the genus Polerovirus are also efficiently vectored by species of the B. tabaci complex (Navas-Castillo et al., 2011; Whitfield et al., 2015; Gilbertson et al., 2015; Ghosh et al., 2019).

Over the last decade, advances in the use of molecular markers lead to a reappraisal of the taxonomic status of B. tabaci (Dinsdale et al., 2010; De Barro et al., 2011). Based on molecular phylogeny of the mitochondrial cytochrome oxidase I (mtCOI) gene, it has been proposed that B. tabaci consists of a complex of more than 40 cryptic (morphologically indistinguishable) species (Dinsdale et al., 2010; De Barro et al., 2011; Lee et al., 2013; Mugerwa et al., 2018; Vyskočilová et al., 2018). Partial or complete reproductive isolation and biological and ecological differences among distinct species within the complex support the proposed classification (Qin et al., 2016; Gilbertson et al., 2015; De Marchi et al., 2017; Malka et al., 2018). The global dissemination of polyphagous and invasive species, such as B. tabaci Middle East-Asia Minor 1 (BtMEAM1) and B. tabaci Mediterranean (BtMED), have caused major changes in the epidemiology of crop-infecting begomoviruses such as Tomato yellow leaf curl virus (TYLCV), currently present in all the main tomato producing areas of the world (Lefeuvre et al., 2010; Mabvakure et al., 2016; Pan et al., 2012). In addition, the dissemination of polyphagous whiteflies has favored the transfer of indigenous begomoviruses from wild reservoir hosts to cultivated plants, as occurred in tomato crops in Brazil after the introduction of BtMEAM1 in the mid-1990's (Ribeiro et al., 1998; Rocha et al., 2013).

Specific associations between endemic, indigenous populations of B. tabaci and begomoviruses have also led to the emergence of severe epidemic in crops (Fauquet and Fargette, 1990; Pan et al., 2018). In Africa, the emergence of CMD seems to have been the result of the transfer of indigenous begomoviruses from wild reservoir hosts to cassava, probably mediated by endemic populations of B. tabaci that have adapted to feed in cassava since its introduction from South America (Fauquet and Fargette, 1990; Legg and Fauquet, 2004). Although wild reservoir hosts and a possible ancestral progenitor of current

begomoviruses causing CMD have not been found, the absence of cassava-infecting begomoviruses in the Americas supports an African origin for those viruses, and the presence of cassava-adapted B. tabaci species being restricted to Africa reinforces that hypothesis. The high species diversity and high level of molecular variation observed in viral populations causing CMD strongly suggests Africa as a diversification center for CMBs, with distinct CMBs recurrently emerging and evolving for a long time (De Bruyn et al., 2012; Ndunguru et al., 2005; De Bruyn et al., 2016; Tiendrebeogo et al., 2012).

Even if CMD in Africa is caused by indigenous viruses, the fact that cassava in South America has not been affected by begomoviruses is puzzling (Carabali et al., 2008; Carabali et al., 2010; Patil and Fauquet, 2009). Although it could be suggested that there are no begomoviruses capable of infecting cassava in the Americas, the high diversity of begomoviruses reported in a broad range of cultivated and non-cultivated plants in several botanical families, including the Euphorbiaceae, make this highly unlikely (Fernandes et al., 2008; Fernandes et al., 2011; Albuquerque et al., 2012a; Albuquerque et al., 2012b; Macedo et al., 2017; Rocha et al., 2013; Castillo-Urquiza et al., 2008; Paz-Carrasco et al., 2014; Rodríguez-Negrete et al., 2019; Mar et al., 2017b). Carabali et al. (2005) suggested that the absence of cassava-infecting begomoviruses in the Americas would be due to lack of competent B. tabaci species that efficiently colonize cassava (Fauquet and Fargette, 1990; Legg and Fauquet, 2004). In Colombia, the inability of B. tabaci MEAM1 to colonize M. esculenta efficiently has been demonstrated under experimental conditions, reinforcing the above hypothesis (Carabali et al., 2010). In addition, a recent study surveying endosymbiont diversity in cassava whiteflies from Colombia failed to detect any whitefly species of the B. tabaci complex (Gómez-Díaz et al., 2019). This study revealed that the most prevalent species in cassava was Trialeurodes variabilis, followed by Aleurotrachelus socialis and Bemisia tuberculata (Gómez-Díaz et al., 2019).

Although whitefly diversity in Brazil has been surveyed extensively in recent years (Rocha et al., 2011; Marubayashi et al., 2013; Marubayashi et al., 2014; Barbosa et al., 2015; Moraes et al., 2017; Moraes et al., 2018), no study has been carried out specifically to explore the composition of whitefly populations colonizing cassava. Those studies carried out in other crops demonstrated that B. tabaci MEAM1 is the predominant species across Brazil in crops such as common bean, cotton, pepper, tomato and soybean. Furthermore, B. tabaci MED, which was recently introduced in Brazil, has rapidly spread and currently is present in five states from the South and Southeast regions (Barbosa et al., 2015; Moraes et al., 2017; Moraes et al., 2018). A small number of whitefly samples from cassava were analyzed in those studies, with

B. tuberculata and Tetraleurodes acaciae prevalent and detected exclusively in cassava (Marubayashi et al., 2014; Moraes et al., 2018). A large survey addressing whitefly diversity in cassava in its domestication center is needed, as it could provide clues to understand the absence of a CMD-like disease in the Americas. Moreover, this knowledge would be useful to anticipate the potential of emergence of begomoviruses in the crop and to help anticipate a management strategy.

Given this context, the objective of this work was to evaluate whitefly diversity in cassava across Brazil. Analysis of mtCOI sequences from 1,385 individuals (nymphs and adults) collected in 12 states demonstrated that the most prevalent species were T. acaciae and B. tuberculata, representing over 75% of the analyzed individuals. Although we detected, for the first time, the presence of whiteflies of the B. tabaci complex colonizing cassava, they were not prevalent, contrary to has been previously reported for other crops such as common bean and tomato. The possible implications of these findings are discussed considering the absence of CMD and the potential for its emergence in cassava fields in Brazil.

Methods

Whitefly and cassava samples

Whiteflies were collected exclusively from cassava (M. esculenta) plants across 12 Brazilian states representative of the five macroregions (North, Northeast, Midwest, Southeast and South; Figure 1) between March 2016 and February 2019 (Table 1). To gather evidence of whether a given population was actually colonizing cassava, adults and nymphs from the same field were collected whenever possible (Table 1). Samples were obtained from commercial and non-commercial (subsistence) crops as indicated in Table 1. Whitefly adults were sampled using a hand-held aspirator and nymphs were collected with the aid of a needle. Insects were preserved in 95% ethanol and stored at −20°C until being used for molecular identification of the species.

To verify the presence of begomoviruses infecting cassava, foliar samples were also collected at some sampled sites. The samples were collected randomly regardless of the presence of virus-like symptoms. The leaves were press-dried and stored at room temperature as herbarium-like samples until being used for DNA extraction.

Whitefly species identification

Whitefly species were identified using a PCR-RFLP analysis of the partial mtCOI fragment followed by sequencing, as previously described (Moraes et al., 2018). Overall, between nine to 31 individuals were analyzed from each sampled site. When enough adults and nymphs were collected at a given site, ten individuals from each stage were analyzed, and when only one stage was obtained, 20 individuals were tested (Table 1). When variation in the RFLP pattern was observed in the first screening, suggesting that more than one species could be present in that site, approximately five additional individuals for each stage were analyzed according to sample availability.

Total DNA was extracted from single individual whiteflies following a Chelex protocol (Moraes et al., 2018)with slight modifications. Briefly, adults or nymphs were ground in 30 µl of Chelex buffer (5% Chelex in 1x Tris-EDTA) using a toothpick in a 200 µl tube. Samples were vortexed for 30 seconds and incubated at 99°C for 8 min in a PTC-100 thermocycler (MJ Research). Next, the tubes were centrifuged at 14,000 g for 5 min and 20 µl of the supernatant was collected and transferred to a clean tube. One microliter of the supernatant was used as a template for PCR amplification of a 800 bp fragment of the mtCOI gene using primers C1-J-2195 and L2-N-3014 (Simon et al., 1994; Frohlich et al., 1999). PCR was performed in a final volume of 25 µl using GoTaq Colorless Master Mix (Promega), following the manufacturer's instructions. The PCR cycles consisted of an initial denaturing step at 95°C for 5 min, followed by 35 cycles at 95°C for 30 sec, 42°C for 45 sec and 72°C for 1 min, with a final extension at 72°C for 10 min. Amplified products were visualized in 0.8% agarose gels stained with ethidium bromide and directly used for RFLP analysis.

For RFLP analysis, a 5 µl aliquote of the PCR product were digested with 0.1 unit of TaqI (Promega) in a final volume of 20 µl at 65°C for 2 hours and visualized in 1.2% agarose gels stained with ethidium bromide. To verify whether the predicted mtCOI restriction pattern corresponded to a given species according to in silico prediction, a subset of PCR products from adults and nymphs representative of distinct patterns from different sampled sites were selected and sequenced. PCR products were purified and sequenced commercially (Macrogen Inc.) in both directions using primers C1-J-2195 and L2-N-3014. For samples MG6 and MT1, an additional digestion with EcoRI and HhaI, respectively, were performed.

For a small subset of samples that failed to yield a PCR product using primers C1-J-2195 and L2-N-3014, a second screen using a recently described primer set with improved specificity for species of the B. tabaci complex and B. afer (2195Bt and C012/Bt-sh2)

(Mugerwa et al., 2018). Specific primers for T. vaporariorum (TvapF and Wfrev) were also tested (Scott et al., 2007).

Samples collected at sites # 8, 10, 11, 12, 44, 54, 55, 56, 57, 58, 59, 60, 61, 62, 64, 65 were processed exactly as described in Moraes et al. (2018).

Sequence comparisons and phylogenetic analysis

Nucleotide sequences were first checked for quality and assembled using Geneious v. 8.1 (Kearse et al., 2012) mtCOI sequences were initially analyzed with the BLASTn algorithm (Altschul et al., 1990) to determine the whitefly species with which they shared greatest similarity. Pairwise comparisons between all mtCOI sequences obtained here and those with higher similarities (as determined by the BLASTn search) were performed with the program SDT v. 1.2 (Muhire et al., 2014) using the MUSCLE alignment option (Edgar, 2004).

For phylogenetic analyses, the final dataset was composed of 142 sequences, 95 obtained in this work and 47 sequences representative of species in the Aleyrodidae family. Sequences were retrieved from GenBank and from the updated mtCOI reference dataset for species of the Bemisia tabaci complex (Boykin et al., 2017). Multiple sequence alignments were prepared using the MUSCLE option in MEGA7 (Kumar et al., 2016). Alignments were manually checked and adjusted when necessary. Phylogenetic trees were constructed using Bayesian inference performed with MrBayes v. 3.0b4 (Ronquist and Huelsenbeck, 2003). The program MrModeltest v. 2.2 (Nylander, 2004) was used to select the nucleotide substitution model with the best fit in the Akaike Information Criterion (AIC). The analyses were carried out running 50,000,000 generations with sampling at every 100 generations and a burn-in of 25%. Trees were visualized and edited using FigTree (tree.bio.ed.ac.uk/software/figtree/) and CorelDRAW X5.

Virus detection in foliar samples

Total DNA was extracted as described (Doyle and Doyle, 1987) and used as a template for PCR using the DNA-A universal primer pair PAL1v1978 and PAR1c496 (Rojas et al., 1993). PCR was performed in a final volume of 25 µl using Taq DNA Polymerase (Invitrogen), following the manufacturer's instructions. The PCR cycles consisted of an initial denaturing step at 95°C for 5 min, followed by 35 cycles at 95°C for 1 min, 52°C for 1 min and 72°C for 1 min, with a final extension at 72°C for 10 min. PCR products were visualized in 0.8% agarose gels stained with ethidium bromide.

Diversity index and statical analysis

Simpson's index of diversity (1-D) was calculated to verify if there was any difference in whitefly diversity across macroregions. This index represents the probability that two randomly chosen individuals in a given sampled site will belong to distinct species (Morris et al., 2014). Simpson's index was chosen because it is easily interpreted, as its value increases with increasing diversity and assigns more weight to more abundant species in a sample. We assume that species colonizing cassava will be in abundance, whereas rare species that briefly visit the plant without colonizing it will be underrepresented. Simpson's index was calculated for each sampled site separately and then pooled according to macroregions. To access the statistical significance of the differences in diversity among regions, the non-parametric Kruskal–Wallis test followed by post hoc multiple comparison test using Fisher's least significant difference, calculated using the function kruskal implemented in the Agricolae package in R software (R Development Core Team, 2007).

Results

High whitefly species richness in cassava in Brazil

To verify the composition of whitefly communities colonizing cassava in Brazil, sampling was performed across the whole country including the main agroecological zones. A total of 66 sites from 12 states were sampled (Figure 1; Table 1). Out of 1,385 individuals submitted to PCR-RFLP analysis, 58 adults and 37 nymphs from different locations and representing distinct restriction patterns were sequenced. The combination of PCR-RFLP followed by sequencing showed reliability and consistence for species identification without misidentification due to incongruence between the two methods.

Based on pairwise comparisons and molecular phylogeny of the mtCOI gene we identified the presence of at least seven species comprising the whitefly community in cassava (Figure 2). Among them, T. acaciae and B. tuberculata, both previously reported in this crop, were the most prevalent. In addition, based on the criterion of 3.5% divergence to differentiate species within the B. tabaci complex, three B. tabaci species were identified for the first time in cassava fields in Brazil (Figure 2). The BtMEAM1 was predominant among them, representing 18% of the total individuals analyzed, followed by BtMED (1.6%) and BtNW (0.21%).

Furthermore, two putative new species were identified (Figure 2), provisionally named WtNEW1 and WtNEW2. The WtNEW1 mtCOI sequence showed highest identity (81.13%)

and clustered close to the T. acaciae clade, comprised of individuals reported here and three other previously reported sequences from cassava in Brazil (Figure 2). For WtNEW2, two mtCOI sequences obtained from one adult and one nymph shared highest identity (81%) with Bemisia afer (AY124567; Figure 2) and clustered as a basal sister clade with genus Bemisia. Although whitefly taxonomy is predominantly based on puparial characters (Hodges and Evans, 2005) and there is no taxonomic criterion established based in mtCOI sequences for most of the groups, as has been proposed for the B. tabaci complex, the level of divergence between the two proposed new species with the closest species is similar to the level of divergence observed between species already described within the Aleyrodidae, as demonstrated in pairwise comparisons and phylogenetic analysis (Figure 2). Together, these results indicate the existence of a high whitefly species richness in cassava in Brazil.

Both the prevalence and the capacity to colonize cassava differ among species

Nymphs were collected for samples identified as T. acaciae, B. tuberculata, BtMEAM1 and the two new putative species (Figure 3A), suggesting that these species may colonize cassava. Nymphs were not obtained at the two sites where BtMED was prevalent (SP1 and SP12). Although it could be suggested that this species has the potential to colonize cassava due to the high incidence of adults at these two sites, the lack of nymphs suggests otherwise. Likewise, at the sites PR4 and MT6, BtMEAM1 predominated among adults, but 100% of the nymphs were B. tuberculata. Further sampling in those sites or free-choice experiments are necessary to confirm the potential of BtMED and MEAM1 to colonize cassava. Considering the whole sampling, we detected only three adults of BtNW, suggesting an inability of this specie to colonize cassava.

To verify if prevalence differs across distinct stage among species, the data were separated according to stage and the proportions of individuals were compared for the three most abundant species (Figure 3B, C). Considering the entire data set, T. acaciae was the prevalent species, followed by B. tuberculata and BtMEAM1 (x_2^2=152.63, P<2.2x10^{-16}). The same was true according to stage, either adults (x_2^2=28.612, P<6.124x10^{-07}) or nymphs (x_2^2=169.44, P<2.2x10^{-16}; Figure 3B). However, caution is needed to interpret these data because only adults were sampled at some sites where BtMEAM1 and B. tuberculata were prevalent (Figure 3A), which could bias the analysis, causing an underestimation of the number of nymphs for those species. Therefore, we also analyzed the data considering only those sites where both adults and nymphs were obtained. Again, T. acaciae was the predominant species followed by B. tuberculata and BtMEAM1 considering either the entire data set (x_2^2=258.61,

P<2.2x10^{-16}) or only nymphs (x_2^2=164.47, P<2.2x10^{-16}). When only adults were considered, T. acaciae was still predominant (x_2^2=113.52, P<2.2x10^{-16}) but no difference between B. tuberculata and BtMEAM1 was observed (x_1^2=0.505, P=0.477; Figure 3C). Moreover, it could be argued that samples from Minas Gerais (MG) were overrepresented in our sampling (Figure 1C), which could also bias the results presented above due to the predominance of T. acaciae in this state (Figure 3A). To test this possibility, we analyzed the data excluding the samples from MG. In this case, when both stages were considered, B. tuberculata was predominant (x_2^2=62.097, P=3.28x10^{-14}) but no difference between T. acaciae and BtMEAM1 was observed (x_1^2=1.9149, P=0.1664). When only one stage was considered, B. tuberculata was predominant followed by BtMEAM1 and T. acaciae (adults: x_2^2=43.943, P=2.87x10^{-10}; nymphs, x_2^2=84.199, P<2.2x10^{-16}). Together, these results indicate that the potential to colonize cassava differs among species, either due to lower preference for the plant or to differences in the competitive ability among species during cassava colonization. In addition, they reinforce the low efficiency of BtMEAM1 to colonize cassava.

Competitive interference does not explain the differences in prevalence

Interestingly, at least two species were detected co-occurring at 51% of the sites (Figure 3A). To verify the possibility of competition among T. acaciae, B. tuberculata and BtMEAM1 to explain the observed differences in prevalence (instead of differences in host preference), the competitive capacity of these three species was inferred based on the analysis of predominance at the sites where they occurred together. Initially, we verified if there were any differences in incidence, defined here as the number of sampled sites where at least one individual belonging to one of the three species was detected (Figure 4A). The results demonstrate that there were no differences in incidence among them (x_2^2=1.247, P=0.537; Figure 4A). In addition, no differences were observed when the proportion of sites where whitefly species occurred alone or in different combinations was compared (x_6^2=3.258, P=0.7758; Figure 4B). However, when we compared the occurrence between BtMEAM1 and non-B. tabaci species at the sites where they occur alone, the number of sites with non-B. tabaci species was higher (x_1^2 =6.533, P=0.01059; Figure 4B). Thus, the competitive capacity was inferred based on the proportion of individuals from each species at the fields where these species were detected co-occurring in different combinations (Figure 4C). Interestingly, at the sites where B tMEAM1 and B. tuberculata were sampled together, B. tuberculata predominated over BtMEAM1, suggesting higher competitive potential (Figure 4C). However, for all other species combinations, no

evidence of differences in competitive capacity were observed (Figure 4C). Together, these results suggest that, rather than incidence or competition, host preference by BtMEAM1 explains its non-prevalence compared to T. acaciae and B. tuberculata, resulting in low colonization rate as indicated by the low number of BtMEAM1 nymphs detected in cassava (Figure 3).

Composition and species diversity of whiteflies differ among Brazilian regions

The predominance of species composing the whitefly community across macroregions varied considerably (Figure 5A). While T. acaciae predominated in the North, Southeast and Northeast, it was not detected in the Midwest. In addition, B. tuberculata was detected in all regions, and was prevalent in the South and Midwest. BtMEAM1, although not prevalent in any of the regions, was also detected in all regions. Although the number of species detected was higher in the Southeast, where six species out seven were detected, whitefly diversity was significantly higher in the Northeast according to Simpson's index of diversity (Figure 5B), with no differences among the other four regions.

Discussion

Vectors play an essential role during the life cycle of plant viruses, directly affecting their ecology and evolution (Gallet et al., 2018; Sacristan et al., 2003; Gutierrez and Zwart, 2018). Usually, a group of plant viruses establishes a very specific interaction with only one or a few related species of vectors, making virus ecology strongly dependent on that of its vector (Gallet et al., 2018). It has been suggested that the natural host range of a virus is dependent on its vector's host range, as the most viruses have greater specificity for the vector than for the plant host (Elena et al., 2014; Dietzgen et al., 2016). Indeed, the existence of a vector which is not only for transmission but is also able to colonize potential reservoir and recipient new hosts is a primary ecological factor driving host range expansion of viruses. Thus, vectors play an essential role during viral disease emergence and epidemics (Navas-Castillo et al., 2011; Elena et al., 2014; Fereres, 2015; Gilbertson et al., 2015).

Our country-wide survey of whitefly populations associated with cassava in Brazil uncovered a high degree of species diversity and indicated that T. acaciae and B. tuberculata are the prevalent species across the country. Non-B. tabaci species, including B. tuberculata, have been shown to be prevalent also in Colombia (Gómez-Díaz et al., 2019). In contrast, in Africa, endemic species of the B. tabaci complex are prevalent in cassava (Legg et al., 2014;

Ghosh et al., 2015; Tocko-Marabena et al., 2017). Previous studies surveying whitefly diversity in South American countries failed to detect T. acaciae and B. tuberculata in hosts other than cassava, indicating a very narrow host range which may in fact be restricted to cassava (Marubayashi et al., 2014; Moraes et al., 2018; Alemandri et al., 2015).

In contrast with previous studies, species of the B. tabaci complex were reported for the first time colonizing cassava in Brazil. BtMEAM1 was the third most prevalent species, representing 18% of the genotyped individuals, and with similar incidence to T. acaciae and B. tuberculata. The failure of previous studies to detect BtMEAM1 in cassava may have been due to the small number of samples analyzed. The wide distribution and prevalence of BtMEAM1 in several crops across the main agroecological zones in Brazil has been well established (Moraes et al., 2018), indicating a high adaptation to distinct climate conditions and hosts. It will be important to monitor BtMEAM1 populations in cassava over the next years, to assess its possible adaptation to this host. The fact that we collected BtMEAM1 nymphs at several locations suggests that this process may already be under way. We also detected BtMED, a worrying result given the recent introduction of this species in the Brazil and its potential to displace other species, including BtMEAM1 (Liu et al., 2012; Sun et al., 2013; Watanabe et al., 2019). BtMED has disseminated quickly in the country, mainly in association with ornamental plants in greenhouses (Moraes et al., 2018). Our work is the first report of its detection in in cassava. However, we cannot infer its potential to effectively colonize this host since only adults were collected. The third species detected is the indigenous BtNW. Although BtMEAM1 partially displaced BtNW in brazil, this species can still be sporadically detected, mostly in association with non-cultivated hosts (Marubayashi et al., 2013; Marubayashi et al., 2014; Moraes et al., 2018). It has been recurrently detected in Euphorbia heterophylla, suggesting a potential to colonize other species in the family Euphorbiaceae. However, the very low frequency with which it was detected, and the absence of nymphs indicate that BtNW is poorly adapted to cassava.

The identification of two putative new species highlights the remarkable diversity of whiteflies, as new species have not been detected from other crops in Brazil. Interestingly, one of the new species was collected in the state of Mato Grosso, which corresponds to the region considered to be the domestication center of cassava(Olsen and Schaal, 1999; Leotard et al., 2009; Clement et al., 2016; Watling et al., 2018). Further studies are needed to explore the high degree of plant biodiversity in this region (Nassar, 2001; Olsen, 2004), which might reveal a similar diversity of whiteflies which may be specifically adaptated to those wild plant species due to long-term co-evolution. The close phylogenetic relationship of the new species with non-

B. tabaci whiteflies suggests that they are not virus vectors, but this should be demonstrated experimentally.

The whitefly species richness in cassava is just beginning to be assessed and, in all likelihood, may be greater than reported here. Based on morphological characters, Alonso et al. (2012) reported the presence of Aleurothrixus aepim and Trialeurodes manihoti colonizing cassava in the state of Rio de Janeiro. Although we did not analyze samples from that region, the failure to detect these species in other states suggests a restricted occurrence. Additional studies are necessary to assess the actual whitefly diversity in this crop in Brazil.

Host suitability has been shown to be an important factor influencing the competitive capacity among species of the B. tabaci complex (Luan et al., 2012; Sun et al., 2013). Sun et al. (2013) demonstrated that displacement capacity between two invasive B. tabaci species was dependent on the host suitability. While BtMEAM1 displaced BtMED on cabbage and tomato, BtMED displaced BtMEAM1 on pepper. Interestingly, Luan et al. (2012) demonstrated that even in a host plant poorly suitable for B. tabaci MEAM1, it was able to displace an indigenous species challenger. These authors demonstrated that even though host suitability may affect the speed of displacement, it may not affect the direction, as B. tabaci MEAM1 always won the challenge (Luan et al., 2012). Although there is no experimental study addressing the ability of BtMEAM1 to outcompete species which are adaptated to cassava, a lower competitive ability could, at least in part, explain its low prevalence observed here and elsewhere (Tocko-Marabena et al., 2017; Tajebe et al., 2015; Ghosh et al., 2015). Interestingly, two or more species occurring sympatrically were detected in 51% of the fields analyzed in our study. In sites where BtMEAM1 and B. tuberculata co-occurred, B. tuberculata predominated, indicating a higher competitive capacity. Nonetheless, in all other combinations of co-occurring species, no differences in prevalence were observed. Thus, competitive capacity by itself is unlikely to explain the low prevalence of BtMEAM1, or the differences observed between T. acaciae and B. tuberculata.

Host adaptation may be a more important component affecting the low predominance of BtMEAM1 in cassava, as previously suggested (Carabali et al., 2005). The inability of BtMEAM1 and BtMED to colonize domesticated cassava efficiently has been demonstrated under experimental conditions (Carabali et al., 2010; Carabali et al., 2008; Vyskočilová et al., 2019; Milenovic et al., 2019). Carabali et al. (2008), evaluating the colonization potential of BtMEAM1 in three commercial cassava genotypes, demonstrated that only in one of them did BtMEAM1 complete its development cycle from eggs to adult, and even then, at very low rates (0.003%). Using an electrical penetration graph assay, Milenovic et al. (2019) demonstrated

the inability of BtMED to feed in cassava plants. Adults of this species spent a very short time ingesting cassava phloem sap compared to sap from a suitable host, suggesting that they would die by starvation in the field. Furthermore, the low efficiency of whiteflies of the BtMED mitochondrial subgroups Q1 and Q2 in using cassava as a host has also been demonstrated (Vyskočilová et al., 2019). Oviposition and adult survival rates were very low, and development from eggs to adults was not observed. Although these studies were conducted under experimental conditions, the low predominance of BtMEAM1 and BtMED shown here and in other field surveys in Africa (Tocko-Marabena et al., 2017; Tajebe et al., 2015; Ghosh et al., 2015) strongly indicates a low adaptation of these species to cassava.

Nevertheless, our results indicate an ongoing adaptation process of BtMEAM1 to cassava, with the detection of nymphs and adults in the same field. Interestingly, Carabali et al. (2005) demonstrated a gradual increase in the rate of reproduction and development of BtMEAM1 after successive passages on plants phylogenetically related to the genus Manihot (Euphorbia pulcherrima and Jatropha gossypiifolia), indicating the potential of this whitefly species to become adapted to cassava through intermediate hosts. Furthermore, successful reproductive capacity in the wild relative M. esculenta ssp. flabellifolia, indicates that this plant may constitute an intermediate host leading to adaptation (Carabali et al., 2010). This plant has been reported to be widely spread in the Amazon basin and the Midwest region of Brazil (Olsen, 2004). Interestingly, our data showed the higher prevalence of BtMEAM1 to be in the Midwest. Although we cannot establish a cause and effect relationship, these results suggest that M. esculenta ssp. flabellifolia could be acting as an intermediate host mediating adaptation.

In Brazil, cassava is predominantly grown as a subsistence crop, usually side by side with other vegetables and with a high incidence of weeds. Growing cassava in a heterogenous environment, especially in the presence of related plants, may increase the adaptation potential of BtMEAM1 and other species of the complex such as BtMED, which we also detected in the open field. A high diversity of plants in cassava fields may permit an overlapping of ecological niches for distinct whitefly species, which under enough selection pressure may gradually adapt to new hosts. The sympatric occurrence of T. acaciae, B. tuberculata and (to a lesser extent) BtMEAM1, supports the role of botanical heterogeneity in shaping the composition of whitefly populations associated with cassava. A similar pattern was observed in Colombia, with 66% of the surveyed sites showing at least two species occurring sympatrically (Gómez-Díaz et al., 2019). Moreover, a predominance of one species in a given developmental stage and a different one in another stage (eg, nymphs vs adults) at the same site suggests that other hosts may sustain reproduction and development, with adults migrating to cassava. The detection of only

BtMEAM1 adults at some sites reinforces the possibility of migration from other hosts. Although Moraes et al. (2018) did not indicate the cropping system in the fields surveyed in their study, only one whitefly species was detected at most of the sampled sites, in association with crops such as common bean, cotton, soybean and tomato. These crops are usually cultivated in monoculture. Thus, management strategies aiming at eliminating possible intermediate hosts could help to reduce the adaptation potential of whitefly populations.

Euphorbia heterophylla (family Euphorbiaceae) is an invasive weed widely spread across Brazil and associated with several crops (Wilson, 1981; Mar et al., 2017b). The presence of E. heterophylla plants in association with cassava (Figure 1A) and the fact that it was the most suitable host for BtMEAM1 in Brazil out seven tested (Sottoriva et al., 2014) shows its potential to act as an intermediate host mediating BtMEAM1 adaptation. E. heterophylla has been frequently associated with the begomovirus Euphorbia yellow mosaic virus (EuYMV) (Mar et al., 2017b). Interestingly, Barreto et al. (2013) demonstrated that this plant is also a host of Tomato severe rugose virus (ToSRV), which even at a very low titre was transmitted to tomato plants, demonstrating the potential of E. heterophylla to act as a reservoir host. Surprisingly, considering that E. heterophylla and tomato belong to distinct botanical families, EuYMV is able to infect tomato (Barreto et al., 2013). The closer botanical relationship between E. heterophylla and cassava may indicate a higher potential of EuYMV to infect cassava. The presence of E. heterophylla-infected in cassava fields, as observed in this study (Figure 1A), its suitability as a host for BtMEAM1, and the high efficiency of EuYMV transmission by BtMEAM1 (Mar et al., 2017a), suggest that it may also be a potential reservoir of begomoviruses to cassava. Experimental studies may help to determine the potential for this occur.

The emergence of begomoviruses in tomato crops in Brazil followed the introduction of BtMEAM1 (Ribeiro et al., 1998; Rocha et al., 2013), demonstrating the role of vector populations in promoting viral host range expansion. Thus, the adaptation of whiteflies to cassava could facilitate the emergence of begomoviruses in this crop. The establishment of management strategies to prevent or at least delaying the adaptation process is therefore necessary. Bemisia tabaci species may disperse across long distances international trade routes (Hadjistylli et al., 2016). Thus, preventing the introduction of cassava-adapted B. tabaci species from Africa should also be a priority.

References

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

Albuquerque L. C., Varsani A., Fernandes F. R., Pinheiro B., Martin D. P., Ferreira P. T. O., Lemos T. O., Inoue-Nagata A. K. (2012b) Further characterization of tomato-infecting begomoviruses in Brazil. Archives of Virology, **157**, 747-752.

Alemandri V., Vaghi Medina C. G., Dumon A. D., Arguello Caro E. B., Mattio M. F., Garcia Medina S., Lopez Lambertini P. M., Truol G. (2015) Three members of the Bemisia tabaci (Hemiptera: Aleyrodidae) cryptic species complex occur sympatrically in Argentine horticultural crops. Journal of Economic Entomology, **108**, 405-413.

Alonso R. d. S., Racca-Filho F., Lima A. F. d. (2012) Occurrence of whiteflies (Hemiptera: Aleyrodidae) on cassava (Manihot esculenta Crantz) crops under field conditions in the state of Rio de Janeiro, Brazil. EntomoBrasilis, **5**, 2.

Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. (1990) Basic local alignment search tool. Journal of Molecular Biology, **215**, 403-410.

Alves A. A. C., Setter T. L. (2004) Response of cassava leaf area expansion to water deficit: cell proliferation, cell expansion and delayed development. Annals of Botany, **94**, 605-613.

Barbosa L. F., Yuki V. A., Marubayashi J. M., De Marchi B. R., Perini F. L., Pavan M. A., Barros D. R., Ghanim M., Moriones E., Navas-Castillo J., Krause-Sakate R. (2015) First report of Bemisia tabaci Mediterranean (Q biotype) species in Brazil. Pest Management Science, **71**, 501-504.

Barreto S. S., Hallwass M., Aquino O. M., Inoue-Nagata A. K. (2013) A study of weeds as potential inoculum sources for a tomato-infecting begomovirus in central Brazil. Phytopathology, **103**, 436-444.

Boykin L. M., Savill A., De Barro P. (2017) Updated mtCOI reference dataset for the Bemisia tabaci species complex. F1000Res, **6**, 1835.

Carabali A., Belloti A. C., Montoya-Lerma J. (2010) Biological parameters of Bemisia tabaci (Gennadius) biotype B (Hemiptera: Aleyrodidae) on Jatropha gossypiifolia, commercial (Manihot esculenta) and wild cassava (Manihot flabellifolia and M. carthaginensis) (Euphorbiaceae). Neotropical Entomology, **39**, 562-567.

Carabali A., Bellotti A. C., Montoya-Lerma J., Cuellar M. E. (2005) Adaptation of Bemisia tabaci biotype B (Gennadius) to cassava, Manihot esculenta (Crantz). Crop Protection, **24**, 643-649.

Carabali A., Montoya-Lerma J., Bellotti A. C. (2008) Development and reproduction of Bemisia tabaci "B" (Hemiptera : Aleyrodidae) on cassava (Manihot esculenta) genotypes. Revista Colombiana De Entomologia, **34**, 28-32.

Carter S. E., Fresco L. O., Jones P. G., Fairbairn J. N. (1997) Introduction and diffusion of cassava in Africa. IITA Research Guide no. 49, Ibadan, Nigeria: International Institute of Tropical Agriculture (IITA).

Castillo-Urquiza G. P., Beserra Jr. J. E. A., Bruckner F. P., Lima A. T. M., Varsani A., Alfenas-Zerbini P., Zerbini F. M. (2008) Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil. Archives of Virology, **153**, 1985-1989.

Clement C. R., Rodrigues D. P., Alves-Pereira A., Mühlen G. S., Cristo-Araújo M. d., Moreira P. A., Lins J., Reis V. M. (2016) Crop domestication in the upper Madeira River basin. Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas, **11**, 193-205.

De Barro P. J., Liu S. S., Boykin L. M., Dinsdale A. B. (2011) Bemisia tabaci: A statement of species status. Annual Review of Entomology, **56**, 1-19.

De Bruyn A., Harimalala M., Zinga I., Mabvakure B. M., Hoareau M., Ravigne V., Walters M., Reynaud B., Varsani A., Harkins G. W., Martin D. P., Lett J. M., Lefeuvre P. (2016) Divergent evolutionary and epidemiological dynamics of cassava mosaic geminiviruses in Madagascar. BMC Evolutionary Biology, **16**, 182.

De Bruyn A., Villemot J., Lefeuvre P., Villar E., Hoareau M., Harimalala M., Abdoul-Karime A. L., Abdou-Chakour C., Reynaud B., Harkins G. W., Varsani A., Martin D. P., Lett J. M. (2012) East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus. BMC Evolutionary Biology, **12**, 228.

De Marchi B. R., Marubayashi J. M., Favara G. M., Yuki V. A., Watanabe L. F. M., Barbosa L. F., Pavan M. A., Krause-Sakate R. (2017) Comparative transmission of five viruses by Bemisia tabaci NW2 and MEAM1. Tropical Plant Pathology, **42**, 495-499.

Dietzgen R. G., Mann K. S., Johnson K. N. (2016) Plant virus-insect vector interactions: current and potential future research directions. Viruses, **8**, 303.

Dinsdale A., Cook L., Riginos C., Buckley Y. M., De Barro P. (2010) Refined global analysis of Bemisia tabaci (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. Annals of the Entomological Society of America, **103**, 196-208.

Doyle J. J., Doyle J. L. (1987) A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochemical Bulletin, **19**, 11-15.

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

El-Sharkawy M. A. (2004) Cassava biology and physiology. Plant Molecular Biology, **56**, 481-501.

Elena S. F., Fraile A., Garcia-Arenal F. (2014) Evolution and emergence of plant viruses. Advances in Virus Research, **88**, 161-191.

Fauquet C., Fargette D. (1990) African cassava mosaic virus: etiology, epidemiology and control. Plant Disease, **74**, 404-411.

Fereres A. (2015) Insect vectors as drivers of plant virus emergence. Current Opinion in Virology, **10**, 42-46.

Fernandes F. R., Albuquerque L. C., Giordano L. B., Boiteux L. S., Ávila A. C., Inoue-Nagata A. K. (2008) Diversity and prevalence of Brazilian bipartite begomovirus species associated to tomatoes. Virus Genes, **36**, 251-258.

Fernandes F. R., Albuquerque L. C., Oliveira C. L., Cruz A. R. R., Rocha W. B., Pereira T. G., Naito F. Y. B., Dias N. D., Nagata T., Faria J. C., Zerbini F. M., Aragão F. J. L., Inoue-Nagata A. K. (2011) Molecular and biological characterization of a new Brazilian begomovirus, euphorbia yellow mosaic virus (EuYMV), infecting Euphorbia heterophylla plants. Archives of Virology, **156**, 2063-2069.

Frohlich D. R., Torres-Jerez I. I., Bedford I. D., Markham P. G., Brown J. K. (1999) A phylogeographical analysis of the Bemisia tabaci species complex based on mitochondrial DNA markers. Molecular Ecology, **8**, 1683-1691.

Gallet R., Michalakis Y., Blanc S. (2018) Vector-transmission of plant viruses and constraints imposed by virus-vector interactions. Current Opinion in Virology, **33**, 144-150.

Ghosh S., Bouvaine S., Maruthi M. N. (2015) Prevalence and genetic diversity of endosymbiotic bacteria infecting cassava whiteflies in Africa. Bmc Microbiology, **15**, 93.

Ghosh S., Kanakala S., Lebedev G., Kontsedalov S., Silverman D., Alon T., Mor N., Sela N., Luria N., Dombrovsky A., Mawassi M., Haviv S., Czosnek H., Ghanim M. (2019) Transmission of

a new polerovirus infecting pepper by the whitefly Bemisia tabaci. Journal of Virology, **15**, 00488-00419.

Gilbertson R. L., Batuman O., Webster C. G., Adkins S. (2015) Role of the insect supervectors Bemisia tabaci and Frankliniella occidentalis in the emergence and global spread of plant viruses. Annual Review of Virology, **2**, 67-93.

Gleadow R., Pegg A., Blomstedt C. K. (2016) Resilience of cassava (Manihot esculenta Crantz) to salinity: implications for food security in low-lying regions. Journal of Experimental Botany, **67**, 5403-5413.

Gómez-Díaz J. S., Montoya-Lerma J., Muñoz-Valencia V. (2019) Prevalence and diversity of endosymbionts in cassava whiteflies (Hemiptera: Aleyrodidae) from Colombia. Journal of Insect Science, **19**.

Gutierrez S., Zwart M. P. (2018) Population bottlenecks in multicomponent viruses: first forays into the uncharted territory of genome-formula drift. Current Opinion in Virology, **33**, 184-190.

Hadjistylli M., Roderick G. K., Brown J. K. (2016) Global population structure of a worldwide pest and virus vector: Genetic diversity and population history of the Bemisia tabaci sibling species group. PLoS ONE, **11**, e0165105.

Hodges G. S., Evans G. A. (2005) An iddntification guide to the whiteflies (Hemiptera: Aleyrodidae) of the southeastern United States. Florida Entomologist, **88**, 518-534, 517.

Jacobson A. L., Duffy S., Sseruwagi P. (2018) Whitefly-transmitted viruses threatening cassava production in Africa. Current Opinion in Virology, **33**, 167-176.

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.

Kumar S., Stecher G., Tamura K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution, **33**, 1870-1874.

Lee W., Park J., Lee G. S., Lee S., Akimoto S. I. (2013) Taxonomic status of the Bemisia tabaci complex (Hemiptera: Aleyrodidae) and reassessment of the number of its constituent species. PLoS ONE, **8**, e63817.

Lefeuvre P., Martin D. P., Harkins G., Lemey P., Gray A. J. A., Meredith S., Lakay F., Monjane A., Lett J. M., Varsani A., Heydarnejad J. (2010) The spread of tomato yellow leaf curl virus from the Middle East to the world. PLoS Pathogens, **6**, e1001164.

Legg J., Fauquet C. (2004) Cassava mosaic geminiviruses in Africa. Plant Molecular Biology, **56**, 585-599.

Legg J. P., Lava Kumar P., Makeshkumar T., Tripathi L., Ferguson M., Kanju E., Ntawuruhunga P., Cuellar W. (2015) Cassava virus diseases: Biology, epidemiology, and management. Advances in Virus Research, **91**, 85-142.

Legg J. P., Sseruwagi P., Boniface S., Okao-Okuja G., Shirima R., Bigirimana S., Gashaka G., Herrmann H. W., Jeremiah S., Obiero H., Ndyetabula I., Tata-Hangy W., Masembe C., Brown J. K. (2014) Spatio-temporal patterns of genetic change amongst populations of cassava Bemisia tabaci whiteflies driving virus pandemics in East and Central Africa. Virus Research, **186**, 61-75.

Leotard G., Duputie A., Kjellberg F., Douzery E. J., Debain C., de Granville J. J., McKey D. (2009) Phylogeography and the origin of cassava: new insights from the northern rim of the Amazonian basin. Mol Phylogenet Evol, **53**, 329-334.

Liu B. M., Yan F. M., Chu D., Pan H. P., Jiao X. G., Xie W., Wu Q. J., Wang S. L., Xu B. Y., Zhou X. G., Zhang Y. J. (2012) Difference in feeding behaviors of two invasive whiteflies on host

plants with different suitability: implication for competitive displacement. International Journal of Biological Sciences, **8**, 697-706.

Luan J.-b., Xu J., Lin K.-k., Zalucki M. P., Liu S.-s. (2012) Species exclusion between an invasive and an indigenous whitefly on host plants with differential levels of suitability. Journal of Integrative Agriculture, **11**, 215-224.

Mabvakure B., Martin D. P., Kraberger S., Cloete L., van Brunschot S., Geering A. D., Thomas J. E., Bananej K., Lett J. M., Lefeuvre P., Varsani A., Harkins G. W. (2016) Ongoing geographical spread of Tomato yellow leaf curl virus. Virology, **498**, 257-264.

Macedo M. A., Albuquerque L. C., Maliano M. R., Souza J. O., Rojas M. R., Inoue-Nagata A. K., Gilbertson R. L. (2017) Characterization of tomato leaf curl purple vein virus, a new monopartite New World begomovirus infecting tomato in Northeast Brazil. Archives of Virology, **163**, 737-743.

Malka O., Santos-Garcia D., Feldmesser E., Sharon E., Krause-Sakate R., Delatte H., van Brunschot S., Patel M., Visendi P., Mugerwa H., Seal S., Colvin J., Morin S. (2018) Species-complex diversification and host-plant associations in Bemisia tabaci: A plant-defence, detoxification perspective revealed by RNA-Seq analyses. Molecular Ecology, **27**, 4241-4256.

Mar T. B., Mendes I. R., Lau D., Fiallo-Olive E., Navas-Castillo J., Alves M. S., Zerbini F. M. (2017a) Interaction between the New World begomovirus Euphorbia yellow mosaic virus and its associated alphasatellite: effects on infection and transmission by the whitefly Bemisia tabaci. Journal of General Virology, **98**, 1552-1562.

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. (2017b) Genetic variability and population structure of the New World begomovirus Euphorbia yellow mosaic virus. Journal of General Virology, **98**, 1537-1551.

Marubayashi J. M., Kliot A., Yuki V. A., Rezende J. A., Krause-Sakate R., Pavan M. A., Ghanim M. (2014) Diversity and localization of bacterial endosymbionts from whitefly species collected in Brazil. PLoS ONE, **9**, e108363.

Marubayashi J. M., Yuki V. A., Rocha K. C. G., Mituti T., Pelegrinotti F. M., Ferreira F. Z., Moura M. F., Navas-Castillo J., Moriones E., Pavan M. A., Krause-Sakate R. (2013) At least two indigenous species of the Bemisia tabaci complex are present in Brazil. Journal of Applied Entomology, **137**, 113-121.

Maruthi M. N., Jeremiah S. C., Mohammed I. U., Legg J. P. (2017) The role of the whitefly, Bemisia tabaci (Gennadius), and farmer practices in the spread of cassava brown streak ipomoviruses. Journal of Phytopathology, **165**, 707-717.

Milenovic M., Wosula E. N., Rapisarda C., Legg J. P. (2019) Impact of host plant species and whitefly species on feeding behavior of Bemisia tabaci. Frontiers in Plant Science, **10**, 1.

Montagnac J. A., Davis C. R., Tanumihardjo S. A. (2009) Nutritional value of cassava for use as a staple food and recent advances for improvement. Comprehensive Reviews in Food Science and Food Safety, **8**, 181-194.

Moraes L. A., Marubayashi J. M., Yuki V. A., Ghanim M., Bello V. H., De Marchi B. R., Barbosa L. F., Boykin L. M., Krause-Sakate R., Pavan M. A. (2017) New invasion of Bemisia tabaci Mediterranean species in Brazil associated to ornamental plants. Phytoparasitica, **45**, 517-525.

Moraes L. A., Muller C., Bueno R., Santos A., Bello V. H., De Marchi B. R., Watanabe L. F. M., Marubayashi J. M., Santos B. R., Yuki V. A., Takada H. M., Barros D. R., Neves C. G., Silva F. N., Goncalves M. J., Ghanim M., Boykin L., Pavan M. A., Krause-Sakate R. (2018) Distribution and phylogenetics of whiteflies and their endosymbiont relationships after the Mediterranean species invasion in Brazil. Scientific Reports, **8**, 14589.

Morris E. K., Caruso T., Buscot F., Fischer M., Hancock C., Maier T. S., Meiners T., Müller C., Obermaier E., Prati D., Socher S. A., Sonnemann I., Wäschke N., Wubet T., Wurst S., Rillig M. C. (2014) Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. Ecology and Evolution, **4**, 3514-3524.

Mugerwa H., Seal S., Wang H.-L., Patel M. V., Kabaalu R., Omongo C. A., Alicai T., Tairo F., Ndunguru J., Sseruwagi P., Colvin J. (2018) African ancestry of New World, Bemisia tabaci-whitefly species. Scientific Reports, **8**, 2734.

Muhire B. M., Varsani A., Martin D. P. (2014) SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLoS ONE, **9**, e108277.

Nassar N. (2001) Cassava, Manihot esculenta Crantz and wild relatives: their relationships and evolution. Genetic Resources and Crop Evolution, **48**, 429-436.

Navas-Castillo J., Fiallo-Olivé E., Sánchez-Campos S. (2011) Emerging virus diseases transmitted by whiteflies. Annual Review of Phytopathology, **49**, 219-248.

Ndunguru J., Legg J., Aveling T., Thompson G., Fauquet C. (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. Virology Journal, **2**, 21.

Nylander J. A. A. (2004) MrModeltest v2. In: Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Olsen K. M. (2004) SNPs, SSRs and inferences on cassava's origin. Plant Molecular Biology, **56**, 517-526.

Olsen K. M., Schaal B. A. (1999) Evidence on the origin of cassava: phylogeography of Manihot esculenta. Proceedings of the National Academy of Sciences, USA, **96**, 5586-5591.

Pan H., Chu D., Yan W., Su Q., Liu B., Wang S., Wu Q., Xie W., Jiao X., Li R., Yang N., Yang X., Xu B., Brown J. K., Zhou X., Zhang Y. (2012) Rapid spread of tomato yellow leaf curl virus in China is aided differentially by two invasive whiteflies. PLoS ONE, **7**, e34817.

Pan L. L., Cui X. Y., Chen Q. F., Wang X. W., Liu S. S. (2018) Cotton Leaf Curl Disease: Which Whitefly Is the Vector? Phytopathology, **108**, 1172-1183.

Patil B. L., Fauquet C. M. (2009) Cassava mosaic geminiviruses: actual knowledge and perspectives. Molecular Plant Pathology, **10**, 685-701.

Paz-Carrasco L. C., Castillo-Urquiza G. P., Lima A. T., Xavier C. A., Vivas-Vivas L. M., Mizubuti E. S., Zerbini F. M. (2014) Begomovirus diversity in tomato crops and weeds in Ecuador and the detection of a recombinant isolate of rhynchosia golden mosaic Yucatan virus infecting tomato. Archives of Virology, **159**, 2127-2132.

Qin L., Pan L.-L., Liu S.-S. (2016) Further insight into reproductive incompatibility between putative cryptic species of the Bemisia tabaci whitefly complex. Insect Science, **23**, 215-224.

R Development Core Team (2007) R: A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria.

Rey C., Vanderschuren H. (2017) Cassava mosaic and brown streak diseases: Current perspectives and beyond. Annual Review of Virology, **4**, 429-452.

Ribeiro S. G., Ávila A. C., Bezerra I. C., Fernandes J. J., Faria J. C., Lima M. F., Gilbertson R. L., Zambolim E. M., Zerbini F. M. (1998) Widespread occurrence of tomato geminiviruses in Brazil, associated with the new biotype of the whitefly vector. Plant Disease, **82**, 830.

Rocha C. S., Castillo-Urquiza G. P., Lima A. T. M., Silva F. N., Xavier C. A. D., Hora-Junior B. T., Beserra-Junior J. E. A., Malta A. W. O., Martin D. P., Varsani A., Alfenas-Zerbini P., Mizubuti E. S. G., Zerbini F. M. (2013) Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. Journal of Virology, **87**, 5784-5799.

Rocha K. C. G., Marubayashi J. M., Navas-Castillo J., Yuki V. A., Wilcken C. F., Pavan M. A., Krause-Sakate R. (2011) Only the B biotype of Bemisia tabaci is present on vegetables in Sao Paulo State, Brazil. Scientia Agricola, **68**, 120-123.

Rodríguez-Negrete E. A., Morales-Aguilar J. J., Domínguez-Duran G., Torres-Devora G., Camacho-Beltrán E., Leyva-López N. E., Voloudakis A. E., Bejarano E. R., Méndez-Lozano J. (2019) High-throughput sequencing reveals differential begomovirus species diversity in non-cultivated plants in Northern-Pacific Mexico. Viruses, **11**, 594.

Rojas M. R., Gilbertson R. L., Russell D. R., Maxwell D. P. (1993) Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. Plant Disease, **77**, 340-347.

Ronquist F., Huelsenbeck J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, **19**, 1572-1574.

Sacristan S., Malpica J. M., Fraile A., Garcia-Arenal F. (2003) Estimation of population bottlenecks during systemic movement of Tobacco mosaic virus in tobacco plants. Journal of Virology, **77**, 9906-9911.

Scott I. A. W., Workman P. J., Drayton G. M., Burnip G. M. (2007) First record of Bemisia tabaci biotype Q in New Zealand. New Zealand Plant Protection, **60**, 264-270.

Simon C., Frati F., Beckenbach A., Crespi B., Liu H., Flook P. (1994) Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America, **87**, 651-701.

Sottoriva L. D. M., Lourenção A. L., Colombo C. A. (2014) Performance of Bemisia tabaci (Genn.) biotype B (Hemiptera: Aleyrodidae) on weeds. Neotropical Entomology, **43**, 574-581.

Sun D.-B., Liu Y.-Q., Qin L., Xu J., Li F.-F., Liu S.-S. (2013) Competitive displacement between two invasive whiteflies: insecticide application and host plant effects. Bulletin of Entomological Research, **103**, 344-353.

Tajebe L. S., Boni S. B., Guastella D., Cavalieri V., Lund O. S., Rugumamu C. P., Rapisarda C., Legg J. P. (2015) Abundance, diversity and geographic distribution of cassava mosaic disease pandemic-associated Bemisia tabaci in Tanzania. Journal of Applied Entomology, **139**, 627-637.

Tiendrebeogo F., Lefeuvre P., Hoareau M., Harimalala M. A., De Bruyn A., Villemot J., Traore V. S., Konate G., Traore A. S., Barro N., Reynaud B., Traore O., Lett J. M. (2012) Evolution of African cassava mosaic virus by recombination between bipartite and monopartite begomoviruses. Virology Journal, **9**, 67.

Tocko-Marabena B. K., Silla S., Simiand C., Zinga I., Legg J., Reynaud B., Delatte H. (2017) Genetic diversity of Bemisia tabaci species colonizing cassava in Central African Republic characterized by analysis of cytochrome c oxidase subunit I. PLoS ONE, **12**.

Vasquez-Ordonez A. A., Hazzi N. A., Escobar-Prieto D., Paz-Jojoa D., Parsa S. (2015) A geographic distribution database of the Neotropical cassava whitefly complex (Hemiptera, Aleyrodidae) and their associated parasitoids and hyperparasitoids (Hymenoptera). Zookeys, 75-87.

Vyskočilová S., Seal S., Colvin J. (2019) Relative polyphagy of "Mediterranean" cryptic Bemisia tabaci whitefly species and global pest status implications. Journal of Pest Science, **92**, 1071-1088.

Vyskočilová S., Tay W. T., van Brunschot S., Seal S., Colvin J. (2018) An integrative approach to discovering cryptic species within the Bemisia tabaci whitefly species complex. Scientific Reports, **8**, 10886.

Watanabe L. F. M., Bello V. H., De Marchi B. R., Silva F. B. d., Fusco L. M., Sartori M. M., Pavan M. A., Krause-Sakate R. (2019) Performance and competitive displacement of Bemisia tabaci MEAM1 and MED cryptic species on different host plants. Crop Protection, **124**, 104860.

Watling J., Shock M. P., Mongeló G. Z., Almeida F. O., Kater T., De Oliveira P. E., Neves E. G. (2018) Direct archaeological evidence for Southwestern Amazonia as an early plant domestication and food production centre. PLoS ONE, **13**, e0199868.

Whitfield A. E., Falk B. W., Rotenberg D. (2015) Insect vector-mediated transmission of plant viruses. Virology, **480**, 278-289.

Wilson A. K. (1981) Euphorbia heterophylla: A review of distribution, importance and control. Tropical Pest Management, **27**, 32-38.

Wosula E. N., Chen W., Fei Z., Legg J. P. (2017) Unravelling the genetic diversity among cassava Bemisia tabaci whiteflies using NextRAD sequencing. Genome Biol Evol, **9**, 2958-2973.

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

Table 2. Sampled sites and whiteflies species detected in cassava in Brazil.

Sample	Date of collection	Location	Region	Geographical coordinates		Cultivation system	Number of samples			Whiteflies species[&]	Reference
				Latitude	Longitude		Adults	Nymphs	Total		
AL1	July 2018	Arapiraca, AL	Northeast	09° 48' 27.66"S	36° 36' 40.68"W		15	15	30	Btu, Ta	This study
AL2	April 2018	TeotonioVilela, AL	Northeast	09° 53' 16.86"S	36° 23' 05.52"W		14	15	29	BtM, Btu, Ta	This study
AL3	April 2018	União dos Palmares, AL	Northeast	09° 11' 51.84"S	36° 01"51.90"W		16	15	31	BtM, Ta	This study
AL4	April 2018	Arapiraca, AL	Northeast	09° 47' 29.46"S	36° 25' 36.00"W		15	15	30	BtM, BtNW, Btu, Ta	This study
AL5	April 2018	Arapiraca, AL	Northeast	09° 43' 09.30"S	36° 40' 45.60"W		14	10	24	BtM, BtNW, Btu, Ta	This study
BA1	December 2017	Barra, BA	Northeast	11° 20' 52.37"S	43° 12' 57.44"W		28	0	28	BtM, Btu, Ta	This study
BA2	March 2018	LEM, BA	Northeast	12° 20' 10.00"S	45° 49' 12.00"W		13	0	13	BtM, Ta	This study
BA3	March 2018	Wanderley, BA	Northeast	12° 13' 42.04"S	43° 55' 50.00"W		10	10	20	BtM, Btu	This study
BA4	June 2018	Cristópolis, BA	Northeast	12° 13' 56.08"S	44° 23' 11.50"W		14	0	14	BtM, Ta	This study
DF1	March 2018	Planaltina, DF	Midwest	15° 30' 59.00"S	47° 16' 09.00"W		10	0	10	BtM	This study
DF2	March 2018	Planaltina, DF	Midwest	15° 28' 57.00"S	47° 20' 06.00"W		10	0	10	BtM	This study
DF3	March 2021	Planaltina, DF	Midwest	15° 31' 17.00"S	47° 21' 22.00"W		10	0	10	BtM	This study
ES1	January 2018	Sooretama, ES	Southeast	19° 06' 52.02"S	40° 04' 46.30"W		14	15	29	BtM, Ta	This study
ES2	January 2018	Marilândia, ES	Southeast	19° 24' 22.04"S	40° 32' 21.70"W		15	15	30	BtM, Btu, Ta	This study
ES3	January 2018	Pinheiros, ES	Southeast	18° 40' 83.20"S	40° 28' 63.40"W		11	11	22	Ta	This study
GO1	March 2018	Bela Vista, GO	Midwest	16° 59' 50.00"S	48° 57' 56.00"W		10	11	21	BtM	This study
GO2	March 2018	Itaberaí, GO	Midwest	15° 56' 58.00"S	49° 47' 07.00"W		15	15	30	BtM, Btu	This study
MG1	May 2018	Ouro Fino, MG	Southeast	22° 16' 44.00"S	46° 29' 33.00"W		12	10	22	Ta	This study
MG10	February 2018	Florestal, MG	Southeast	19° 52' 38.00"S	44° 25' 21.00"W		15	15	30	Btu, Ta	This study
MG11	March 2018	Florestal, MG	Southeast	19° 52' 38.00"S	44° 25' 21.00"W		15	15	30	Btu, Ta	This study
MG12	May 2018	Divinópolis, MG	Southeast	20° 06' 21.00"S	44° 55' 36.00"W		15	15	30	Btu, Ta	This study
MG13	August 2018	Viçosa, MG	Southeast	20° 46' 06.00"S	42° 52' 14.00"W		17	0	17	Btu, Ta	This study
MG14	March 2018	Piraúba, MG	Southeast	21° 16' 22.71"S	43° 02' 31.28"W		10	10	20	Ta	This study
MG15	March 2018	Descoberto, MG	Southeast	21° 28' 06.28"S	42° 58' 05.53"W		10	10	20	Ta	This study
MG16	March 2018	Mar de Espanha, MG	Southeast	21° 46' 07.26"S	43° 04' 26.62"W		15	15	30	BtM, Ta	This study

MG17	March 2018	Leopoldina, MG	Southeast	21° 33' 50.15"S	42° 40' 42.97"W	16	13	29	BtM, Btu, Ta	This study
MG18	March 2018	Dona Euzébia, MG	Southeast	21° 19' 18.83"S	42° 48' 37.07"W	15	14	29	Ta	This study
MG19	February 2018	Caparaó, MG	Midwest	20° 31' 48.00"S	41° 54' 00.36"W	15	15	30	BtM, Ta	This study
MG2	July 2018	Pouso Alegre, MG	Southeast	22° 15' 01.00"S	46° 58' 31.00"W	9	0	9	Ta	This study
MG3	February 2018	Careaçu, MG	Southeast	22° 04' 37.00"S	45° 41' 49.00"W	11	11	22	Ta	This study
MG4	June 2018	Lambari, MG	Southeast	21° 56' 05.00"S	45° 15' 49.00"W	15	15	30	Btu, Ta	This study
MG5	February 2018	Lima Duarte, MG	Southeast	21° 50' 35.00"S	43° 47' 01.00"W	10	10	20	Ta	This study
MG6	April 2018	RioPomba, MG	Southeast	21° 15' 50.00"S	43° 09' 59.00"W	10	0	10	Ta	This study
MG7	February 2018	Florestal, MG	Southeast	19° 54' 13.00"S	44° 25' 48.00"W	15	15	30	BtM, Btu, Ta	This study
MG8	February 2018	Florestal, MG	Southeast	19° 51' 20.00"S	44° 23' 58.00"W	15	15	30	Btu, Btu-like	This study
MG9	March 2018	Florestal, MG	Southeast	19° 53' 39.00"S	44° 24' 55.00"W	14	11	25	BtM, BtNW, Btu, Ta	This study
MT1	December 2017	Canarana, MT	Midwest	13° 31' 16.00"S	52° 25' 03.00"W	15	16	31	Btu, Ta	This study
MT2	December 2017	Canarana, MT	Midwest	13° 33' 47.00"S	52° 15' 53.00"W	0	20	20	Btu	This study
MT4	January 2018	Pedra Preta, MT	Midwest	16° 38' 29.00"S	54° 25' 41.00"W	10	10	20	Btu	This study
MT5	January 2018	Pedra Preta, MT	Midwest	16° 39' 07.00"S	54° 22' 27.00"W	10	10	20	BtM, Btu	This study
MT6	January 2018	Pedra Preta, MT	Midwest	16° 39' 28.00"S	54° 20' 20.00"W	10	10	20	BtM, Btu	This study
PA1	January 2018	Brasil Novo, PA	North	03° 12' 23.07"S	52° 30' 13.80"W	15	15	30	Btu, Ta	This study
PA2	January 2018	Vitória doXingu, PA	North	03° 04' 51.01"S	52° 10' 08.80"W	15	16	31	Btu, Ta	This study
PA3	August 2018	Altamira, PA	North	03° 18' 15.03"S	52° 07' 26.00"W	10	10	20	BtM	This study
PA4	January 2018	Altamira, PA	North	03° 09' 14.60"S	52° 07' 49.00"W	11	10	21	Ta	This study
PA5	January 2018	Belém, PA	North	01° 18' 20.00"S	48° 26' 46.00"W	10	12	22	Ta	This study
PI1	April 2018	Picos, PI	Northeast	07° 04' 79.50"S	41° 25' 57.80"W	0	29	29	Btu, Ta	This study
PI2	May 2018	Teresina, PI	Northeast	05° 02' 41.94"S	42° 47' 18.84"W	0	27	27	Btu, Ta	This study
PR1	March 2018	Santo Antôniodo Caiuá, PR	South	22° 41' 07.00"S	52° 19' 06.00"W	30	0	30	Btu, Ta	This study
PR2	March 2018	Santo Antôniodo Caiuá, PR	South	22° 49' 03.00"S	52° 21' 25.00"W	30	0	30	BtM, Btu	This study
PR3	March 2018	Paranavaí, PR	South	23° 06' 03.29"S	52° 29' 10.00"W	10	10	20	BtM, BtMED	This study
PR4	March 2018	Sertanópolis, PR	South	23° 02' 54.00"S	50° 59' 54.00"W	26	0	26	Btu, Ta	This study
SC2	March 2018	Agrônomica, SC	South	27° 34' 40.00"S	48° 32' 08.00"W	10	0	10	BtM	This study
SP1	July 2016	Holambra, SP	Southeast	22° 36' 26.00"S	47° 02' 50.00"W	10	0	10	Btu	Moraes et al., 2018

SP10	March 2016	Casa Branca, SP	Southeast	21° 49' 08.00"S	45° 58' 23.00"W	10	0	10	BtM	Moraes et al., 2018
SP10	July 2016	São Pedro do Turvo, SP	Southeast	22° 36' 32.00"S	49° 45' 29.00"W	10	0	10	BtMED	Moraes et al., 2018
SP11	July 2016	Oleo, SP	Southeast	22° 56' 32.00"S	49° 26' 15.00"W	10	0	10	Btu	Moraes et al., 2018
SP12	July 2016	Mogi Mirim, SP	Southeast	22° 28' 32.30"S	47° 00' 47.60"W	10	0	10	Btu	Moraes et al., 2018
SP2	July 2016	Mogi Mirim, SP	Southeast	22° 28' 08.00"S	46° 56' 25.00"W	10	0	10	Btu	Moraes et al., 2018
SP3	July 2016	Mogi Mirim, SP	Southeast	22° 24' 59.00"S	46° 59' 19.00"W	10	0	10	Btu	Moraes et al., 2018
SP4	February 2019	Mogi Mirim, SP	Southeast	22° 26' 44.00"S	47° 04' 11.00"W	5	5	10	Btu	This study
SP5	July 2017	Mogi Mirim, SP	Southeast	22° 27' 05.00"S	47° 04' 56.00"W	10	0	10	Btu	Moraes et al., 2018
SP6	January 2019	SãoPedro, SP	Southeast	22° 34' 08.00"S	48° 05' 22.00"W	10	0	10	BtMED	This study
SP7	July 2017	Montalvao, SP	Southeast	22° 02' 23.00"S	51° 19' 53.00"W	10	0	10	BtM	Moraes et al., 2018
SP8	July 2016	Pindamonhangaba, SP	Southeast	22° 56' 05.00"S	45° 26' 25.00"W	10	0	10	Btu	Moraes et al., 2018
SP9	January 2019	Casa Branca, SP	Southeast	21° 11' 32.00"S	47° 48' 44.00"W	4	10	14	BtM, Btu, BtMED	This study
Total						819	566	1385		

[*] Brazilian states where samples were collected: AL, Alagoas; BA, Bahia; DF, Distrito Federal; ES, Espírito Santo; GO, Goiás; MG, Minas Gerais; MT, Mato Grosso; PA, Pará; PI, Piauí; PR, Paraná; SC, Santa Catarina; SP, São Paulo.

[&] Btu, Bemisia tuberculata; BtM, Bemisia tabaci MEAM1; BtMED, B. tabaci MED; BtNW, B. tabaci NW; Ta, Tetraleuroides acaciae.

Figure legends

Figure 1. A. Clockwise from top-left: Adults and nymphs of Bemisia tuberculata colonizing cassava in Mogi Mirim, São Paulo state. Growth of sooty mould fungus on the leaf surface due to the deposition of honeydew by whiteflies. Presence of begomovirus-infected Blainvillea rhomboidea (family Asteraceae) in a cassava field in Minas Gerais state. Presence of begomovirus-infected Euphorbia heterophylla (family Euphorbiaceae) in a cassava field in Minas Gerais state. (**B**) Map of Brazil showing the locations where whiteflies samples were collected. The map is colored according to the regions as indicated in the legend. Blacks dots correspond to the sampled sites. (**C**) Number of adults and nymphs analyzed from each sampled site according to state. AL, Alagoas; BA, Bahia; DF, Distrito Federal; ES, Espírito Santo; GO, Goiás; MG, Minas Gerais; MT, Mato Grosso; PA, Pará; Piauí; PR, Paraná; SC, Santa Catarina; SP, São Paulo.

Figure 2. Bayesian phylogenetic tree based on partial nucleotide sequences of the mitochondrial cytochrome oxidase (mtCOI) gene of representative individuals of each whitefly species detected in this study and reference sequences retrieved from GenBank. The tree was rooted with the aphid Aphys gossypii. Bayesian posterior probabilities are shown at the nodes. The scale bar represents the number of nucleotide substitutions per site. Nodes with posterior probability values between 0.60 and 0.80 are indicated by empty circles and nodes with values equal to or greater than 0.81 are indicated by filled circles. Clades highlighted with different colors indicate the species detected in this study. Branches highlighted in red indicate the putative new species detected here.

Figure 3. Composition of whitefly populations colonizing cassava in Brazil. **A.** Species composition at each sampled site according to stage of development (adult and nymphs). Asterisk indicate that nymphs was not detected. **B, C, D.** Species distribution of the 1,385 individuals genotyped in this study considering the samples from all sites (**B**) or only samples from sites where both adults and nymphs were sampled (**C**) or without samples from Minas Gerais state (**D**).

Figure 4. A. Incidence of Trialeurodes acaciae, Bemisia tuberculata and B. tabaci MEAM1 in cassava fields in Brazil, measured as the percentage of sampled sites where at least one individual belonging to each one of the three species was detected. Other species detected at low incidence

are shown together as "others". **B.** Venn diagram showing the proportion of sites where each one of the three whitefly species occur alone or in different combinations. **C.** Competitive capacity inferred based on the prevalence of individuals from each of two species in fields where those two species were detected co-occurring. The horizontal line inside the box corresponds to the median. The asterisk indicates a significant difference according to the non-parametric Kruskal-Wallis test ($p < 0.05$).

Figure 5. Composition and species diversity of whitefly populations differ among Brazilian regions. **A.** Pie charts represent the distribution of the 1,385 individuals genotyped in this study in the five geographic regions of Brazil. **B.** Boxplots correspond to Simpson's index of diversity (1-D) calculated for each geographic region. The index was first calculated for each sampled site and grouped by geographic region. Different letters indicate significant differences between groups according to the non-parametric Kruskal-Wallis test followed by post hoc multiple comparison test ($p < 0.05$).

Figure 1

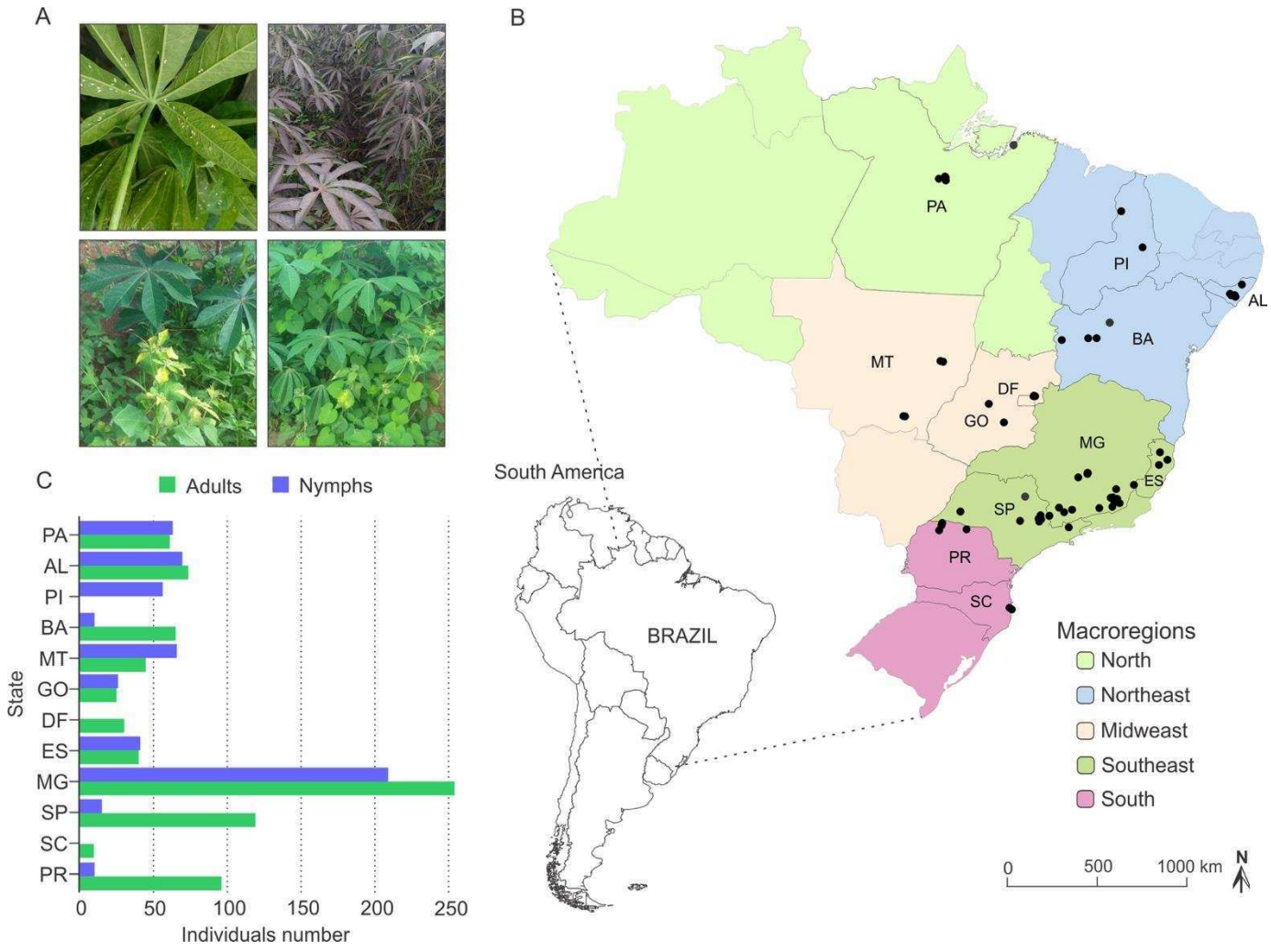

A
B
C
Adults
Nymphs
State
PA
AL
PI
BA
MT
GO
DF
ES
MG
SP
SC
PR
0
50
100
150
200
250
Individuals number
South America
BRAZIL
Macroregions
North
Northeast
Midweast
Southeast
South
0
500
1000 km
N

Figure 2

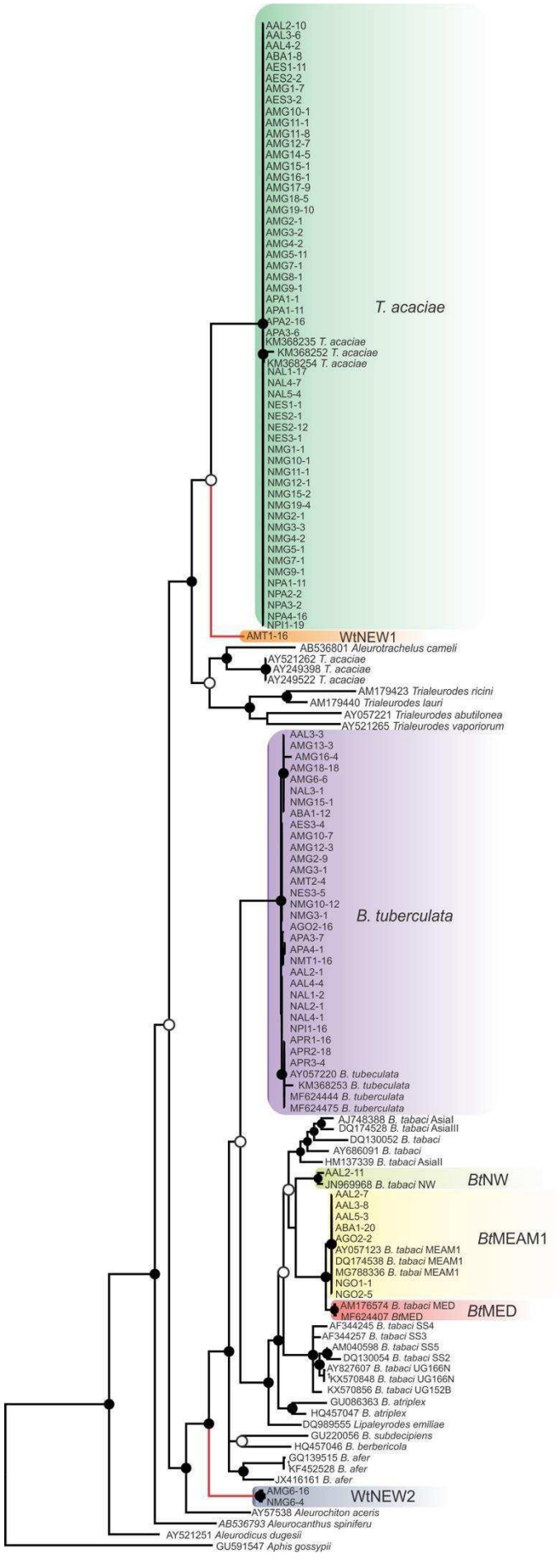

T. acaciae
WtNEW1
B. tuberculata
BtNW
BtMEAM1
BtMED
WtNEW2
0.2

Figure 3

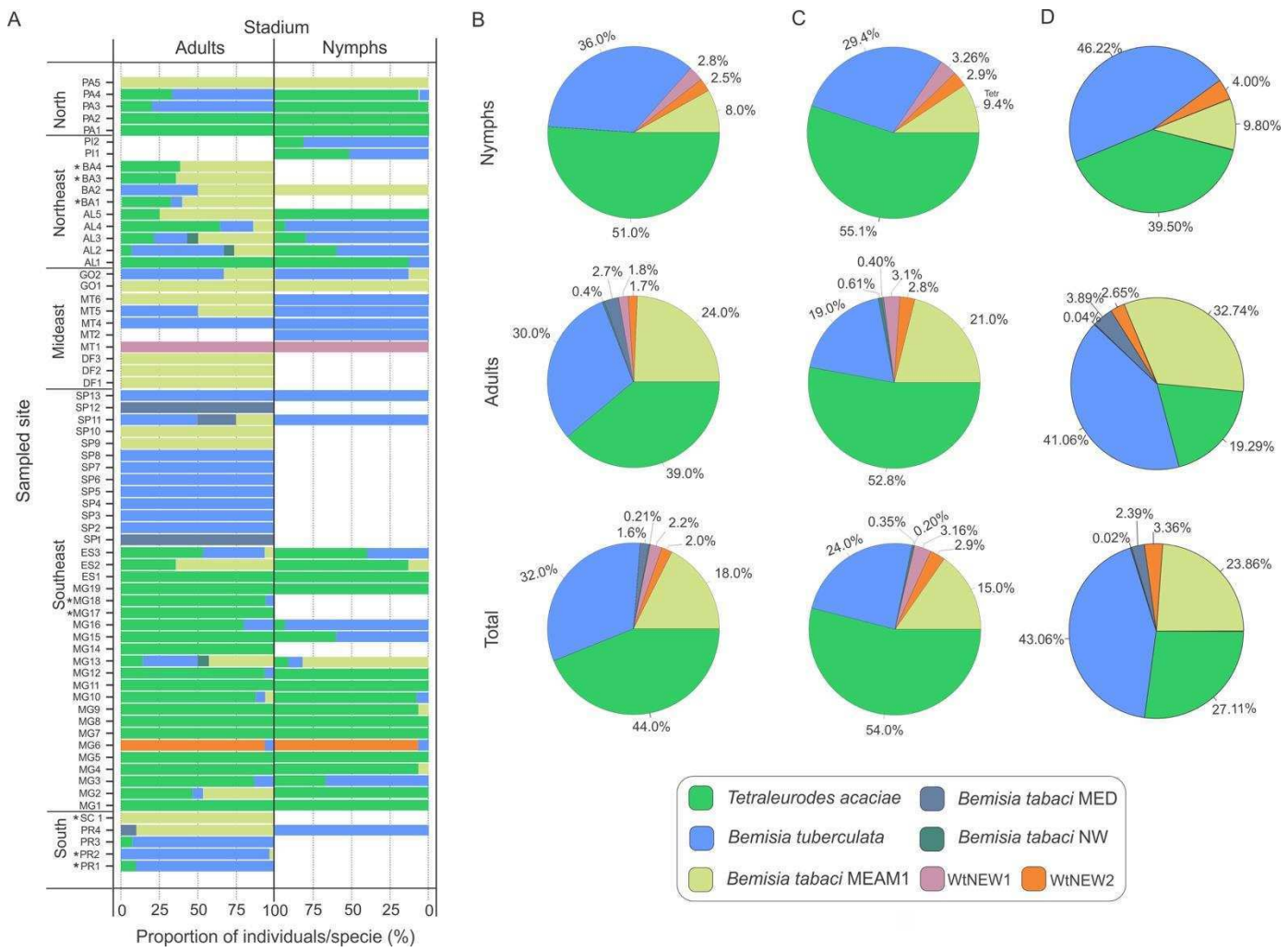

A
Stadium
Adults
Nymphs
Sampled site
North
Northeast
Mideast
Southeast
South
PA5
PA4
PA3
PA2
PA1
PI2
PI1
*BA4
*BA3
BA2
*BA1
AL5
AL4
AL3
AL2
AL1
GO2
GO1
MT6
MT5
MT4
MT2
MT1
DF3
DF2
DF1
SP13
SP12
SP11
SP10
SP9
SP8
SP7
SP6
SP5
SP4
SP3
SP2
SP1
ES3
ES2
ES1
MG19
*MG18
*MG17
MG16
MG15
MG14
MG13
MG12
MG11
MG10
MG9
MG8
MG7
MG6
MG5
MG4
MG3
MG2
MG1
*SC 1
PR4
PR3
*PR2
*PR1
0 25 50 75 100 75 50 25 0
Proportion of individuals/specie (%)
B
C
D
Nymphs
Adults
Total
36.0%
2.8%
2.5%
8.0%
51.0%
29.4%
3.26%
2.9%
Tetr
9.4%
55.1%
46.22%
4.00%
9.80%
39.50%
2.7%
1.8%
0.4%
1.7%
24.0%
30.0%
39.0%
0.40%
3.1%
0.61%
2.8%
19.0%
21.0%
52.8%
3.89%
2.65%
0.04%
32.74%
41.06%
19.29%
0.21%
1.6%
2.2%
2.0%
32.0%
18.0%
44.0%
0.35%
0.20%
24.0%
3.16%
2.9%
15.0%
54.0%
2.39%
0.02%
3.36%
23.86%
43.06%
27.11%
Tetraleurodes acaciae
Bemisia tuberculata
Bemisia tabaci MEAM1
Bemisia tabaci MED
Bemisia tabaci NW
WtNEW1
WtNEW2

Figure 4

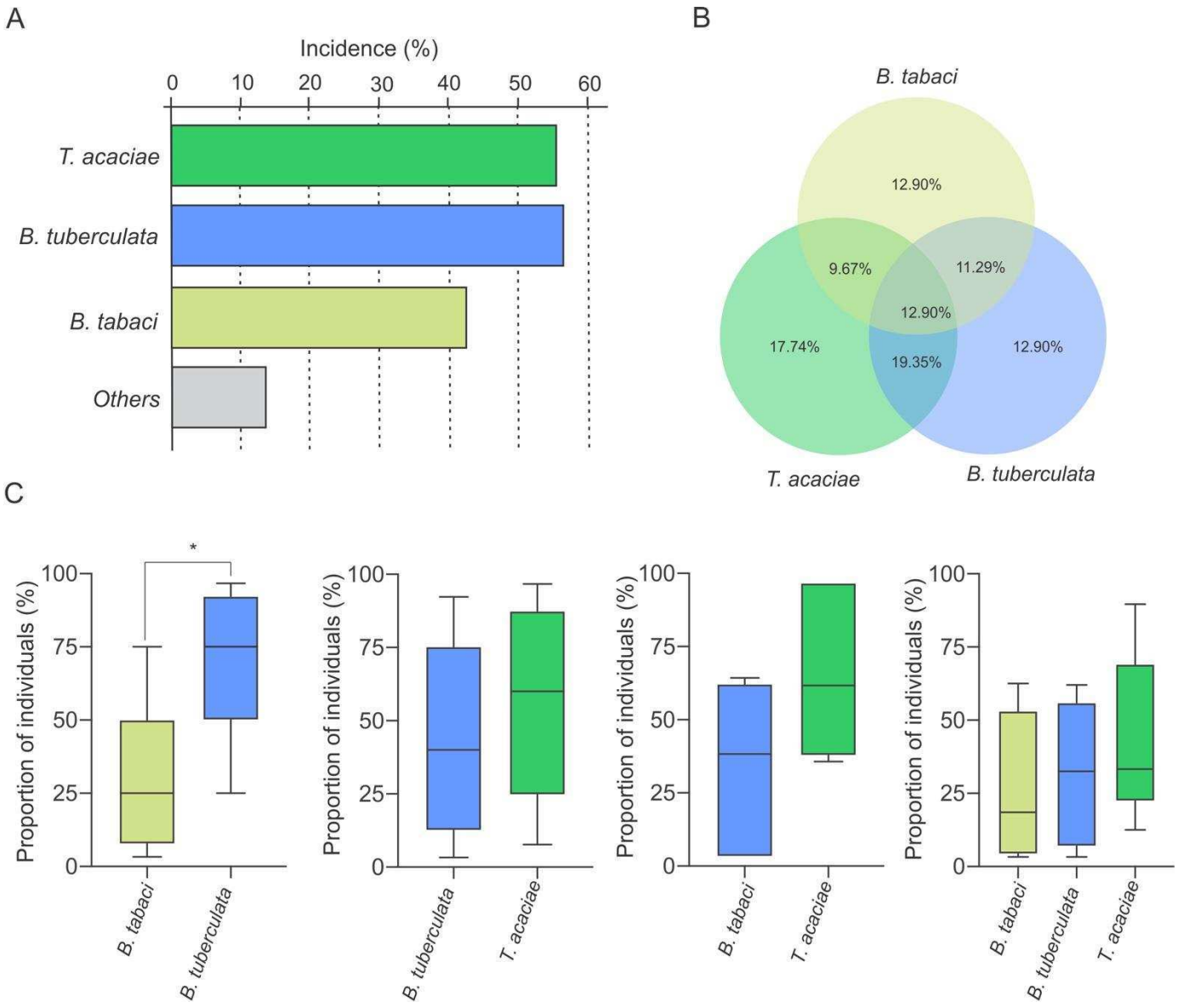
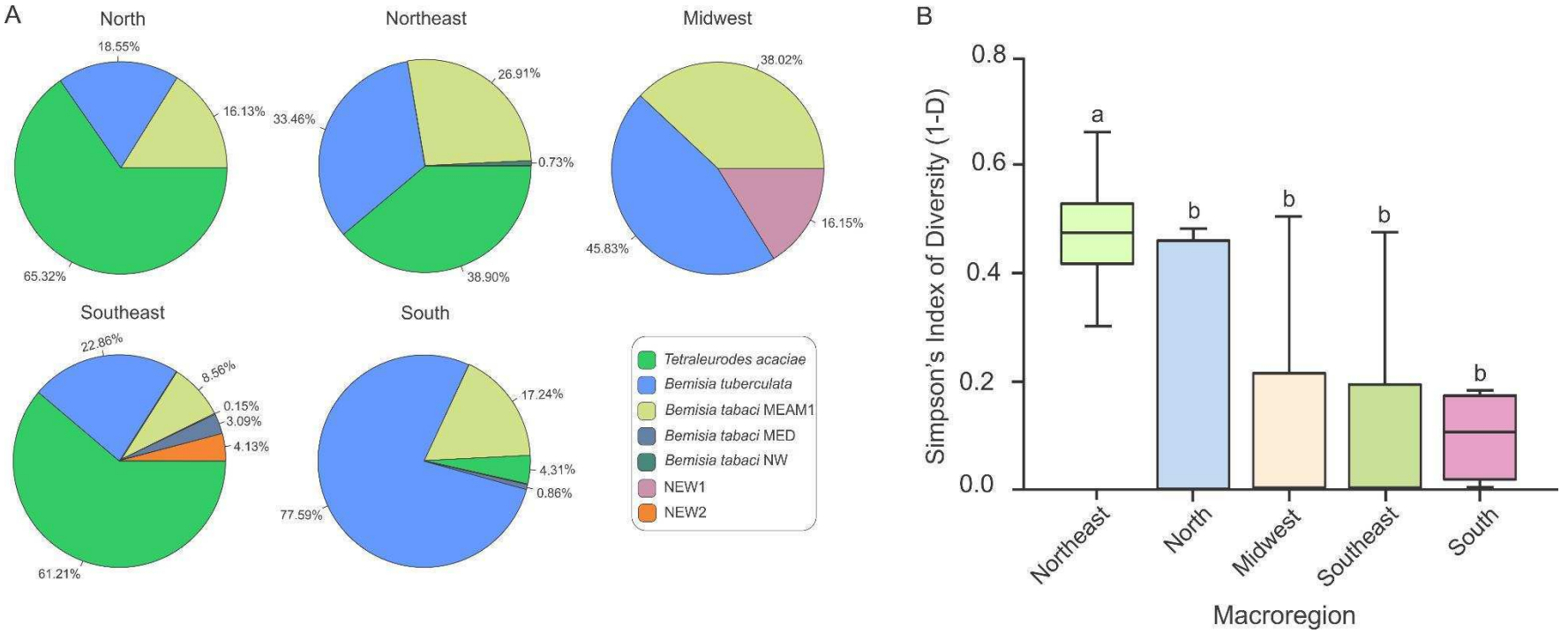

A
Incidence (%)
0
10
20
30
40
50
60
T. acaciae
B. tuberculata
B. tabaci
Others
B
B. tabaci
12.90%
9.67%
11.29%
12.90%
17.74%
19.35%
12.90%
T. acaciae
B. tuberculata
C
Proportion of individuals (%)
100
75
50
25
0
*
B. tabaci
B. tuberculata
B. tuberculata
T. acaciae
B. tabaci
T. acaciae
B. tabaci
B. tuberculata
T. acaciae

Figure 5


A
North
18.55%
16.13%
65.32%
Northeast
26.91%
33.46%
0.73%
38.90%
Midwest
38.02%
45.83%
16.15%
Southeast
22.86%
8.56%
0.15%
3.09%
4.13%
61.21%
South
17.24%
4.31%
0.86%
77.59%
Tetraleurodes acaciae
Bemisia tuberculata
Bemisia tabaci MEAM1
Bemisia tabaci MED
Bemisia tabaci NW
NEW1
NEW2
B
Simpson's Index of Diversity (1-D)
0.0
0.2
0.4
0.6
0.8
a
b
b
b
b
Northeast
North
Midwest
Southeast
South
Macroregion

CONCLUSIONS

- 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.

- Overall, we demonstrate here that two AtCDAs have evolved distinct functions upon CabLCV infection, independent of deamination activity. AtCDA seems to have evolved distinct functions for distinct viruses. In addition, it is demonstrated for the first time that AtCDLA4 may constitute a branch in antiviral immunity in plants.

- The prevalence of ToSRV, at least in relation to ToYSV, in tomato crops in the Southeast and Midwest regions of Brazil can be attributed to a higher transmission efficiency by the insect vector, as the well-adaptated ToYSV was not transmissible under our experimental conditions.

- The results presented here expand our knowledge of whitefly diversity in cassava and support the hypothesis that begomovirus epidemics have not occurred in cassava in Brazil due to the absence of competent vector populations. However, they indicate an ongoing adaptation process of B. tabaci Middle East-Asia Minor 1 to cassava, increasing the likelihood of begomovirus emergence in this crop in the future