TALITA BERNARDON MAR

GENETIC VARIABILITY AND POPULATION STRUCTURE OF THE BEGOMOVIRUS *Euphorbia yellow mosaic virus* AND ITS ASSOCIATED ALPHASATELLITE

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

VIÇOSA
MINAS GERAIS – BRASIL
2017

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

T
M298g
2017

Mar, Talita Bernardon, 1987-
Genetic variability and population structure of the begomovirus *Euphorbia yellow mosaic virus* and its associated alphasatellite / Talita Bernardon Mar. – Viçosa, MG, 2017.
vii, 123f. : il. (algumas color.) ; 29 cm.

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

1. Genética vegetal. 2. Vírus - Genética. 3. Melhoramento genético. I. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-graduação em Genética e Melhoramento. II. Título.

CDD 22 ed. 581.35

TALITA BERNARDON MAR

GENETIC VARIABILITY AND POPULATION STRUCTURE OF THE BEGOMOVIRUS *Euphorbia yellow mosaic virus* AND ITS ASSOCIATED ALPHASATELLITE

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

APROVADA: 15 de fevereiro de 2017.

Douglas Lau

Jorge Luis Badel Pacheco

Poliane Alfenas Zerbini

Anésia Aparecida dos Santos

Francisco Murilo Zerbini Júnior
(Orientador)

AGRADECIMENTOS

Ao Professor Francisco Murilo Zerbini, pela oportunidade, orientação, disponibilidade e confiança. Também pelo total apoio e paciência nos momentos decisivos;

Ao Professor Eduardo Seiti Gomide Mizubuti, pelos ensinamentos;

Ao Pesquisador Douglas Lau, pelo incentivo e confiança;

Aos colegas do Laboratório de Virologia Vegetal Molecular, pela amizade, em especial Larissa, Fernanda, Josiane, Hermano e Igor.

Aos funcionários do Programa de Pós-Graduação em Genética e Melhoramento Marco Tulio e Odilon, pela disponibilidade;

À minha família pelo apoio, compreensão e incentivo;

A todos os que contribuíram direta ou indiretamente para a realização deste trabalho.

À CAPES e ao CNPq, pela concessão da bolsa de estudo.

SUMÁRIO

ABSTRACT

MAR, Talita Bernardon, D.Sc., Universidade Federal de Viçosa, February, 2017. **Genetic variability and population structure of the begomovirus *Euphorbia yellow mosaic virus* and its associated alphasatellite.** Advisor: Francisco Murilo Zerbini Júnior. Co-advisors: Eduardo Seiti Gomide Mizubuti and Alison Talis Martins Lima.

The emergence of begomoviruses (whitefly-transmitted viruses classified in the genus *Begomovirus*, family *Geminiviridae*) in Brazil probably occurred by horizontal transfer of viruses infecting non-cultivated plants after the introduction of *Bemisia tabaci* MEAM1. Recently, Euphorbia yellow mosaic alphasatellite (EuYMA) was found in association with *Euphorbia yellow mosaic virus* (EuYMV) infecting *Euphorbia heterophylla*, an important invasive species in soybean and other crops in Brazil. We assessed the genetic variability and population structure of begomoviruses infecting *E. heterophylla*, and the geographical range of alphasatellites in samples collected throughout nine Brazilian states from 2009 to 2014. Only EuYMV was detected (with one exception), and a total of 158 and 57 haplotypes were compared in DNA-A and DNA-B datasets, respectively. EuYMA was detected in only six samples collected in the states of Rio Grande do Sul and Paraná. Compared with other begomoviruses, EuYMV has a low degree of genetic variation, and few intraspecific recombination events. The application of principal component analysis allowed the differentiation of six viral subpopulations according to sampling locations, in agreement with phylogenetic analysis. The polymorphisms of 23 sites that mostly contributed to the geographical structure were shown to hold supporting information to discriminate between the viral subpopulations. Biological and molecular aspects of the association between the alphasatellite EuYMA and EuYMV were studied following the construction of infectious clones of both agents. Phenotypic differences of EuYMV infection in the presence of EuYMA were observed in *Arabidopsis thaliana*, *Nicotiana benthamiana* and *E. heterophylla*. Transmission assays indicated that EuYMA negatively affects the transmission of EuYMV by *Bemisia tabaci* MEAM1. Together, the

results indicate that EuYMV displays a lower degree of genetic variation compared to other begomoviruses infecting cultivated and non-cultivated plants in Brazil but, similar to other begomoviruses, segregates according to sampling location, and that EuYMA is capable of modulating symptoms, viral accumulation and whitefly transmission of EuYMV, potentially interfering with virus dissemination in the field.

RESUMO

MAR, Talita Bernardon, D.Sc., Universidade Federal de Viçosa, fevereiro de 2017. **Variabilidade genética e estrutura de populações do begomovírus *Euphorbia yellow mosaic virus* e alfassatélite associado.** Orientador: Francisco Murilo Zerbini Júnior. Coorientadores: Eduardo Seiti Gomide Mizubuti e Alison Talis Martins Lima.

A emergência dos begomovírus (virus transmitidos por mosca-branca classificados no gênero *Begomovirus*, família *Geminiviridae*) no Brasil provavelmente ocorreu pela transferência horizontal de vírus presentes em plantas não-cultivadas após a introdução da *Bemisia tabaci* MEAM1. Recentemente, Euphorbia yellow mosaic alphasatellite (EuYMA) foi encontrado em associação com *Euphorbia yellow mosaic virus* (EuYMV) infectando *Euphorbia heterophylla*, uma importante invasora na cultura da soja no Brasil. A variabilidade genética e a estrutura da população de begomovírus infectando *E. heterophylla*, e novas ocorrências de EuYMA, foram investigados em amostras coletadas em nove estados Brasileiros entre 2009 e 2014. Apenas EuYMV foi detectado (com uma exceção), e um total de 158 e 57 haplótipos foram comparados em conjuntos de dados correspondentes ao DNA-A e DNA-B, respectivamente. EuYMA foi detectado em apenas seis amostras coletadas nos estados do Rio Grande do Sul e Paraná. Comparado com populações de outros begomovírus, o EuYMV possui um menor grau de variabilidade genética, e poucos eventos de recombinação intraespecíficos. Análise de componentes principais permitiu a diferenciação de seis subpopulações virais de acordo com os locais de amostragem, corroborando as análises filogenéticas. Os polimorfismos de 23 sítios que mais contribuíram para a estrutura geográfica foram descritos como suficientemente informativos para discriminar entre as subpopulações virais. A caracterização biológica e molecular da associação do alfassatélite EuYMA com o EuYMV foi realizada após a construção de clones infeciosos de ambos os agentes. Diferenças fenotípicas na infecção por EuYMV na presença do EuYMA foram observadas em *Arabidopsis thaliana*, *Nicotiana benthamiana* e *E. heterophylla*. Ensaios de transmissão indicaram que o EuYMA afeta

negativamente a transmissão de EuYMV por *Bemisia tabaci* MEAM1. Em conjunto, os resultados indicam que o EuYMV apresenta menor grau de variabilidade genética comparado a outros begomovírus que infectam plantas cultivadas e não-cultivadas no Brasil, porém, similar ao descrito para outros begomovírus, segrega de acordo com o local de amostragem. Além disso, o EuYMA é capaz de modular os sintomas, o acúmulo viral e a transmissão pela mosca-branca no contexto da infecção pelo EuYMV, potencialmente interferindo na disseminação do vírus no campo.

INTRODUCTION

The *Geminiviridae* family includes viruses whose genomes are composed of one or two molecules of circular single strand DNA encapsidated by a single structural protein into twinned quasi-icosahedral particles. The family comprises the genera *Begomovirus, Becurtovirus, Capulavirus, Curtovirus, Eragrovirus, Grablovirus, Mastrevirus, Topocuvirus* and *Turncurtovirus*, based on the type of insect vector, host range, phylogenetic relationship and genome organization (Varsani *et al.*, 2017; Zerbini *et al.*, 2017). Begomoviruses from the "New World" (the Americas) have two genomic components termed DNA-A and DNA-B, and are transmitted by the whitefly *Bemisia tabaci* to dicot plants. The DNA-A contains genes involved in replication and encapsidation of the viral progeny. The DNA-B contains genes required for intra- and intercellular movement in the plant (Lazarowitz, 1992). Both genomic components are required for systemic infection of the host.

Geminivirus populations, including begomoviruses, possess a high genetic diversity. The genus *Begomovirus* includes plant viruses of economical importance for Brazilian agriculture and a number of novel begomoviruses associated with tomato plants and weeds have been characterized over the last 25 years (Ribeiro *et al.*, 1994; Ribeiro *et al.*, 1996; Faria *et al.*, 1997; Galvão *et al.*, 1998; Castillo-Urquiza *et al.*, 2007; Cotrim *et al.*, 2007). The emergence of begomoviruses in Brazil probably occurred by horizontal transfer of indigenous viruses infecting non-cultivated plants after the introduction of *Bemisia tabaci* Middle East-Asia Minor 1 (MEAM1; previously known as *B. tabaci* biotype B) in the mid-1990's (Melo, 1992; Jones, 2003; Dinsdale *et al.*, 2010; Navas-Castillo *et al.*, 2011).

Alphasatellites (previously named DNA 1) are circular, single-stranded DNA molecules which replicate independently but require a helper begomovirus for systemic spread in the plant and for insect transmission (Zhou, 2013). The association of some bipartite New World begomoviruses with DNA satellites was recently demonstrated in non-cultivated plants in Brazil (Paprotka *et al.*, 2010) and Cuba (Jeske *et al.*, 2014), and in watermelon in Venezuela (Romay *et al.*, 2010). Also, sap transmission of a begomovirus/alphasatellite complex was

demonstrated for an alphasatellite described in Venezuela (Romay *et al.*, 2015). In Brazil, alphasatellites were detected in *Euphorbia heterophylla* and *Cleome affinis* plants sampled in Miranda (state of Mato Grosso do Sul) (Paprotka *et al.*, 2010), and in *Leonurus sibiricus* and *Sida* sp. plants sampled in Dourados (state of Mato Grosso do Sul) and Chapada (state of Rio Grande do Sul), respectively (Ferro *et al.*, 2017). They were all found to be phylogenetically related to alphasatellites reported in association with other bipartite New World begomoviruses (Ferro *et al.*, 2017).

Euphorbia heterophylla is an important invasive species in soybean and other crops in Brazil and Paraguay, where its center of origin is located (Wilson, 1981). Its occurrence is common in the southern, southeastern and midwestern regions of Brazil (Cronquist, 1981). There are reports of *E. heterophylla* herbicide-resistant genotypes in the state of Rio Grande do Sul, where it is present in 74% of soybean producing areas (Vidal & Winkler, 2004; Christoffoleti, 2008). Reports of possible begomoviruses infecting *E. heterophylla* in Brazil date back to the 1950's (Costa & Bennett, 1950). In 2011, *Euphorbia yellow mosaic virus* (EuYMV) was reported in *E. heterophylla* plants collected in the state of Goiás (Fernandes *et al.*, 2011). Later, the virus was found in Brazil in the non-cultivated plants *Sida santaremnensis* (Tavares *et al.*, 2012), *Macroptilium atropurpureus* (Silva *et al.*, 2012) and *Crotalaria juncea* (Barreto *et al.*, 2013). Experimentally, *Arabidopsis thaliana* is also a host of the virus. Barreto *et al.* (2013) demonstrated that *E. heterophylla* is one of the hosts of *Tomato severe rugose virus* (ToSRV), the most important begomovirus species in Brazilian tomato production, although with a low viral titer. Tests of free-choice, adult preference and oviposition indicated that *E. heterophylla* was the most suitable host for *Bemisia tabaci* MEAM1 among seven non-cultivated plants tested (Sottoriva *et al.*, 2014).

Given the importance of *E. heterophylla* as an invasive species, the high preference of the whitefly for this host, the possibility of co-infection with ToSRV (and possibly other crop-infecting begomoviruses), the reports of alphasatellites in association with EuYMV, and the effects of EuYMV in the host's recombination rate (Richter *et al.*, 2016), it is clear that the study of viral populations in this non-cultivated plant can provide valuable clues in terms of the likelihood of these populations to evolve and infect crops. Thus, the objectives of this work were to

assess the genetic variability and population structure of EuYMV in Brazil, and to study its interaction with alphasatellites at the biological and molecular levels.

References

BARRETO, S. S.; HALLWASS, M.; AQUINO, O. M.; INOUE-NAGATA, A. K. A study of weeds as potential inoculum sources for a tomato-infecting begomovirus in central Brazil. **Phytopathology**, v. 103, n. 5, p. 436-444, 2013.

CASTILLO-URQUIZA, G. P.; BESERRA JUNIOR, J. E. A.; ALFENAS-ZERBINI, P.; VARSANI, A.; LIMA, A. T. M.; BARROS, D. R.; ZERBINI, F. M. Genetic diversity of begomoviruses infecting tomato in Paty do Alferes, Rio de Janeiro state, Brazil. **Virus Reviews and Research**, v. 12, p. 233, 2007.

CHRISTOFFOLETI, P. J. **Aspectos da resistência de plantas daninhas a herbicidas**. 3. ed. Piracicaba: HRAC-BR, 2008. 120 p.

COSTA, A. S.; BENNETT, C. W. Whitefly transmitted mosaic of *Euphorbia prunifolia*. **Phytopathology**, v. 40, p. 266-283, 1950.

COTRIM, M. A.; KRAUSE-SAKATE, R.; NARITA, N.; ZERBINI, F. M.; PAVAN, M. A. Genetic diversity ot tomato-infecting begomoviruses in Central São Paulo state (in Portuguese). **Summa Phytopathologica**, v. 33, p. 300-303, 2007.

CRONQUIST, A. **An integrated system of classification of flowering plants.** New York: Columbia University Press, 1981. 1262 p.

DINSDALE, A.; COOK, L.; RIGINOS, C.; BUCKLEY, Y. M.; DE BARRO, P. 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**, v. 103, n. 2, p. 196-208, 2010.

FARIA, J. C.; SOUZA-DIAS, J. A. C.; SLACK, S.; MAXWELL, D. P. A new geminivirus associated with tomato in the State of São Paulo, Brazil. **Plant Disease**, v. 81, p. 423, 1997.

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. Molecular and biological characterization of a new Brazilian begomovirus, euphorbia yellow mosaic virus (EuYMV), infecting *Euphorbia heterophylla* plants. **Archives of Virology**, v. 156, n. 11, p. 2063-2069, 2011.

FERRO, C. G.; SILVA, J. P.; XAVIER, C. A. D.; GODINHO, M. T.; LIMA, A. T. M.; MAR, T. B.; LAU, D.; ZERBINI, F. M. 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**, p. 10.1111/aab.12329, 2017.

GALVÃO, R. M.; FERNANDES, A. V.; ALMEIDA, J. D.; ALFENAS, P. F.; ANDRADE, E. C.; FONTES, E. P. B. Molecular characterization of two new

tomato-infecting geminiviruses and the *Sida*-infecting geminiviruses complex from Brazil. In: INTERNATIONAL WORKSHOP ON BEMISIA AND GEMINIVIRAL DISEASES, 1998, San Juan - Puerto Rico. **Anais.** San Juan - Puerto Rico, 1998. p. L-93.

JESKE, H.; KOBER, S.; SCHÄFER, B.; STROHMEIER, S. Circomics of Cuban geminiviruses reveals the first alpha-satellite DNA in the Caribbean. **Virus Genes**, v. 49, n. 2, p. 312-324, 2014.

JONES, D. R. Plant viruses transmitted by whiteflies. **European Journal of Plant Pathology**, v. 109, n. 3, p. 195-219, 2003.

LAZAROWITZ, S. G. Geminiviruses: Genome structure and gene function. **Critical Reviews in Plant Sciences**, v. 11, p. 327-349, 1992.

MELO, P. C. T. **Mosca branca ameaça produção de hortaliças**. Campinas, SP, Brazil: Asgrow do Brasil Sementes Ltda., Technical Bulletin 1992.

NAVAS-CASTILLO, J.; FIALLO-OLIVÉ, E.; SÁNCHEZ-CAMPOS, S. Emerging virus diseases transmitted by whiteflies. **Annual Review of Phytopathology**, v. 49, p. 219-248, 2011.

PAPROTKA, T.; METZLER, V.; JESKE, H. The first DNA 1-like alphasatellites in association with New World begomoviruses in natural infections. **Virology**, v. 404, n. 2, p. 148-157, 2010.

RIBEIRO, S. G.; BEZERRA, I. C.; LIMA, M. F.; ÁVILA, A. C.; GIORDANO, L. B. Occurrence of geminivirus in tomato plants in Bahia. In: VIII ENCONTRO NACIONAL DE VIROLOGIA (RESUMOS), 1996, São Lourenco, MG. **Anais.** São Lourenco, MG: SBV, 1996. p. 290.

RIBEIRO, S. G.; MELLO, L. V.; BOITEUX, L. S.; KITAJIMA, E. W.; FARIA, J. C. Tomato infection by a geminivirus in the Federal District, Brazil. **Fitopatologia Brasileira**, v. 19, p. 330, 1994.

RICHTER, K. S.; SERRA, H.; WHITE, C. I.; JESKE, H. The recombination mediator RAD51D promotes geminiviral infection. **Virology**, v. 493, p. 113-127, 2016.

ROMAY, G.; CHIRINOS, D.; GERAUD-POUEY, F.; DESBIEZ, C. Association of an atypical alphasatellite with a bipartite New World begomovirus. **Archives of Virology**, v. 155, n. 11, p. 1843-1847, 2010.

ROMAY, G.; LECOQ, H.; DESBIEZ, C. Melon chlorotic mosaic virus and associated alphasatellite from Venezuela: genetic variation and sap transmission of a begomovirus–satellite complex. **Plant Pathology**, v. 64, n. 5, p. 1224-1234, 2015.

SILVA, S. J. C.; CASTILLO-URQUIZA, G. P.; HORA-JUNIOR, B. T.; ASSUNÇÃO, I. P.; LIMA, G. S. A.; PIO-RIBEIRO, G.; MIZUBUTI, E. S. G.; ZERBINI, F. M. Species diversity, phylogeny and genetic variability of begomovirus populations infecting leguminous weeds in northeastern Brazil. **Plant Pathology**, v. 61, p. 457-467, 2012.

SOTTORIVA, L. D. M.; LOURENÇÃO, A. L.; COLOMBO, C. A. Performance of *Bemisia tabaci* (Genn.) biotype B (Hemiptera: Aleyrodidae) on weeds. **Neotropical Entomology**, v. 43, n. 6, p. 574-581, 2014.

TAVARES, S. S.; RAMOS-SOBRINHO, R.; GONZALEZ-AGUILERA, J.; LIMA, G. S. A.; ASSUNÇÃO, I. P.; ZERBINI, F. M. Further molecular characterization of weed-associated begomoviruses in Brazil with an emphasis on *Sida* spp. **Planta Daninha**, v. 30, p. 305-315, 2012.

VARSANI, A.; ROUMAGNAC, P.; FUCHS, M.; NAVAS-CASTILLO, J.; MORIONES, E.; IDRIS, A.; BRIDDON, R. W.; RIVERA-BUSTAMANTE, R. F.; ZERBINI, F. M.; MARTIN, D. P. *Capulavirus* and *Grablovirus*: two new genera in the family *Geminiviridae*. **Archives of Virology**, v. 162, p. doi:10.1007/s00705-00017-03268-00706, 2017.

VIDAL, R. A.; WINKLER, L. M. *Euphorbia heterophylla* L. resistant to herbicide inhibitors of acetolactate synthase: II - Geographic distribution and genetic characterization of biotypes from Rio Grande do Sul plains. **Revista Brasileira de Agrociência**, v. 10, p. 461-465, 2004.

WILSON, A. K. *Euphorbia heterophylla*: A review of distribution, importance and control. **Tropical Pest Management**, v. 27, p. 32-38, 1981.

ZERBINI, F. M.; BRIDDON, R. W.; IDRIS, A.; MARTIN, D. P.; MORIONES, E.; NAVAS-CASTILLO, J.; RIVERA-BUSTAMANTE, R.; VARSANI, A.; ICTV CONSORTIUM. ICTV Virus Taxonomy Profile: *Geminiviridae*. **Journal of General Virology**, v. 98, *in press*, 2017.

ZHOU, X. Advances in understanding begomovirus satellites. **Annual Review of Phytopathology**, v. 51, n. 1, p. 357-381, 2013.

CHAPTER 1

GENETIC VARIABILITY AND POPULATION STRUCTURE OF THE NEW WORLD BEGOMOVIRUS *Euphorbia yellow mosaic virus*

Mar, T.B., Xavier, C.A.D., Lima, A.T.M., Nogueira, A., Silva, J.C.F., Ramos-Sobrinho, R., Lau, D., Zerbini, F.M. Genetic variability and population structure of the new world begomovirus *Euphorbia yellow mosaic virus*. Journal of General Virology, *submitted*.

GENETIC VARIABILITY AND POPULATION STRUCTURE OF THE NEW WORLD BEGOMOVIRUS *Euphorbia yellow mosaic virus*

Talita B. Mar[1,2], César A.D. Xavier[1,2], Alison T. M. Lima[3], Angélica M. Nogueira[1,2], José C.F. Silva[2], Roberto Ramos-Sobrinho[4], Douglas Lau[5], F. Murilo Zerbini[1,2]*

[1]Dep. de Fitopatologia/BIOAGRO and [2]National Research Institute for Plant-Pest Interactions, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil
[3]Instituto de Ciências Agrárias, Universidade Federal de Uberlândia, Uberlândia, MG, 38400-902, Brazil
[4]Centro de Ciências Agrárias/Fitossanidade, Universidade Federal de Alagoas, Rio Largo, AL, 57100-000, Brazil
[5]Embrapa Trigo, Rodovia BR-285, 3081, Passo Fundo, RS, 99001-970, Brazil

*Corresponding author:
Telephone: (+55-31) 3899-2935
E-mail: zerbini@ufv.br

Running title: Genetic variability of *Euphorbia yellow mosaic virus*

Word in abstract: 248
Words in main text: 5,952
Number of tables: 3
Number of figures: 5

The sequences reported in this paper have been deposited in GenBank under access numbers KY559430 - KY559646.

Abstract

The emergence of begomoviruses (whitefly-transmitted viruses classified in the genus *Begomovirus*, family *Geminiviridae*) in Brazil probably occurred by horizontal transfer from non-cultivated plants after the introduction of *Bemisia tabaci* MEAM1. The center of diversity of *Euphorbia heterophylla* (Euphorbiaceae) is located in Brazil and Paraguay, where it is an invasive species in soybean and other crops. Reports of possible begomovirus infection of *E. heterophylla* in Brazil date back to the 1950's. In 2011, *Euphorbia yellow mosaic virus* (EuYMV) was described in symptomatic plants collected in the Brazilian state of Goiás. Here we assess the genetic variability and population structure of begomoviruses infecting *E. heterophylla* in samples collected throughout nine Brazilian states from 2009 to 2014. A total of 158 and 57 haplotypes were compared in DNA-A and DNA-B datasets, respectively. Analysis comparing population structure in a large sampled area enabled to differentiate two subpopulations. However, the application of DAPC allowed the differentiation of six subpopulations according to sampling locations and in agreement with phylogenetic analysis. In general, negative selection was predominant in all six subpopulations. Interestingly, we were able to reconstruct the phylogeny based on the information from 23 sites that mostly contributed to the geographical structure proposed, demonstrating that these polymorphisms hold supporting information to discriminate between subpopulations. These sites were mapped in the genome and compared at the level of amino acid changes, providing insights into how genetic drift and selection contribute to maintain the patterns of begomovirus population variability from a geographical structuring point of view.

Keywords: begomovirus, geminivirus, genetic variability, non-cultivated hosts, population structure

Introduction

The *Geminiviridae* family includes viruses whose genomes are composed of one or two molecules of circular, single-stranded DNA encapsidated by a single structural protein into twinned, quasi-icosahedral particles. The family includes the genera *Begomovirus, Becurtovirus, Capulavirus, Curtovirus, Eragrovirus, Grablovirus*, *Mastrevirus*, *Topocuvirus* and *Turncurtovirus*, based on the type of insect vector, host range, genome organization and phylogenetic relationships [1-4]. Begomoviruses are transmitted by the whitefly *Bemisia tabaci* and infect dicot plants. The genus *Begomovirus* can be divided into "Old World" (OW; Europe, Africa, Asia, and Australia) and "New World" (NW; the Americas) lineages based on genome organization, phylogenetic relationships and geographical distribution [5, 6]. Most NW begomoviruses have two components, named DNA-A and DNA-B. The DNA-A contains genes involved in replication and encapsidation of the viral progeny. The DNA-B contains genes required for intra- and intercellular movement in the plant. Both genomic components are required for systemic infection of the host [6].

Geminivirus populations, including begomoviruses, possess high genetic variability, mostly due to their high nucleotide substitution rates, similar to those estimated for RNA viruses (in the order of 10^{-4} substitutions per site per year) [7, 8], and the frequent occurrence of recombination [9] and pseudorecombination between bipartite viruses [10]. The emergence of begomoviruses in Brazil probably occurred by horizontal transfer of indigenous viruses infecting non-cultivated plants after the introduction of *Bemisia tabaci* Middle East-Asia Minor 1 (MEAM1; previously known as *B. tabaci* biotype B) in the mid-1990's [11-13]. Since the introduction of *B. tabaci* MEAM1, a large number of begomoviruses have been described in association with tomato and with many non-cultivated plants [14-18]. The presence of several viruses transmitted by the same vector to a wide range of hosts facilitates mixed infections, where novel recombinant variants with increased fitness can be formed [19-21]. Evidence of the emergence of tomato-infecting begomoviruses in Brazil, involving both recombination and pseudorecombination processes, has been reported by Silva *et al.* [16].

Euphorbia heterophylla is an invasive species in soybean and other crops in Brazil and Paraguay, its center of origin [22]. Its occurrence is common in the

southern, southeastern and midwestern regions of Brazil [23]. There are reports of *E. heterophylla* herbicide-resistant genotypes in the state of Rio Grande do Sul, where it is present in 74% of soybean producing areas [24, 25]. Reports of possible begomoviruses infecting *E. heterophylla* in Brazil date back to the 1950's [26]. In 2011, *Euphorbia yellow mosaic virus* (EuYMV) was reported in *E. heterophylla* plants collected in the state of Goiás [27]. Later, the virus was found in Brazil in the non-cultivated plant species *Sida santaremnensis* [28], *Macroptilium atropurpureus* [15] and *Crotalaria juncea* [29]. Experimentally, *Arabidopsis thaliana* is also a host of the virus, in which it was demonstrated to enhance recombination rates [30]. Barreto *et al.* [29] demonstrated that *E. heterophylla* is one of the hosts of *Tomato severe rugose virus* (ToSRV), although with a low viral titer. Tests of free-choice, adult preference and oviposition indicated that *E. heterophylla* was the most suitable host for *Bemisia tabaci* MEAM1 among seven non-cultivated plants tested [31].

Several non-cultivated plants have been reported as begomoviruses hosts. These plants may serve as virus reservoirs, from which they can be transmitted to cultivated plants, and also as "mixing vessels" that increase the probability of interspecific recombination and pseudorecombination [16, 29, 32]. Given the high preference of whiteflies for *E. heterophylla*, the possibility of co-infection with ToSRV (and possibly with other crop-infecting begomoviruses), and the effects of EuYMV in the host's recombination rate, the study of EuYMV populations in this non-cultivated plant could provide valuable clues in terms of the probability of it infecting crop plants due its relatively high rate of evolution.

We assessed the begomovirus diversity associated with *E. heterophylla* plants, and estimated the genetic variability of EuYMV populations infecting this host. The genetic structure of EuYMV populations was investigated in a large-scale sampling, with isolates collected in locations throughout Brazil from 2009 to 2014, using Bayesian clustering and multivariate statistical analyses. Also, we highlighted the genome sites that mostly contributed to the geographical structure proposed, the processes that maintained genetic structure, and described how the genetic variability is distributed within EuYMV subpopulations.

Results

Viral detection and recombination analysis. We investigated 165 symptomatic *E. heterophylla* samples, collected over a period of six years in locations throughout Brazil (states of Amazonas, Goiás, Mato Grosso do Sul, Minas Gerais, Paraná, Pernambuco, Rio Grande do Sul and Santa Catarina, and the Federal District). Based on the detection of an 2,600 bp band after digestion of the rolling-circle amplification (RCA) products with restriction enzymes, 142 samples were preliminarily positive for the presence of a begomovirus (data not shown; also, over the years we have tested symptomless samples and found them to be consistently virus-free). Full-length DNA-A and DNA-B components were cloned from 129 and 41 RCA-positive samples, respectively. In this total of 133 samples cloned, EuYMV was the only virus detected (except for one sample from which a virus corresponding to a new species was cloned).

Based on the criteria of <91% nucleotide sequence identity for the DNA-A, recently updated by the *Geminiviridae* Study Group of the ICTV [3], a virus corresponding to a new species was cloned from a sample collected in Itaroquém, RS in 2011. Pairwise sequence comparisons of the DNA-A from isolate BR:Itr962:11 indicated a maximum nucleotide sequence identity of ~81% with *Tomato dwarf leaf virus* (ToDfLV, JN564749), *Tomato mottle wrinkle virus* (ToMoWV, JQ714137) *Melochia mosaic virus* (MelMV, KT201151) and *Tomato leaf deformation virus* (ToLDeV, JX501509, JX501508, JX501503, JX501502) (data not shown). Probably, BR:Itr962:11 emerged by recombination. Strongly supported recombinant events were detected by six methods (RDP, Geneconv, Bootscan, Maxichi, Chimera and 3Seq), involving the isolates BR:Str42:09 and BR:Iju350:09 as major parents and *Tomato severe leaf curl virus* (TSLCV, AF130415) as the minor parent.

Based on BLAST analysis and pairwise sequence comparisons we identified 150 full-length DNA-A (139 haplotypes) and 57 full-length DNA-B (52 haplotypes) components with >96% and >94% identity amongst themselves and with EuYMV sequences retrieved from GenBank (data not shown). EuYMV sequences determined in this study were combined with those retrieved from GenBank for a total of 158 and 57 haplotypes in the DNA-A and DNA-B datasets, respectively (Suppl. Tables S1; Suppl. Table S2).

As expected, the DNA-B dataset was more variable [33], even with a smaller number of isolates. The average number of nucleotide (nt) differences between isolates was 62.8 and 101.8 for the DNA-A and DNA-B, respectively, and the average pairwise number of nt differences was 0.02 and 0.04 for the DNA-A and DNA-B, respectively (Table 1).

No recombination events were detected in the intraspecific DNA-A dataset. In contrast, five recombinant isolates were detected in the intraspecific DNA-B dataset (Suppl. Table S3). Events 1 and 3 involved breakpoints located within the *NSP* and/or *MP* genes, while event 2 presented breakpoints in non-coding regions. Topological differences between trees based on nt sequences of the *NSP* and *MP* genes further supported the recombinant origin of these isolates (Suppl. Fig. S1b). Although recombination events were not assigned to long branches as expected [34], this is probably due to high degree of similarity between recombinants and parental sequences.

Phylogenetic analysis. Bayesian-inferred trees based on nt sequences of the *CP*, *Rep*, *MP* and *NSP* genes and on the full-length DNA-A and DNA-B clearly clustered EuYMV isolates according to sampling locations (Fig. 1; Suppl. Fig. S1). Both DNA-A and DNA-B trees showed two major clades supporting the existence of different subpopulations.

The DNA-A tree showed strong support for isolates from AM, GO, MG, PB and PE (cluster I) comprising a separate subpopulation from those of PR, RS and SC (cluster II). Isolates from MS were considered to be in cluster I (see below). Cluster I is supported by strong posterior probabilities, except for isolate BR:5:LEA:08 in the DNA-A dataset. This isolate could be a recombinant, since it is related (in a poorly supported clade) to RS isolates in the *Rep* gene phylogeny (Fig. 1; Suppl. Fig. S1). In addition, three well-supported subclusters including isolates from PE, MG and GO indicate further geographical structuring. This was not observed in cluster II, where isolates from PR, RS and SC are mixed in well-supported subclades.

The majority of isolates from MS clustered together in a well-supported clade with isolate BR:Amp1068:11, which could be a migrant. However, BR:Mir1:07 and BR:Pop709:10 are in a branch containing isolates from PB in the DNA-A and *Rep* phylogenies. Observing the incongruence between *Rep* and *CP* phylogenies, these isolates could be recombinants not detected by RDP4. The *Rep*

phylogeny clustered isolates from MS in a well-supported clade with the isolates from PB and with 17 isolates from RS. Interestingly, these same 17 isolates from RS clustered together in the DNA-A tree.

Even with few isolates representing each group, the DNA-B dataset broadly supports the same clusters (Fig. 1; Suppl. Fig. S1), except for isolates BR:Vic54:14 and BR:Sau937:11 which could be migrants and the recombinant BR:Ats3.2:09. In the DNA-B tree, BR:Mir1:07 and BR:Mir2:07 clustered with PR and RS isolates.

Genetic structure and variability of the EuYMV population. The clusters inferred by Bayesian trees suggested the existence of two different subpopulations according to sampling locations. To evaluate spatial processes driving population structure we compared genetic divergence with geographical distances between isolates using Mantel's test. A significant correlation ($r=0.5698$; $p<0.001$) between the number of sites that differ between each pair of sequences with the geographic distance between isolates reinforces the existence of subpopulations assigned to geographical regions (Fig. 2a). Also, both the DNA-A and DNA-B of cluster I isolates displayed slightly larger genetic variability than those of cluster II isolates (Table 1), with significant differences between π values of 0.022 and 0.019 for the DNA-A of cluster I and cluster II isolates, respectively, and 0.036 and 0.033 for the DNA-B of cluster I and cluster II isolates, respectively (Table 1; Fig. 2c). This is consistent with nucleotide diversity calculated on a sliding window across the DNA-A and DNA-B (Fig. 1). DNA-B clusters were more variable than DNA-A clusters (Fig. 2c).

Differences in genetic variability between the phylogenetically-inferred clusters supported population subdivision, which was confirmed by Nst statistics (Table 2). AMOVA analysis attributed 39% and 27% of total variation in the DNA-A and DNA-B datasets, respectively, to the variation between cluster I and cluster II isolates (Table 3). To exclude the possibility of a bias due to the different years when samples were collected in each region, AMOVA analysis attributing collection year as a component of subdivision was also performed, with the results indicating that this parameter does not explain the variation among the isolates (Suppl. Table S4).

A Bayesian clustering method was also applied to infer distinct subpopulations. Due to the low evidence of recombination, we tested the degree

of linkage equilibrium. Although we found evidence of significant linkage disequilibrium (LD; $p<0.001$), the corresponding standardized indexes of association (I^s_A) were 0.0147 and 0.0158 considering only segregating sites, and 0.0045 and 0.0051 considering the whole genome for the DNA-A and DNA-B datasets, respectively. These values were lower than those estimated for begomovirus datasets (0.0367) considered to be effectively in linkage equilibrium [35].

The results of STRUCTURE were consistent with phylogenetic analysis, and the EuYMV population was estimated to be composed of two subpopulations (clusters I and II; Suppl. Fig. S2). In general, the genetic composition of each DNA-A component assigned into a cluster was not admixed, reflecting the absence of intraspecific recombination according to RDP analysis (Suppl. Fig. S2). We detected only 27 out of 158 isolates with more than 10% of admixing in their genetic composition. Seven of them, located in cluster I, were BR:5:LEA:08, the isolate from AM and six isolates from MS. The MS isolates showed approximately 67% of genetic composition from cluster I, except for BR:Mir1:07 and BR:Pop709:10 which showed 99% and 86% of genetic composition from cluster I, respectively. Therefore, we considered these individuals to be members of cluster I instead of cluster II (as suggested by phylogenetic analysis). A similar pattern of admixed individuals from MS was identified in the DNA-B dataset (plus one sequence from MG; Suppl. Fig. S2). The other 19 DNA-A admixed isolates, located in cluster II, include BR:Amp1068:11 and BR:Nom682:10, plus the same 17 isolates from RS which were located in a well-supported clade with MS and PB isolates in the *Rep* phylogeny.

In the DNA-B dataset, nine admixed isolates were identified in cluster II: BR:Ara1133:11, the recombinant BR:Ats3.2:09 and seven individuals from PR and RS which were clustered with MS isolates in the DNA-B phylogenetic tree. These results point to gene flow (between MS/PB and RS/SC/PR isolates) and recombination contributing to increased genetic variability in these subpopulations.

Comparing the phylogenetic analysis and STRUCTURE results, we were able to explain the population structure based on two main groups. However, it is still remarkable that four well supported subclusters of isolates from PB, MG, GO and MS are present in the DNA-A tree (Fig. 1). To better understand the

population structure and find the most suitable subdivision to explain the data, we performed a Discriminant Analysis of Principal Components (DAPC) in the same datasets. Since DAPC requires prior groups, we inferred genetic clusters using *K*-mean runs with sequentially increasing number of *K*. The higher likelihood was obtained for six subpopulations (*K*=6). The subpopulations were checked for their association with geographical location. A correlation was found between the subpopulations inferred by *K*-means and the geographical distribution of sampling locations (Fig. 3a). Four groups were comprised mainly by individuals from GO, MG, MS and PB, and "group 4" was comprised by individuals from PR, SC and RS. Interestingly, the same 17 admixed isolates from RS were identified as a different subpopulation ("group 5"). Detection of individuals from group 5 increased over the years (one detection in 2009, two in 2010, nine in 2011 and five in 2014), indicating that the population may be expanding. Accordingly, we propose six subpopulations (pop1-6) based on geographical distribution (Fig. 3d), and checked how this fits in comparison with *K*-means (Fig. 3b).

The DAPC ability to discriminate between groups is heavily dependent on the number of principal components (PCs) retained. The trade-off between power of discrimination and over-fitting was measured by repeated DAPC analysis using randomized groups, computing a-scores for each group (Suppl. Fig. S3c). For cross-validation analysis the dataset was divided into two sets: a training set and a validation set. DAPC was carried out in the training set with different numbers of PCs retained, and the ability to predict the membership of the validation set was used to confirm the number of PCs to be retained (Suppl. Fig. S3d). We also checked the percentage of successful reassignment after randomization. Retaining 12 PCs allowed for the discrimination between the six proposed subpopulations (Fig. 3c). Given that the number of investigated clusters is relatively low, all the discriminant functions (DA) were retained. DAPC classification was consistent with the proposed subpopulations, except for two discrepant isolates: BR:Amp1068:11 and BR:Vic9:10 were initially placed into pop1 and pop4, while DAPC classification assigned them to pop3 and pop6, respectively (Suppl. Fig. S3a). We also checked the proper assignment of each isolate into a subpopulation by spatial analysis of molecular variance (SAMOVA), which considered populations that are geographically homogeneous and maximally differentiated

from each other. SAMOVA attributed BR:Man2:13 to a unique subpopulation and pop1 and pop2 to the same group (Suppl. Table S5).

The DAPC results were consistent with Nst population differentiation statistics (Table 2), and indicated genetic differentiation amongst isolates sampled from the geographical locations pointed in Fig. 3d. The correlation between Nst and geographic distance among subpopulations was also significant (r=0.7619; p<0.008) according to Mantel's test (Fig. 2b). Differences in genetic variability also support the DAPC subdivision hypothesis (Table 1; Fig. 2d). AMOVA analysis attributed 57% of total variation to the variation among subpopulations (Table 3). Taking into account the number of isolates, pop3 is the most variable (seven isolates with π=0.021) (Table 1). Hierarchical AMOVA, considering STRUCTURE and DAPC results, attributed 19% of total variation to the variation between cluster I and II isolates, and 41% to the variation among the six subpopulations (Suppl. Table S4).

It was not possible to extrapolate the same groups for the DNA-B dataset, except for pop5 (Fig. 3d), probably because there were not enough isolates representing each subpopulation. Four subpopulations explained the DNA-B dataset, and they reflected exactly the four main clades in the phylogenetic tree presented in Fig. 1 (data not shown).

Amino acid sites under selection. The kind of selection acting on the coding regions of each subpopulation was investigated by neutrality tests. All values were negative, and statistically significant deviations from neutrality were observed for the different genes, except for pop2 and pop3 (Suppl. Table S6). These results indicate purifying selection acting on each population, or a recent population expansion. In agreement with previous studies [32, 34], all datasets showed d_N/d_S ratios lower than 1, except for the *Trap* gene in pop6, again indicating the predominance of purifying (negative) selection in all subpopulations. In fact, for all genes in all subpopulations, a higher number of sites under negative selection was detected (Suppl. Table S7).

A wide variation of the d_N/d_S ratios for each gene/population indicated different selective constraints (Table 4). The d_N/d_S values for the *Trap* gene in all subpopulations and for the *CP* gene in pop6 were considerably high, but even with a reduced number of negatively selected sites, the proportion of individual sites detected to be under negative selection was higher than those under positive

selection (Suppl. Table S7). Interestingly, we found evidence of positive selection for the *Trap* gene in pop1 by PARRIS, and one positively selected site (codon 73) was detected by SLAC, FEL and IFEL. Also, an evidence of positive selection was detected in *Rep* sites (codons 39 and 54 in pop5 and pop1, respectively) by four methods (SLAC, FEL, IFEL and REL) (data not shown). Even with evidence of positive selection in these three sites, negative selection showed to be predominant in all subpopulations.

Sites contributing to genetic divergence among subpopulations. We aimed to assess the 57% of variability between subpopulations (Table 3) identifying genome sites that most contribute to the geographical structure defined by DAPC analysis (Fig. 3). Sites that presented substantial difference across subpopulations largely contribute to the discrimination between subpopulations [36]. A plot with each site's contribution was used to assess sites that best discriminate the proposed subpopulations, allowing us to point out the coding regions that drive genetic divergence among isolates from different sampling locations. A total of 23 sites most reflected the geographical subdivision considering all discriminant functions (Fig. 4a). Different sites were identified by each discriminant function (DF), except sites #1254, #1259 (identified by DFs 3 and 5) and #1969 (identified by DFs 1 and 2). Since each DF corresponds to distinct levels of differentiation between each subpopulation, we consider all contributing sites in the analysis. Interestingly, we were able to discriminate between subpopulations based solely on the information from the most contributing sites (Fig. 5). The information from these polymorphisms could be important to explain the genetic differentiation between subpopulations, and may contain valuable clues to understand the main evolutionary mechanisms (selection and genetic drift) underlying the differentiation.

There were 12 sites covering the CP (#381, 507, 624, 792, 834 and 879), Ren (#1254), Trap (#1309) and Rep (#1670, 1697, 1940 and 1997) coding regions displaying synonymous substitutions (Suppl. Table S8) with different allele frequencies in each sampling location (Fig. 4b). For these sites the negative selection maintained the same amino acid being translated in all subpopulations (Suppl. Table S7). However, we found different alleles either fixed in the third position of the codon or changing in frequency, yielding random fluctuations in the number of variants present in each subpopulation (Fig. 4b). For example,

codon 208 encodes glutamic acid, but in the third position (site #834) the allele G was fixed in all subpopulations, except in pop3 which presented allele A (Fig. 4b). The mechanism underlying these changes could be genetic drift.

Sites #1254 and #1309, located in the overlapping *Trap* and *Ren* genes, were particularly informative. For example, #1254 displays a non-synonymous substitution under positive selection corresponding to codon 80 in *Trap*, and a synonymous substitution encoding valine corresponding to codon 34 of *Ren* (Suppl. Table S8). Two possible residues can be translated in the *Trap* gene, hydrophobic glycine (allele C) or positively charged arginine (alleles G or T), with the predominance of arginine in pop3 and pop4, an overlap of both amino acids in pop5, and glycine in pop1, pop2 and pop6 (Fig. 4b; Fig. 5).

The polymorphisms lead to non-synonymous substitutions in the remaining 11 informative sites (Suppl. Table S8). Polymorphic sites #1056, 2007, 2037 and 2244 display non-synonymous substitutions but involve exchanges between amino acids with the same biochemical properties. For example, codon 149 (site #2037 in *Rep*) encodes positively charged amino acids, with a predominance of arginine (allele T) in all subpopulations except pop1, where lysine (allele C) was predominant (Fig. 4b; Fig. 5; Suppl. Table S8). Also, strong evidence for negative selection in codon 149 was found in pop1 and pop5, and weak evidence for positive selection in pop4 (where the frequency of allele C was >1).

However, the substitutions in sites #1259, 1718, 1943, 1945, 1967, 1968 and 1969 leads to exchanges between different amino acid types (Fig. 4b; Suppl. Table S8). For example, polymorphisms in sites #1967, 1968 and 1969 corresponding to codon 172 in *Rep* yielded translation of either hydrophilic non-charged (metionine or histidine) or hydrophobic (leucine or isoleucine) amino acids. Histidine, methionine and isoleucine were prevalent in pop1, pop2 and pop5, respectively. A balance between the frequencies of methionine and histidine (five isolates encoding leucine and two encoding metionine) was found in pop3, while leucine was predominant in pop4 and pop6 (Fig. 4b; Fig. 5). Translation of amino acids from different classes in different subpopulations suggests that different protein features are being selected at each sampling location, contributing to the genetic divergence among isolates from different subpopulations.

Discussion

The introduction of *B. tabaci* MEAM1 in Brazil has facilitated the transfer of indigenous begomoviruses from non-cultivated plants to economically important crops, something that has been surveyed extensively [14-16, 28, 29, 32, 37, 38]. While most studies have focused on viral species diversity, others have investigated the genetic structure of begomovirus populations infecting both cultivated and non-cultivated hosts in different geographical regions of the country [15, 16, 32, 34, 39]. Nevertheless, very few studies have investigated the genetic variability of viruses that remain restricted to non-cultivated hosts. The knowledge of population dynamics of these viruses is important since they may act as reservoirs of virus diversity, where intraspecific recombination and pseudo-recombination may occur [16, 29, 32].

Since the first report of EuYMV infecting *E. heterophylla* in Brazil [27], the virus has been described in other non-cultivated plants [15, 28, 29]. So far, only Barreto *et al.* [29] found a begomovirus other than EuYMV infecting *E. heterophylla* plants (ToSRV), and even then in a mixed infection with EuYMV and accumulating at a low titer. Likewise, in our study, which comprised an extensive survey in sampling sites covering states from the south to the north of Brazil, EuYMV was the only begomovirus found.

Previous studies found a high level of genetic variability in begomovirus populations infecting non-cultivated hosts. Compared with *Macroptilium yellow spot virus* (MaYSV; π=0.06580 for the full-length of DNA-A), EuYMV has a low degree of variability [34, 39]. We found that the EuYMV DNA-B showed higher genetic variability comparing with the DNA-A, in agreement with previous studies which suggested that bipartite begomoviruses components have distinct evolutionary histories [33].

A number of previous studies have indicated that EuYMV belongs to a phylogenetic lineage distinct from other begomoviruses isolated in Brazil. Silva *et al.* [15] isolated EuYMV from *Macroptilium atropurpureum*, and found EuYMV clustering with viruses from Central America. Tavares *et al.* [28] isolated EuYMV from *Sida santaremnensis*, and indicated that EuYMV is related to a group comprised mostly by begomovirus found in Central and North America.

Similarly, Fernandes *et al.* [27] isolated EuYMV from *E. heterophylla*, and described EuYMV isolates as more closely related to Peruvian and North American begomoviruses than to Brazilian begomoviruses, hypothesizing that EuYMV was introduced into Brazil. These authors placed EuYMV in the so called *Squash leaf curl virus* (SqLCV) clade [27]. Some species of this clade have distinct but overlapping host ranges, and Idris *et al.* [40] demonstrated that members of the SqLCV clade can form viable pseudorecombinants able to extend their host ranges. Rocha *et al.* [32] indicated the clustering of EuYMV with Central and North American begomovirus, and also pointed to a recombinant origin of EuYMV, detecting an interspecific recombination event with a major parent from Central America. In our DNA-A dataset, which is comprised of EuYMV sequences only, no recombination events were detected. Thus, although recombinantion seems to be involved in the origin of EuYMV, it appears that no intraspecies recombination has occurred in the DNA-A once the viral population established itself in Brazil. However, it must be pointed out that the low degree of genetic variability amongst EuYMV isolates makes it difficult to distinguish between recombinants and parental sequences.

In contrast to the DNA-A dataset, five recombinant isolates were detected in the DNA-B dataset, with two events involving breakpoints located within coding regions, which are considered to be "cold spots" for recombination among begomoviruses [41]. These recombination events could explain the higher variability of the DNA-B compared with the DNA-A.

Information on the intraspecific genetic variability of begomoviruses is relevant to provide clues about virulence and dispersion [15, 32, 35, 39, 42-46]. We assessed the genetic variability of EuYMV, investigated the factors that determine the genetic structure and described how the genetic variability is distributed within the subpopulations. The distribution of variability, generated by mutation or recombination, in the viral populations depends on genetic drift and selection, but it is often difficult to differentiate between the effects of genetic drift and those of selection [43]. Because EuYMV is essentially the only virus found in *E. heterophylla* and is only rarely found in other hosts, it is a good viral system to study the effects of genetic drift and selection in the population structure, since it suffers less influence of adaptive selection to different hosts. We tried to highlight how these evolutionary processes could shape the genetic

variability of EuYMV populations, using the information of variability between subpopulations to identify genome sites that most contribute to the geographical structure proposed, and studying the effects of selection considering the contribution of polymorphic sites within coding regions.

A great effort has been applied to define begomovirus population differentiation according to geographical origin. Some studies, using Bayesian clustering approaches implemented in the program STRUCTURE, aimed at identifying population structure at a global [35] or local [32] scale. We provided the first analysis comparing population structure in a large sampled area using two different methods: Bayesian clustering and multivariate statistical analysis. For simulated data, multivariate analysis using DAPC proved as accurate as STRUCTURE [47]. Using STRUCTURE we were able to differentiate two subpopulations based on sampling locations, which corresponded to the two major clades inferred by phylogenetic analysis. However, the application of DAPC seemed to be more suitable for our dataset, since it allowed the differentiation of six subpopulations also according to sampling locations and in agreement with phylogenetic analysis.

Similar to begomoviruses, the cryptic species complex of *B.tabaci* exhibits a strong geographical pattern in a global context [48]. However, studies analyzing the population structure within each species of *B. tabaci* in a local scale failed to find evidence of isolation by distance. Such non-geographical structuring could indicate long distance migrations [49, 50]. Low genetic diversity was observed in NW populations in comparison with OW populations [48]. In Brazil, *B. tabaci* MEAM1 is prevalent but the New World 1 (NW1) and New World 2 (NW2) species have also been reported in different regions [51], and the Mediterranean (MED) species was recently reported in the southern region [52]. The lack of information about the genetic structure of *B. tabaci* populations in Brazil prevented us from comparing the viral population structure in relation to (or as a function of) that of *B. tabaci*.

The genetic variability of pop3 was higher compared with other subpopulations defined by DAPC. Pop3 comprised mainly individuals from MS, where the agricultural landscape is composed by a great diversity of hosts. Environmental heterogeneity could also modulate the genetic structure of both host and virus. The effect of landscape heterogeneity on the genetic structure of

viral populations was demonstrated for two bipartite begomoviruses, *Pepper golden mosaic* (PepGMV) and *Pepper huasteco yellow vein virus* (PHYVV) comparing neighbouring populations of wild and human-managed chiltepin (*Capsicum annuum glabriusculum*), a host with a large degree of genetic variability and strong spatial structure [53]. The prevalence of each virus was significantly higher in cultivated compared to wild populations. Nevertheless, the effect of ecosystem biodiversity on the genetic diversity depended on the virus species. The loss of biodiversity at higher levels of habitat anthropization was associated with increased genetic diversity of PepGMV but not of PHYVV [54]. A possible influence of *E. heterophylla* diversity and agro-climate environment in the EuYMV population structure will be addressed in our future studies.

DAPC analysis allowed the visual assessment of between-population differentiation and of the contribution of individual alleles to population structuring. Genome sites that were most markedly different across the proposed subpopulations were highlighted as most contributing to genetic divergence. These sites were mapped in the genome and compared at the level of amino acid changes. We found each gene/subpopulation under different selective pressures, albeit with a tendency of purifying selection acting upon each subpopulation. Many more sites were evolving under negative selection than sites evolving under positive selection, in agreement with previous studies that demonstrate negative selection predominating in coding regions during the evolution of plant viruses [32, 34, 42, 43]. In general, negative selection was predominant in all six subpopulations.

We were able to reconstruct the phylogenetic analysis using only the most contribution sites, demonstrating that these polymorphisms hold supporting information to discriminate between subpopulations and may contain valuable clues to understand the main evolutionary mechanisms underlying genetic differentiation. Considering the predominance of negative selection acting on the EuYMV populations, the polymorphisms were ranked according to three descriptions: synonymous substitutions, non-synonymous substitutions with the same type of amino acid, and non-synonymous substitutions with a different type of amino acid. The mechanism underlying substitutions was most likely genetic drift, with different alleles being fixed or changing in frequency, yielding random fluctuations in the number of variants present in each subpopulation. On the other

hand, different types of amino acids being translated in each subpopulation could be evidence of distinct features being selected at each sampling location, contributing to the genetic divergence among isolates from different subpopulations. The importance of these polymorphisms in the adaptability of each isolate to different environmental conditions could be assessed by directional mutation in infectious clones.

In conclusion, EuYMV has a lower degree of genetic variability compared with other begomovirus populations infecting non-cultivated plants, and only a few intraspecific recombination events (restricted to the DNA-B dataset) were detected. EuYMV displays a different pattern from that commonly observed in begomovirus populations (high variability and frequent recombination events), but nevertheless, similar to other begomoviruses, segregates according to sampling location. It could be a good system to study the standing genetic variability of begomovirus populations and how genetic drift and selection contribute to maintain the patterns of begomovirus population diversity from a geographical structuring perspective.

Methods

Sampling. Samples of *E. heterophylla* plants showing typical symptoms of begomovirus infection (yellow mosaic, leaf curling, and stunting) were collected in locations throughout the states of Amazonas (AM), Goiás (GO), Mato Grosso do Sul (MS), Minas Gerais (MG), Paraíba (PB), Pernambuco (PE), Paraná (PR), Rio Grande do Sul (RS) and Santa Catarina (SC) from 2009 to 2014 (Suppl. Tables S1, S2). Samples were collected in crop fields (mostly common bean, maize, soybean and wheat) and also in natural wild habitats. For each sample the following information was recorded: date of collection, GPS coordinates of sampling location and symptoms (description and digital image of the sample at the time of collection). The samples were stored in plastic bags and transported to the laboratory where they were press-mounted until DNA extraction [55].

Cloning and sequencing of full-length begomovirus genomes. Viral genomes were amplified using rolling-circle amplification (RCA) according to Inoue-Nagata *et al.* [56]. Aliquots of the amplification products were subjected to cleavage with restriction enzymes to obtain fragments of approx. 2,600

nucleotides (nt), corresponding to one genomic copy of each DNA component. These fragments were cloned into the pBLUESCRIPT-KS+ (Stratagene) plasmid vector and completely sequenced at Macrogen, Inc. (Seoul, South Korea). Species assignment was based on the ICTV-established cut-off of 91% nt sequence identity for the full-length DNA-A [3] using pairwise comparisons performed with SDT v. 1.2 [57].

Recombination and phylogenetic analysis. Multiple sequence alignments were performed using the MUSCLE algorithm [58]. Partitions in viral genomes with conflicting evolutionary histories were detected using RDP, Geneconv, Bootscan, Maximum Chi Square, Chimaera, SisterScan and 3Seq methods as implemented in RDP4 program [59], using default settings for each analysis method and a Bonferroni-corrected *p* value of 0.05. Only recombination events detected by at least four methods were considered reliable. Recombinant blocks were eliminated from the data sets. Bayesian analysis was performed with MrBayes 3.0 [60] using the models selected by MrModeltest2.2 [61] in the Akaike Information Criterion (AIC). Two independent analyses were conducted, each running at least 10,000,000 generations. Phylogenetic trees were visualized and edited using FigTree 1.3 (tree.bio.ed.ac.uk/software/ figtree).

Genetic structure of the viral population. The main descriptors of molecular variability were estimated using DnaSP v. 5 [62]. The average pairwise number of nucleotide differences per site (nucleotide diversity, π) was estimated using a sliding window of 100 nt, with a step size of 10 nt. The statistical significance of the differences amongst the mean nucleotide diversity obtained from different datasets was calculated according to Lima *et al.* [63]. The correlation between distance matrices was tested by Mantel's test using vegan package in R software [64]. Two methods were used to infer population subdivision and to assign individuals to each subpopulation. To support the assumption of linkage equilibrium for STRUCTURE analysis, a null hypothesis of linkage equilibrium was tested by Monte Carlo simulations using the program LIAN v. 3.7 [65] as proposed by Prasanna *et al.* [35]. Bayesian clustering using STRUCTURE [66] was performed in ten independent runs of 1,000,000 Markov chain Monte Carlo (MCMC) replications with a *burn-in* of 10,000 runs for each *k* value varying from 1 to 10. The suitable *k* values were determined by the higher likelihood of [Ln P (D)] [67]. For a better visualization of the genetic structure,

the ancestry coefficients were plotted in a map generated in Philcarto software [68]. Multivariate statistical analysis using Discriminant Analysis of Principal Components (DAPC) was performed with the package adegenet implemented in R software [47]. Preliminary subpopulations were inferred by *k*-means run sequentially with increasing values of *k,* retaining all principal components (PCs), and different clustering solutions were compared using Bayesian Information Criterion (BIC). After assigning individuals to each inferred subpopulation, the optimum PC values were investigated to assess the percentage of successful reassignment, a-scores and cross-validation. Wright's F fixation index [69] was estimated for the distinct inferred subpopulations with 10,000 replications in DnaSP v. 5 [62] to support each inferred subpopulation. In addition, different population subdivision hypotheses were tested by analysis of molecular variance (AMOVA) implement in Arlequin v. 3.5 [70] and the proper assignment of each isolate into a subpopulation were checked by spatial analysis of molecular variance (SAMOVA), defining populations that are geographically homogeneous and maximally differentiated from each other [71]. To provide insights into the underlying causes of maintenance of begomovirus geographical structuring we inspected the associated site loadings. As proposed by Jombart [47], the contributions of sites were plotted for a single DAPC discriminant function at a time. We scattered the density of individuals in each subpopulation, and computed the contribution of sites given a threshold of 0.01. Assuming only sites covering coding regions, the frequencies of the 23 most contributing sites were plotted according to geographical structuring, mapped in the genome and compared at the level of amino acid changes. We also reconstructed a Bayesian-inferred tree based on variable amino acids of the most contributing sites.

Detection of negative and positive selection. Three types of neutrality tests were used to test for the occurrence of selection in populations: Tajima's D, Fu and Li's D* and F*. The analyses were performed with DnaSP v. 5 [62]. Amino acid sites under selection were investigated with the maximum likelihood based-methods available in the DataMonkey webserver [72], after determining nt substitution models in MODELTEST and screening for recombination using GARD [73]. The SLAC method was used to estimate the mean ratios of non-synonymous to synonymous substitutions (d_N/d_S) for each open reading frame in the DNA-A components of each EuYMV subpopulation.

Funding information

This work was funded by CNPq (grant 401838/2013-7 to FMZ), FAPEMIG (grant CAG-APQ-02037-13 to FMZ) and Embrapa (grant SEG 02.08.01.008.00.00). TBM was the recipient of a CNPq doctoral fellowship.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Acknowledgements

The authors wish to thank Jesus Navas-Castillo, Elvira Fiallo-Olivé and Eduardo S.G. Mizubuti for helpful discussions and critical reading of the manuscript.

Ethical statement

The research reported in this paper did not involve animals or humans.

References

1. Brown JK, Fauquet CM, Briddon RW, Zerbini FM, Moriones E, Navas-Castillo J. Family *Geminiviridae*. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, editors. Virus Taxonomy Ninth Report of the International Committee on Taxonomy of Viruses. London, UK: Elsevier Academic Press; 2012. p. 351-73.
2. Varsani A, Navas-Castillo J, Moriones E, Hernández-Zepeda C, Idris A, Brown JK, et al. Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. Arch Virol. 2014;159:2193-203.
3. Brown JK, Zerbini FM, Navas-Castillo J, Moriones E, Ramos-Sobrinho R, Silva JC, et al. Revision of *Begomovirus* taxonomy based on pairwise sequence comparisons. Arch Virol. 2015;160(6):1593-619.
4. Varsani A, Roumagnac P, Fuchs M, Navas-Castillo J, Moriones E, Idris A, et al. *Capulavirus* and *Grablovirus*: two new genera in the family *Geminiviridae*. Arch Virol. 2017;162:doi:10.1007/s00705-017-3268-6.

5. Harrison BD, Robinson DJ. Natural genomic and antigenic variation in white-fly transmitted geminiviruses (begomoviruses). Annu Rev Phytopathol. 1999;39:369-98.
6. Rojas MR, Hagen C, Lucas WJ, Gilbertson RL. Exploiting chinks in the plant's armor: Evolution and emergence of geminiviruses. Annu Rev Phytopathol. 2005;43:361-94.
7. Duffy S, Holmes EC. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus *Tomato yellow leaf curl virus*. J Virol. 2008;82(2):957-65.
8. Duffy S, Holmes EC. Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. J Gen Virol. 2009;90:1539-47.
9. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. Virology. 1999;265:218-24.
10. 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*. J Gen Virol. 2006;87:3687-96.
11. Jones DR. Plant viruses transmitted by whiteflies. Eur J Plant Pathol. 2003;109(3):195-219.
12. Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P. Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. Ann Entomol Soc Am. 2010;103(2):196-208.
13. Navas-Castillo J, Fiallo-Olivé E, Sánchez-Campos S. Emerging virus diseases transmitted by whiteflies. Annu Rev Phytopathol. 2011;49:219-48.
14. Castillo-Urquiza GP, Beserra Jr. JEA, Bruckner FP, Lima ATM, Varsani A, Alfenas-Zerbini P, et al. Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil. Arch Virol. 2008;153:1985-9.
15. Silva SJC, Castillo-Urquiza GP, Hora-Junior BT, Assunção IP, Lima GSA, Pio-Ribeiro G, et al. Species diversity, phylogeny and genetic variability

of begomovirus populations infecting leguminous weeds in northeastern Brazil. Plant Pathol. 2012;61:457-67.

16. Silva FN, Lima ATM, Rocha CS, Castillo-Urquiza GP, Alves 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. Virol J. 2014;11:66.
17. Pinto VB, Silva JP, Fiallo-Olivé E, Navas-Castillo J, Zerbini FM. Novel begomoviruses recovered from *Pavonia* sp. in Brazil. Arch Virol. 2016;161:735-9.
18. Fiallo-Olivé E, Zerbini FM, Navas-Castillo J. Complete nucleotide sequences of two new begomoviruses infecting the wild malvaceous plant *Melochia* sp. in Brazil. Arch Virol. 2015;160(12):3161-4.
19. Awadalla P. The evolutionary genomics of pathogen recombination. Nat Rev Genet. 2003;4(1):50-60.
20. Sattar MN, Kvarnheden A, Saeed M, Briddon RW. Cotton leaf curl disease - an emerging threat to cotton production worldwide. J Gen Virol. 2013;94(Pt 4):695-710.
21. Lefeuvre P, Moriones E. Recombination as a motor of host switches and virus emergence: Geminiviruses as case studies. Curr Opin Virol. 2015;10:14-9.
22. Wilson AK. *Euphorbia heterophylla*: A review of distribution, importance and control. Trop Pest Manag. 1981;27:32-8.
23. Cronquist A. An integrated system of classification of flowering plants. New York: Columbia University Press; 1981. 1262 p.
24. Vidal RA, Winkler LM. *Euphorbia heterophylla* L. resistant to herbicide inhibitors of acetolactate synthase: II - Geographic distribution and genetic characterization of biotypes from Rio Grande do Sul plains. Rev Bras Agroc. 2004;10:461-5.
25. Christoffoleti PJ. Aspectos da resistência de plantas daninhas a herbicidas. 3 ed. Piracicaba: HRAC-BR; 2008. 120 p.
26. Costa AS, Bennett CW. Whitefly transmitted mosaic of *Euphorbia prunifolia*. Phytopathology. 1950;40:266-83.

27. Fernandes FR, Albuquerque LC, Oliveira CL, Cruz ARR, Rocha WB, Pereira TG, et al. Molecular and biological characterization of a new Brazilian begomovirus, euphorbia yellow mosaic virus (EuYMV), infecting *Euphorbia heterophylla* plants. Arch Virol. 2011;156(11):2063-9.

28. 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 Dan. 2012;30:305-15.

29. 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(5):436-44.

30. Richter KS, Ende L, Jeske H. Rad54 is not essential for any geminiviral replication mode in planta. Plant Mol Biol. 2015;87(1-2):193-202.

31. Sottoriva LDM, Lourenção AL, Colombo CA. Performance of *Bemisia tabaci* (Genn.) biotype B (Hemiptera: Aleyrodidae) on weeds. Neotrop Entomol. 2014;43(6):574-81.

32. Rocha CS, Castillo-Urquiza GP, Lima ATM, Silva FN, Xavier CAD, Hora-Junior BT, et al. Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. J Virol. 2013;87(10):5784-99.

33. Briddon RW, Patil BL, Bagewadi B, Nawaz-ul-Rehman MS, Fauquet CM. Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. BMC Evol Biol. 2010;10:97.

34. Lima ATM, Sobrinho RR, Gonzalez-Aguilera J, Rocha CS, Silva SJC, Xavier CAD, et al. Synonymous site variation due to recombination explains higher genetic variability in begomovirus populations infecting non-cultivated hosts. J Gen Virol. 2013;94:418-31.

35. Prasanna HC, Sinha DP, Verma A, Singh M, Singh B, Rai M, et al. The population genomics of begomoviruses: Global scale population structure and gene flow. Virol J. 2010;7:220.

36. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010;11(1):94.

37. Albuquerque LC, Varsani A, Fernandes FR, Pinheiro B, Martin DP, Ferreira PTO, et al. Further characterization of tomato-infecting begomoviruses in Brazil. Arch Virol. 2012;157(4):747-52.

38. Fernandes FR, Cruz ARR, Faria JC, Zerbini FM, Aragão FJL. Three distinct begomoviruses associated with soybean in central Brazil. Arch Virol. 2009;154(9):1567-70.

39. Ramos-Sobrinho R, Xavier CAD, Pereira HMB, Lima GSA, Assunção IP, Mizubuti ESG, et al. Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts. J Gen Virol. 2014;95(11):2540-52.

40. Idris AM, Mills-Lujan K, Martin K, Brown JK. Melon chlorotic leaf curl virus: Characterization and differential reassortment with closest relatives reveal adaptive virulence in the Squash leaf curl virus clade and host shifting by the host-restricted Bean calico mosaic virus. J Virol. 2008;82(4):1959-67.

41. Lefeuvre P, Lett JM, Varsani A, Martin DP. Widely conserved recombination patterns among single-stranded DNA viruses. J Virol. 2009;83(6):2697-707.

42. Nouri S, Arevalo R, Falk BW, Groves RL. Genetic structure and molecular variability of *Cucumber mosaic virus* isolates in the United States. PLoS ONE. 2014;9(5):e96582.

43. García-Arenal F, Fraile A, Malpica JM. Variability and genetic structure of plant virus populations. Annu Rev Phytopathol. 2001;39:157-86.

44. González-Aguilera J, Tavares SS, Sobrinho RR, Xavier CAD, Dueñas-Hurtado F, Lara-Rodrigues RM, et al. Genetic structure of a Brazilian population of the begomovirus *Tomato severe rugose virus* (ToSRV). Trop Plant Pathol. 2012;37(5):346-53.

45. Yang XL, Zhou MN, Qian YJ, Xie Y, Zhou XP. Molecular variability and evolution of a natural population of *Tomato yellow leaf curl virus* in Shanghai, China. J Zhejiang Univ-Sc B. 2014;15(2):133-42.

46. Acosta-Leal R, Duffy S, Xiong Z, Hammond RW, Elena SF. Advances in plant virus evolution: Translating evolutionary insights into better disease management. Phytopathology. 2011;101(10):1136-48.

47. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24(11):1403-5.

48. 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(11):e0165105.

49. Dalmon A, Halkett F, Granier M, Delatte H, Peterschmitt M. Genetic structure of the invasive pest *Bemisia tabaci*: evidence of limited but persistent genetic differentiation in glasshouse populations. Heredity. 2007;100(3):316-25.

50. Tahiri A, Halkett F, Granier M, Gueguen G, Peterschmitt M. Evidence of gene flow between sympatric populations of the Middle East-Asia Minor 1 and Mediterranean putative species of *Bemisia tabaci*. Ecol Evol. 2013;3(8):2619-33.

51. Marubayashi JM, Yuki VA, Rocha KCG, Mituti T, Pelegrinotti FM, Ferreira FZ, et al. At least two indigenous species of the *Bemisia tabaci* complex are present in Brazil. J Appl Entomol. 2013;137(1-2):113-21.

52. Barbosa LF, Yuki VA, Marubayashi JM, De Marchi BR, Perini FL, Pavan MA, et al. First report of *Bemisia tabaci* Mediterranean (Q biotype) species in Brazil. Pest Manag Sci. 2015;71(4):501-4.

53. Rodelo-Urrego M, Pagán I, González-Jara P, Betancourt M, Moreno-Letelier A, Ayllón MA, et al. Landscape heterogeneity shapes host-parasite interactions and results in apparent plant-virus codivergence. Mol Ecol. 2013;22(8):2325-40.

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

55. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem Bull. 1987;19:11-5.

56. Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T. A simple method for cloning the complete begomovirus genome using the bacteriophage phi29 DNA polymerase. J Virol Met. 2004;116(2):209-11.

57. Muhire BM, Varsani A, Martin DP. SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLoS ONE. 2014;9:e108277.

58. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinf. 2004;5:1-19.

59. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015;1:vev003.

60. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(1):1572-4.

61. Nylander JAA. MrModeltest v2. Program distributed by the author Evolutionary Biology Centre, Uppsala University2004.

62. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP: DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 2003;19:2496-7.

63. Lima ATM, Silva JCF, Silva FN, Castillo-Urquiza GP, Silva FF, Seah YM, et al. The diversification of begomovirus populations is predominantly driven by mutational dynamics. Virus Evol. 2017;3:in press.

64. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. vegan: Community Ecology Package. R package version 2.0-7. Available at http://CRANR-projectorg/package=vegan. 2013.

65. Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics. 2000;16(9):847-9.

66. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. Mol Ecol Res. 2009;9(5):1322-32.

67. Earl DA, Vonholdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Cons Genet Res. 2012;4(2):359-61.

68. Waniez P. Philcarto: histoire de vie d'un logiciel de cartographie. Eur J Geo. 2010;497.

69. Weir BS. Genetic data analysis II: Methods for discrete population genetic data. Sunderland, Massachusetts: Sinauer Associated Inc; 1996. 445 p.

70. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res. 2010;10(3):564-7.

71. Dupanloup I, Schneider S, Excoffier L. A simulated annealing approach to define the genetic structure of populations. Mol Ecol. 2002;11(12):2571-81.

72. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. 2010;26(19):2455-7.

73. Martin DP, Posada D, Crandall KA, Williamson C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retrov. 2005;21(1):98-102.

Table 1. Genetic variability of *Euphorbia yellow mosaic virus* (EuYMV).

Population	No. of sequences	Genome size (nt)	s*	Eta	k	π	h	Hd
DNA-A								
Phylogeny/ STRUCTURE								
cluster I	68	2595	434	491	56.2	0.022	66	0.99
cluster II	90	2601	517	605	48.3	0.019	89	1.00
DAPC								
pop1	73	2612	452	520	40.8	0.016	73	1.00
pop2	17	2606	159	167	40.5	0.016	16	0.99
pop3	7	2606	144	154	55.7	0.021	7	1.00
pop4	21	2603	161	170	27.4	0.011	19	0.99
pop5	31	2633	246	266	33.2	0.013	31	1.00
pop6	9	2608	48	48	11.0	0.004	9	1.00
TOTAL	158	2589	699	865	62.8	0.024	154	1.00
DNA-B								
cluster I	21	2558	462	528	93.1	0.036	20	1.00
cluster II	36	2554	560	661	84.9	0.033	36	1.00
TOTAL	57	2635	739	925	101.8	0.040	56	0.99

*s, total number of polymorfic segregating sites; Eta, total number of mutations; k, average number of nucleotide differences between sequences; π, nucleotide diversity; h, haplotype number; Hd, haplotype diversity

Table 2. Results of subdivision tests performed on *Euphorbia yellow mosaic virus* (EuYMV) subpopulations.

Population		Nst*
DNA-A		
cluster I	cluster II	0.323
pop1	pop2	0.408
pop1	pop3	0.359
pop1	pop4	0.573
pop1	pop5	0.504
pop1	pop6	0.721
pop2	pop3	0.346
pop2	pop4	0.590
pop2	pop5	0.526
pop2	pop6	0.715
pop3	pop4	0.421
pop3	pop5	0.320
pop3	pop6	0.607
pop4	pop5	0.521
pop4	pop6	0.765
pop5	pop6	0.706
DNA-B		
cluster I	cluster II	0.274

* 0 to 0.05 - little genetic differentiation
0.05 to 0.15 - moderate differentiation
0.15 to 0.25 - great differentiation
> 0.25 high differentiation

Table 3. Analysis of molecular variance (AMOVA) performed on *Euphorbia yellow mosaic virus* (EuYMV) subpopulations.

Analysis	Source of variation	D.f.	Square sum	Variance components	% of variation
DNA-A					
cluster I-II	Between populations	1	1444.006	18.26883 Va	38.85
	Within populations	156	4484.988	28.74992 Vb	61.15
	Total	157	5928.994	47.01875	
	Fst : 0.388				
DNA-B					
cluster I-II	Between populations	1	556.697	19.08365 Va	27.43
	Within populations	55	2776.286	50.47792 Vb	72.57
	Total	56	3332.982	69.56157	
	Fst : 0.274				
DNA-A					
pop1-6	Among populations	5	2999.992	25.75366 Va	57.20
	Within populations	152	2929.002	19.26975 Vb	42.80
	Total	157	5928.994	45.02340	
	Fst : 0.572				

Figure legends

Fig. 1. Midpointed-rooted Bayesian inference trees based on full-length nucleotide sequences of EuYMV DNA-A (**a**) and DNA-B (**b**) components. Nodes to the right of branches with posterior probabilities equal to or higher than 0.8 are indicated by filled circles and those with values lower than 0.8 and higher than 0.5 by empty circles. Isolate colors based on geographical region of origin: AM, Amazonas; GO, Goiás; MG, Minas Gerais; MS, Mato Grosso do Sul; PB, Paraíba; PE, Pernambuco; PR, Paraná; RS, Rio Grande do Sul; SC, Santa Catarina. Populations inferred by STRUCTURE are indicated by black/gray vertical bars. Mean pairwise number of nucleotide differences per site (nucleotide diversity - π) for the DNA-A (**c**) and DNA-B (**d**), calculated on a 10 nucleotide sliding window. A linearized representation of each DNA component, with genes indicated by black bars, is presented at the bottom of each graph.

Fig. 2. Genetic divergence among geographical subpopulations of EuYMV. Mantel test correlating **(a)** genetic distance estimated by the number of sites that differ between each pair of DNA-A sequences with the geographical distance between isolates and **(b)** genetic distance between pop1 to pop6 expressed as pairwise Nst, with the geographical distance between subpopulations estimated by the mean geographical coordinates. Correlation coefficients and significances are indicated with r and p, respectively. Statistical significance of the differences amongst the average pairwise number of nucleotide differences per site (nucleotide diversity, π) calculated for **(c)** DNA-A and DNA-B clusters inferred by Bayesian trees and **(d)** subpopulations inferred by DAPC. Confidence intervals which include the value "zero" denote no statistically significant difference between the means.

Fig. 3. Multivariate statistical clustering analysis of population subdivision using Discriminant Analysis of Principal Components (DAPC) for EuYMV DNA-A. **(a)** Comparison between groups inferred by *K*-means (columns) and sampling locations (rows: AM, Amazonas; GO, Goiás; MG, Minas Gerais; MS, Mato Grosso do Sul; PB, Paraíba; PE, Pernambuco; PR, Paraná; RS, Rio Grande do Sul; SC, Santa Catarina.); **(b)** Comparison between groups inferred by *K*-means

(columns) and geographical subdivision (rows); **(c)** Percentage of successful reassignment after randomization retaining 12 principal components; **(d)** DAPC scatterplot with ellipses representing subpopulations. Dots represent each isolate, colored based on the geographical region located on the map and in the DNA-A (larger) and DNA-B (smaller) phylogenetic trees.

Fig. 4. Contribution of EuYMV DNA-A sites to the genetic divergence among geographical subpopulations. **(a)** Densities of isolates shown colored based on geographical subpopulation and the most contributing sites, given a threshold of 0.01 in each discriminant function of DAPC. The height of each bar is proportional to the contribution of the corresponding site. Sites above the threshold located inside the coding regions shown in red. Site positions in each protein are summarized in the table. **(b)** Changes in frequencies of the most contributing sites that better reflect the geographical structure in each subpopulation (pop1 to pop6).

Fig. 5. Bayesian inference tree based on the most contributing sites of EuYMV DNA-A. Nodes to the right of branches with posterior probabilities equal to or higher than 0.8 are indicated by filled circles and those with values lower than 0.8 and higher 0.5 by empty circles. Isolate colors based on DAPC-inferred subpopulations. **(a)** Tree based on codon sequences of the most contributing sites. **(b)** Tree based on variable sites of amino acid sequences of the most contributing sites, with branches colored according to amino acid sequences. Amino acids correspond to (amino acid site/protein): 1, 223/CP; 2, 80/Rep; 3, 149/Rep; 4, 160/Rep; 5, 172/Rep; 6, 180/Rep; 7, 255/Rep; 8, 16/Ren; 9, 78/Trap; 10, 33/Ren; 11, 80/Trap; 12, 100/Ren. Small letters correspond to amino acid that occurs in less than three sequences, and stripes correspond to amino acids which occur with similar frequencies.

Figure 1

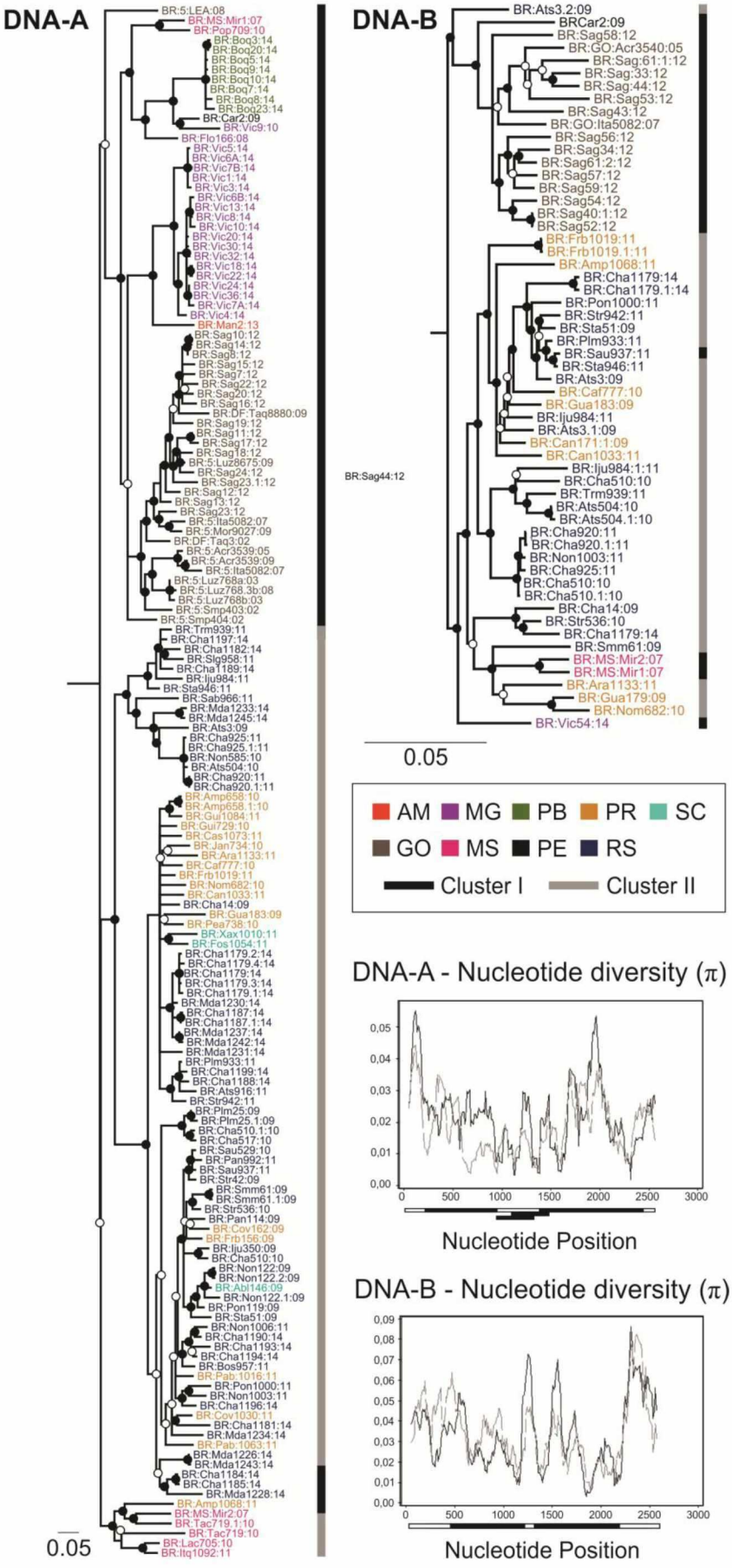

DNA-A
DNA-B
0.05
AM MG PB PR SC
GO MS PE RS
Cluster I Cluster II
DNA-A - Nucleotide diversity (π)
Nucleotide Position
DNA-B - Nucleotide diversity (π)
Nucleotide Position

Figure 2

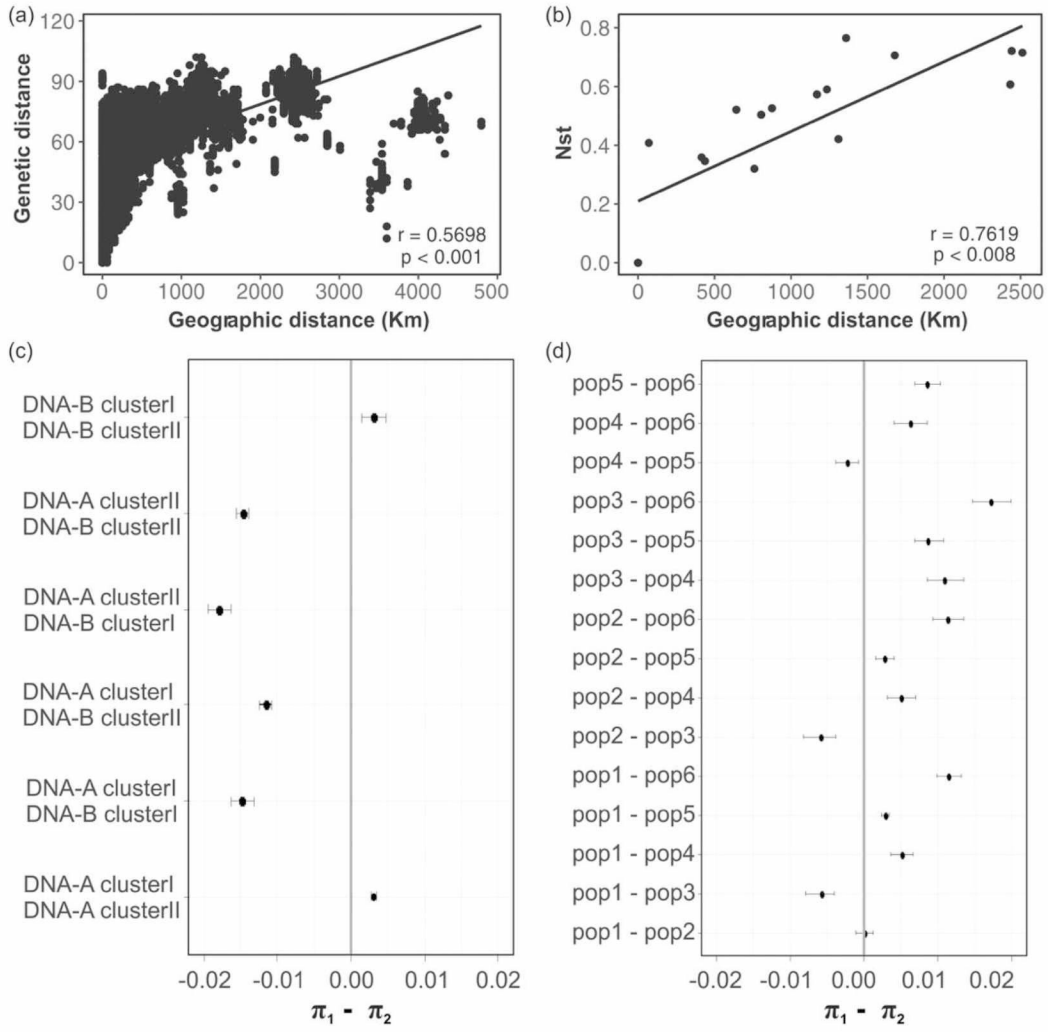

(a)
Genetic distance
120
90
60
30
0
0
1000
2000
3000
4000
500
r = 0.5698
p < 0.001
Geographic distance (Km)
(b)
Nst
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
r = 0.7619
p < 0.008
Geographic distance (Km)
(c)
DNA-B clusterI
DNA-B clusterII
DNA-A clusterII
DNA-B clusterII
DNA-A clusterII
DNA-B clusterI
DNA-A clusterI
DNA-B clusterII
DNA-A clusterI
DNA-B clusterI
DNA-A clusterI
DNA-A clusterII
-0.02
-0.01
0.00
0.01
0.02
π1 - π2
(d)
pop5 - pop6
pop4 - pop6
pop4 - pop5
pop3 - pop6
pop3 - pop5
pop3 - pop4
pop2 - pop6
pop2 - pop5
pop2 - pop4
pop2 - pop3
pop1 - pop6
pop1 - pop5
pop1 - pop4
pop1 - pop3
pop1 - pop2
-0.02
-0.01
0.00
0.01
0.02
π1 - π2

Figure 3

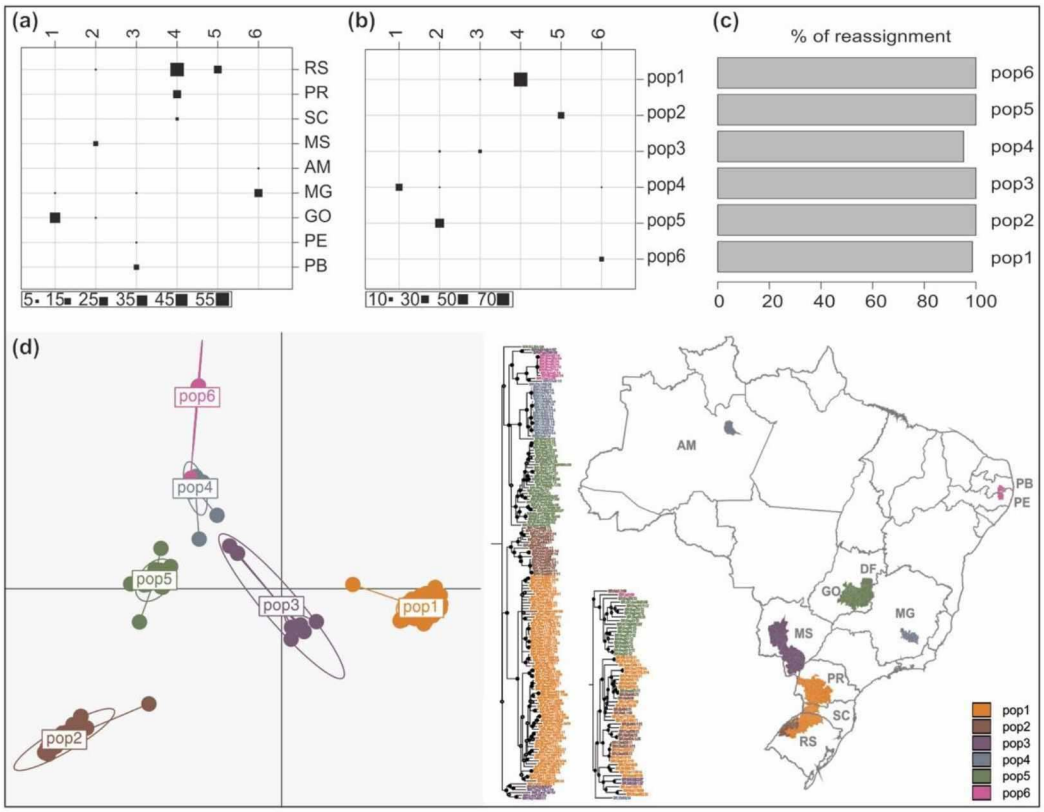
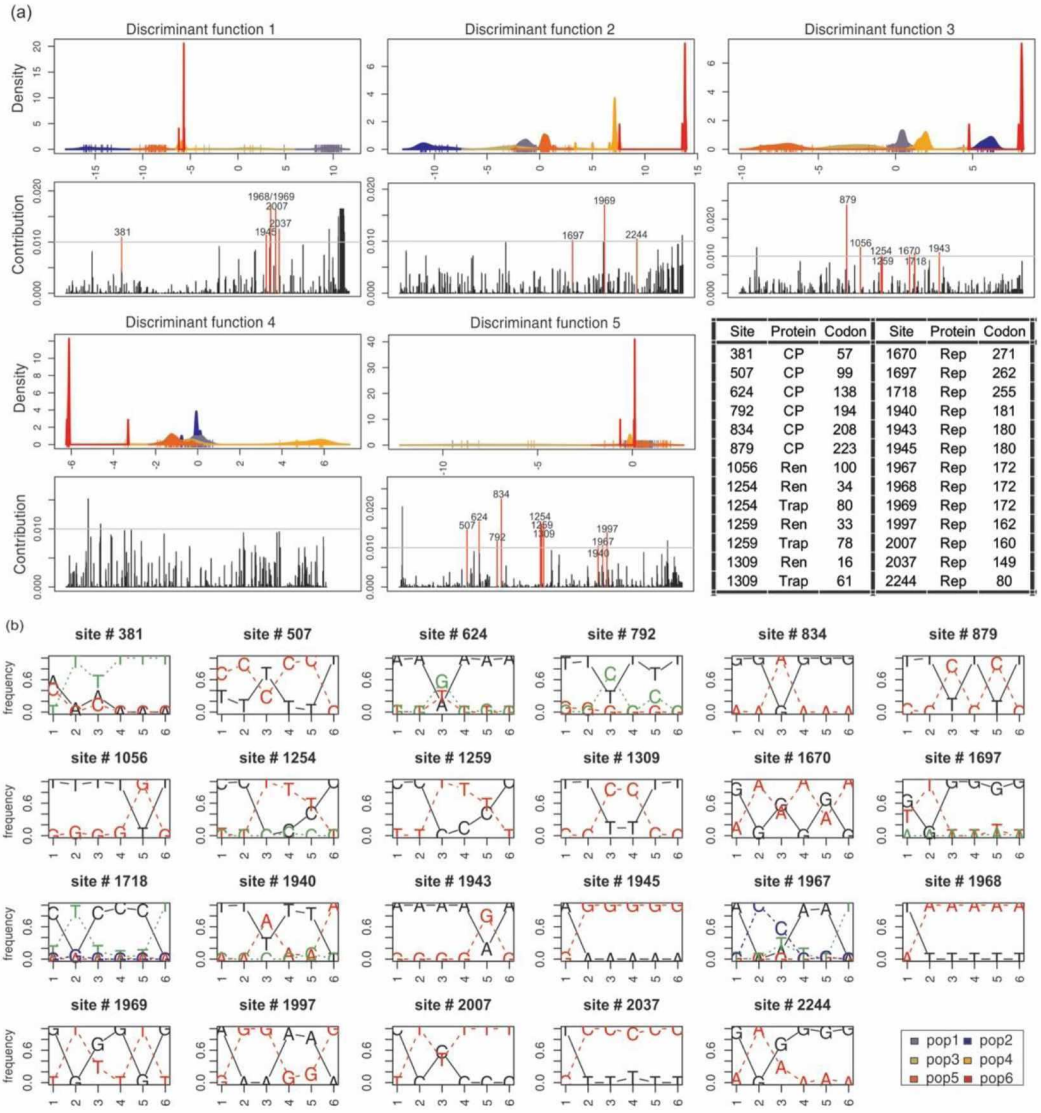
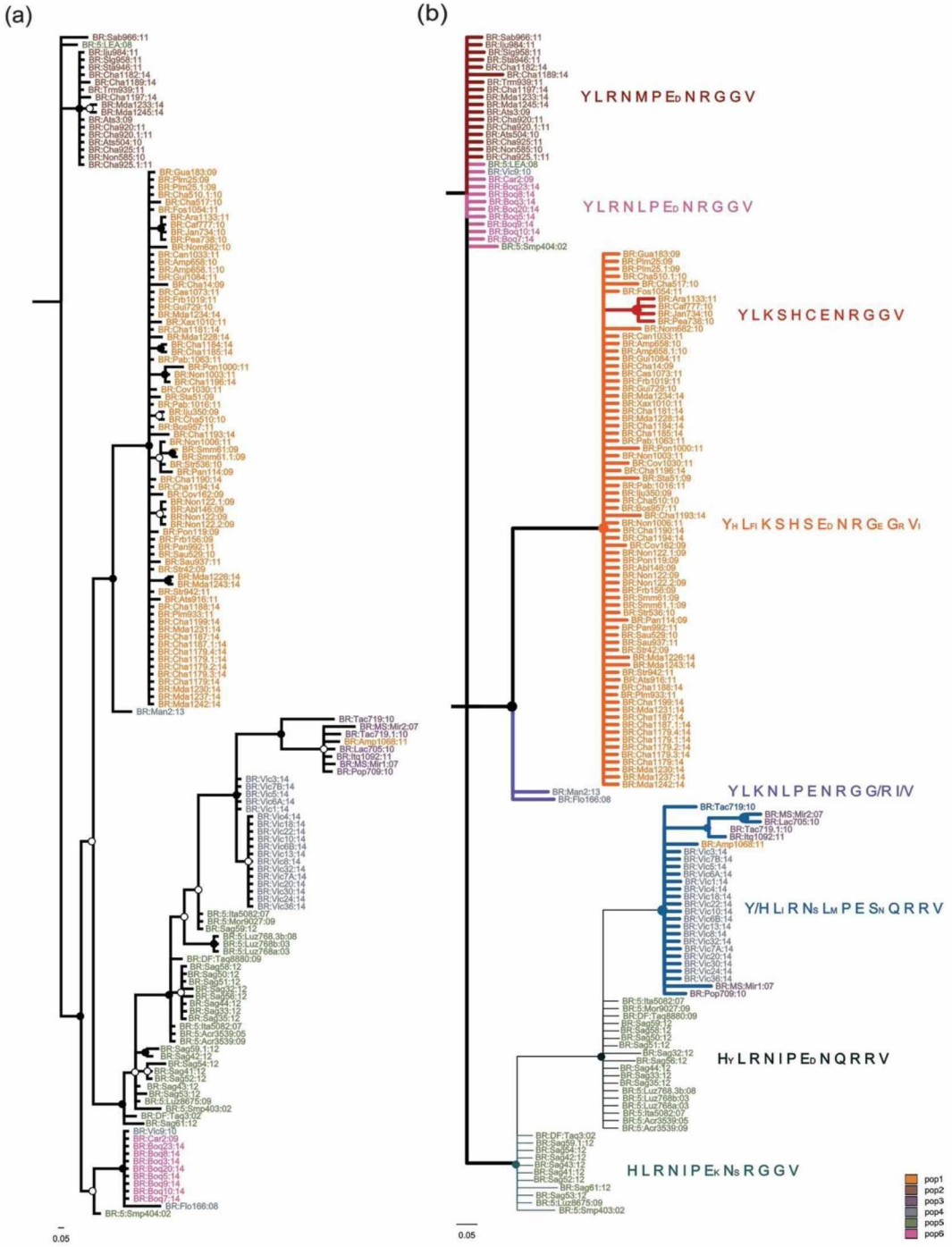

(a)
1
2
3
4
5
6
RS
PR
SC
MS
AM
MG
GO
PE
PB
5 15 25 35 45 55
(b)
1
2
3
4
5
6
pop1
pop2
pop3
pop4
pop5
pop6
10 30 50 70
(c)
% of reassignment
pop6
pop5
pop4
pop3
pop2
pop1
0
20
40
60
80
100
(d)
pop6
pop4
pop5
pop3
pop1
pop2
AM
PB
PE
DF
GO
MG
MS
PR
SC
RS
pop1
pop2
pop3
pop4
pop5
pop6

Figure 4



Site	Protein	Codon	Site	Protein	Codon
381	CP	57	1670	Rep	271
507	CP	99	1697	Rep	262
624	CP	138	1718	Rep	255
792	CP	194	1940	Rep	181
834	CP	208	1943	Rep	180
879	CP	223	1945	Rep	180
1056	Ren	100	1967	Rep	172
1254	Ren	34	1968	Rep	172
1254	Trap	80	1969	Rep	172
1259	Ren	33	1997	Rep	162
1259	Trap	78	2007	Rep	160
1309	Ren	16	2037	Rep	149
1309	Trap	61	2244	Rep	80

Figure 5


(a)
(b)
YLRNMPED NRGGV
YLRNLPED NRGGV
YLKSHCENRGGV
YH LFI K S H S ED N R GE GR VI
YLKNLPENRGG/R I/V
Y/H LI R NS LM P E SN Q R R V
HY L R N I P ED N Q R R V
H L R N I P EK NS R G G V
0.05
0.05
pop1
pop2
pop3
pop4
pop5
pop6

Supplementary material

Supplementary Table S1. Full-length sequences corresponding to *Euphorbia yellow mosaic virus* (EuYMV) DNA-A components obtained from *E. heterophylla* samples collected in Brazil. Sequences in parenthesis correspond to a second isolate of a haplotype, and therefore were not included in the data sets. Additional sequences retrieved from GenBank are underlined.

Sample code	Isolate name	Location	Host	Geographical coordinates		GenBank access #
Amazonas (AM)						
2013						
2	BR-Man2-13	Manaus	*E. heterophylla*	02°92'53"S	59°22'91"W	KY559430
Goiás (GO) / Federal District (DF)						
2002						
-	BR-DF-Taq3-02	Taquara/Planaltina	*E. heterophylla*	29°39'02"S	50°46'50"W	JF756670
-	BR-GO-Smp403-02	São Miguel do Passa Quatro	*E. heterophylla*	17°03'25"S	48°39'26"W	JF756672
-	BR-GO-Smp-404-02	São Miguel do Passa Quatro	*E. heterophylla*	17°03'25"S	48°39'26"W	JF756675
2003						
-	BR-GO-Luz768a-03	Luziânia	*Crotalaria sp.*	16°14'04"S	47°55'00"W	JX415184
-	(BR-768Tom8a-08)	Luziânia	*Solanum lycopersicum*	16°14'04"S	47°55'00"W	JX415192
-	(BR-GO-Luz768c-03)	Luziânia	*Crotalaria sp.*	16°14'04"S	47°55'00"W	JX415184
-	BR-GO-Luz768b-03	Luziânia	*Crotalaria sp.*	16°14'04"S	47°55'00"W	JX415185
2005						
-	BR-GO-Acr3539-05	Acreúna	*Crotalaria sp.*	17°23'47"S	50º22'31"W	JX415189
2007						
-	BR-DF-Ita5082-07	Itaberaí	*E. heterophylla*	16º01'13"S	49º48'37"W	JF756671
2008						
-	BR-GO-Luz768.3b-08	Luziânia	*Crotalaria juncea*	16°14'04"S	47°55'00"W	JX415191
-	BR-GO-LEA-08	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	FJ619507
2009						

-	BR-GO-Acr3539-09	Acreúna	*E. heterophylla*	17°23'47"S	50º22'31"W	JX415200
-	BR-DF-Taq8880-09	Taquara/Planaltina	*E. heterophylla*	29°39'02"S	50°46'50"W	JF756673
-	BR-GO-Mor9027-09	Morrinhos	*E. heterophylla*	17°43'52"S	49º05'58"W	JF756674
-	BR-GO-Luz8675-09	Luziânia	*E. heterophylla*	16°14'04"S	47°55'00"W	JF756676
2011						
-	BR-DF-Pla5818-07	Planaltina	*E. heterophylla*	29°39'02"S	50°46'50"W	JF756669
2012						
GO-32	BR-Sag32-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559431
GO-33	BR-Sag33-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559432
GO-35	BR-Sag35-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559433
GO-34	(BR-Sag34-12)	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559434
GO-41	BR-Sag41-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559435
GO-42	BR-Sag42-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559436
GO-43	BR-Sag43-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559437
GO-44	BR-Sag44-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559438
GO-50	BR-Sag50-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559439
GO-51	BR-Sag51-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559440
GO-52	BR-Sag52-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559441
GO-53	BR-Sag53-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559442
GO-54	BR-Sag54-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559443
GO-56	BR-Sag56-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559444
GO-57	(BR-Sag57-12)	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559445
GO-58	BR-Sag58-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559446
GO-59	BR-Sag59-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559447
	BR-Sag59.1-12					KY559448
GO-61	BR-Sag61-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559449
Mato Grosso do Sul (MS)						
2007						

-	BR-MS-Mir1-07	Miranda	*E. heterophylla*	20°14'25"S	56°22'55"W	FN435995
-	BR-MS-Mir2-07	Miranda	*E. heterophylla*	20°14'25"S	56°22'55"W	FN435997
2010						
705	BR-Lac705-10	Laguna Carapã	*E. heterophylla*	22°25'35"S	55°21'30"W	KY559450
709	BR-Pop709-10	Ponta Porã	*E. heterophylla*	22°36'29"S	55°38'52"W	KY559451
719	BR-Tac719-10	Tacuru	*E. heterophylla*	23°38'17"S	54°58'17"W	KY559452
	BR-Tac719.1-10					KY559453
2011						
1092	BR-Itq1092-11	Itaquiraí	*E. heterophylla*	23°12'37"S	54°12'32"W	KY559454
Minas Gerais (MG)						
2008						
-	BR-Flo166-08	Florestal	*E. heterophylla*	19°53'12"S	44°25'56"W	KC706530
2010						
-	BR-Vic9-10	Viçosa	*Sida santaremensis*	20°45'14"S	42°52'54"W	JX871379
2014						
1	BR-Vic1-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559455
3	BR-Vic3-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559456
4	BR-Vic4-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559457
5	BR-Vic5-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559458
8	BR-Vic8-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559459
10	BR-Vic10-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559460
13	BR-Vic13-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559461
18	BR-Vic18-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559462
20	BR-Vic20-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559463
22	BR-Vic22-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559464
24	BR-Vic24-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559465
30	BR-Vic30-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559466
32	BR-Vic32-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559467

36	BR-Vic36-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559468
6A	BR-Vic6A-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559469
6B	BR-Vic6B-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559470
7A	BR-Vic7A-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559471
7B	BR-Vic7B-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559472
Paraná (PR)						
2009						
156	BR-Frb156-09	Francisco Beltrão	*E. heterophylla*	26°04'01"S	53°00'42"W	KY559473
162	BR-Cov162-09	Coronel Vivida	*E. heterophylla*	25°58'13"S	52°27'02"W	KY559474
183	BR-Gua183-09	Guamiranga	*E. heterophylla*	25°11'15"S	50°52'36"W	KY559475
2010						
658	BR-Amp658-10	Ampére	*E. heterophylla*	25°57'06"S	53°24'25"W	KY559476
	BR-Amp658.1-10					KY559477
682	BR-Nom682-10	Nova Mercedes	*E. heterophylla*	24°30'54"S	54 07'00"W	KY559478
729	BR-Gui729-10	Guaíra	*E. heterophylla*	24°14'04"S	54°11'57"W	KY559479
734	BR-Jan734-10	Janiópolis	*E. heterophylla*	24°08'02"S	52°47'22"W	KY559480
738	BR-Pea738-10	Peabiru	*E. heterophylla*	23°52'26"S	52°19'05"W	KY559481
777	BR-Caf777-10	Cafelândia	*E. heterophylla*	24°39'21"S	53°03'07"W	KY559482
2011						
1016	BR-Pab-1016-11	Pato Branco	*E. heterophylla*	26°18'13"S	52°39'19"W	KY559483
1019	BR-Frb1019-11	Francisco Beltrão	*E. heterophylla*	26°04'01"S	53°00'42"W	KY559484
1030	BR-Cov1030-11	Coronel Vivida	*E. heterophylla*	25°58'13"S	52°27'02"W	KY559485
1033	BR-Can1033-11	Candói	*E. heterophylla*	25°32'07"S	52°00'03"W	KY559486
1063	BR-Pab-1063-11	Pato Branco	*E. heterophylla*	26°18'47"S	52°49'23"W	KY559487
1068	BR-Amp1068-11	Ampére	*E. heterophylla*	25°57'06"S	53°24'25"W	KY559488
1073	BR-Cas1073-11	Cascavel	*E. heterophylla*	24°50'06"S	53°37'59"W	KY559489
1084	BR-Gui1084-11	Guaíra	*E. heterophylla*	24°14'04"S	54°11'57"W	KY559490
1133	BR-Ara1133-11	Araruna	*E. heterophylla*	24°03'50"S	52°33'52"W	KY559491

Pernambuco (PE)						
2009						
-	BR-Car2-09	Caruaru	*Macroptilium atropurpureum*	08º02'23"S	34º54'55"W	JN419000
2014						
3PB1H	BR-Boq3-14	Boqueirão	*E. heterophylla*	07°28'49"S	36°08'2"W	KY559492
5PB1H	BR-Boq5-14	Boqueirão	*E. heterophylla*	07°28'49"S	36°08'2"W	KY559493
7PB1H	BR-Boq7-14	Boqueirão	*E. heterophylla*	07°28'49"S	36°08'2"W	KY559494
8PBII	BR-Boq8-14	Boqueirão	*E. heterophylla*	07°28'49"S	36°08'2"W	KY559495
9PB1H	BR-Boq9-14	Boqueirão	*E. heterophylla*	07°28'49"S	36°08'2"W	KY559496
14PB1H	(BR-Boq14-14)	Boqueirão	*E. heterophylla*	07°28'49"S	36°08'2"W	KY559497
10PB1H	BR-Boq10-14	Boqueirão	*E. heterophylla*	07°28'49"S	36°08'2"W	KY559498
18PBII	(BR-Boq18-14)	Boqueirão	*E. heterophylla*	07°28'49"S	36°08'2"W	KY559499
20PBII	BR-Boq20-14	Boqueirão	*E. heterophylla*	07°28'49"S	36°08'2"W	KY559500
23PB1H	BR-Boq23-14	Boqueirão	*E. heterophylla*	07°28'49"S	36°08'2"W	KY559501
Rio Grande do Sul (RS)						
2009						
3	BR-Ats3-09	Almirante Tamandaré	*E. heterophylla*	28°06'35"S	52°54'29"W	KY559502
14	BR-Cha14-09	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559503
	(BR-Cha14.3a-09)					KY559504
25	BR-Plm25-09	Palmeira das Missões	*E. heterophylla*	27°45'57"S	53°27'29"W	KY559505
	BR-Plm25.1-09					KY559506
42	BR-Str42-09	Santa Rosa	*E. heterophylla*	27°52'52"S	54°26'02"W	KY559507
51	BR-Sta51-09	Santo Ângelo	*E. heterophylla*	28°22'54"S	54°18'17"W	KY559508
61	BR-Smm61-09	São Miguel das Missões	*E. heterophylla*	28°29'35"S	54°33'37"W	KY559509
	BR-Smm61.1-09					KY559510
114	BR-Pan114-09	Panambi	*E. heterophylla*	28°18'24"S	53°29'21"W	KY559511
119	BR-Pon119-09	Nonoai	*E. heterophylla*	27°29'49"S	52°54'07"W	KY559512
122	BR-Non122-09	Nonoai	*E. heterophylla*	27°29'49"S	52°54'07"W	KY559513

	BR-Non122.1-09					KY559514
	BR-Non122.2-09					KY559515
350	BR-Iju350-09	Ijuí	*E. heterophylla*	28°22'57"S	54°02'40"W	KY559516
2010						
504	BR-Ats504-10	Almirante Tamandaré	*E. heterophylla*	28°06'35"S	52°54'29"W	KY559517
510	BR-Cha510-10	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559518
	BR-Cha510.1-10					KY559519
517	BR-Cha517-10	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559520
529	BR-Sau529-10	Santo Augusto	*E. heterophylla*	27°44'12"S	53°51'27"W	KY559521
536	BR-Str536-10	Santa Rosa	*E. heterophylla*	27°52'52"S	54°26'02"W	KY559522
585	BR-Non585-10	Nonoai	*E. heterophylla*	27°19'36"S	52°46'18"W	KY559523
2011						
916	BR-Ats916-11	Almirante Tamandaré	*E. heterophylla*	28°06'35"S	52°54'29"W	KY559524
920	BR-Cha920-11	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559525
	BR-Cha920.1-11					KY559526
	(BR-Cha920.3a-11)					KY559527
922	BR-Pan992-11	Panambi	*E. heterophylla*	28°18'24"S	53°29'21"W	KY559528
925	BR-Cha925-11	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559529
	BR-Cha925.1-11					KY559530
933	BR-Plm933-11	Palmeira das Missões	*E. heterophylla*	27°45'57"S	53°27'29"W	KY559531
937	BR-Sau937-11	Santo Augusto	*E. heterophylla*	27°44'12"S	53°51'27"W	KY559532
939	BR-Trm939-11	Três de Maio	*E. heterophylla*	27°45'39"S	54°15'42"W	KY559533
942	BR-Str942-11	Santa Rosa	*E. heterophylla*	27°45'39"S	54°15'42"W	KY559534
946	BR-Sta946-11	Santo Ângelo	*E. heterophylla*	28°22'54"S	54°18'17"W	KY559535
957	BR-Bos957-11	Bossoroca	*E. heterophylla*	28°43'27"S	54°55'30"W	KY559536
958	BR-Slg958-11	São Luiz Gonzaga	*E. heterophylla*	28°24'57"S	55°00'29"W	KY559537
966	BR-Sab966-11	São Borja	*E. heterophylla*	28°39'06"S	55°52'51"W	KY559538
984	BR-Iju984-11	Ijuí	*E. heterophylla*	28°23'00"S	54°02'56"W	KY559539

1000	BR-Pon1000-11	Pontão	*E. heterophylla*	27°58'09"S	52°43'47"W	KY559540
1003	BR-Non1003-11	Nonoai	*E. heterophylla*	27°29'49"S	52°54'07"W	KY559541
1006	BR-Non1006-11	Nonoai	*E. heterophylla*	27°19'36"S	52°46'18"W	KY559542
2014						
1179	BR-Cha1179-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559543
	BR-Cha1179.1-14					KY559544
	BR-Cha1179.2-14					KY559545
	BR-Cha1179.3-14					KY559546
	BR-Cha1179.4-14					KY559547
	(BR-Cha1179.2a-14)					KY559548
	(BR-Cha1179.7a-14)					KY559549
	(BR-Cha1179.8a-14)					KY559550
1181	BR-Cha1181-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559551
1182	BR-Cha1182-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559552
1184	BR-Cha1184-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559553
1185	BR-Cha1185-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559554
1187	BR-Cha1187-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559555
	BR-Cha1187.1-14					KY559556
1188	BR-Cha1188-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559557
1189	BR-Cha1189-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559558
1190	BR-Cha1190-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559559
1193	BR-Cha1193-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559560
1194	BR-Cha1194-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559561
1196	BR-Cha1196-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559562
1197	BR-Cha1197-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559563
1199	BR-Cha1199-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559564
1226	BR-Mda1226-14	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559565
1228	BR-Mda1228-14	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559566

1230	BR-Mda1230-14	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559567
1231	BR-Mda1231-14	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559568
	(BR-Mda1231.3e-14)					KY559569
1233	BR-Mda1233-14	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559570
1234	BR-Mda1234-14	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559571
1237	BR-Mda1237-14	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559572
1239	(BR-Mda1239-14)	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559573
1242	BR-Mda1242-14	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559574
1243	BR-Mda1243-14	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559575
1245	BR-Mda1245-14	Maximiliano de Almeida	*E. heterophylla*	27°66'91"S	51°79'88"W	KY559576
Santa Catarina (SC)						
2009						
1146	BR-Abl146-09	Abelardo Luz	*E. heterophylla*	26°32'01"S	52°17'09"W	KY559577
2011						
1010	BR-Xax1010-11	Xaxim	*E. heterophylla*	26°56'17"S	52°29'54"W	KY559578
1054	BR-Fos1054-11	Formosa do Sul	*E. heterophylla*	26°34'05"S	52°47'39"W	KY559579

Supplementary Table S2. Full-length sequences corresponding to *Euphorbia yellow mosaic virus* (EuYMV) DNA-B components obtained from *E. heterophylla* samples collected in Brazil. Sequences in parenthesis correspond to a second isolate of a haplotype, and therefore were not included in the data sets. Additional sequences retrieved from GenBank are underlined.

Sample code	Isolate name	Location	Host	Geographical coordinates		GenBank access #
Goiás (GO) / Federal District (DF)						
2005						
-	BR-GO-Acr3540-05	Acreúna	*E. heterophylla*	17°22'44"S	17°22'44"W	JF756678
2007						
-	BR-GO-Ita5082-07	Itaberaí	*E. heterophylla*	16º01'13"S	49º48'37"W	JF756677
2012						
GO-33	BR-Sag-33-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559580
GO-34	BR-Sag34-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559581
GO-40	BR-Sag40-1-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559582
	(BR-Sag40-2-12)					KY559583
GO-43	BR-Sag43-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559584
GO-44	BR-Sag-44-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559585
GO-52	BR-Sag52-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559586
GO-53	BR-Sag53-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559587
GO-54	BR-Sag54-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559588
GO-56	BR-Sag56-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559589
GO-57	BR-Sag57-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559590
GO-58	BR-Sag58-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559591
GO-59	BR-Sag59-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559592
GO-61	BR-Sag-61-1-12	Santo Antônio de Goiás	*E. heterophylla*	16°30'22"S	49°17'06"W	KY559593
	BR-Sag61-2-12					KY559594
Mato Grosso do Sul (MS)						

2007						
-	BR-MS-Mir1-07	Miranda	*E. heterophylla*	20°14'25"S	56°22'55"W	FN435996
-	BR-MS-Mir2-07	Miranda	*E. heterophylla*	20°14'25"S	56°22'55"W	FN435998
Minas Gerais (MG)						
2014						
54	BR-Vic54-14	Viçosa	*E. heterophylla*	20°45'14"S	42°52'54"W	KY559595
Paraná (PR)						
2009						
171	BR-Can171.1-09	Candói	*E. heterophylla*	25°32'07"S	52°00'03"W	KY559596
	(BR-Can171.2-09)					KY559597
179	BR-Gua179-09	Guamiranga	*E. heterophylla*	25°11'15"S	50°52'36"W	KY559598
183	BR-Gua183-09	Guamiranga	*E. heterophylla*	25°11'15"S	50°52'36"W	KY559599
2010						
682	BR-Nom682-10	Nova Mercedes	*E. heterophylla*	24°30'54"S	54 07'00"W	KY559600
	(BR-Nom682.2-10)					KY559601
777	BR-Caf777-10	Cafelândia	*E. heterophylla*	24°39'21"S	53°03'07"W	KY559602
2011						
1019	BR-Frb1019-11	Francisco Beltrão	*E. heterophylla*	26°04'01"S	53°00'42"W	KY559603
	BR-Frb1019.1-11					KY559604
1033	BR-Can1033-11	Candói	*E. heterophylla*	25°32'07"S	52°00'03"W	KY559605
1068	BR-Amp1068-11	Ampére	*E. heterophylla*	25°57'06"S	53°24'25"W	KY559606
1133	BR-Ara1133-11	Araruna	*E. heterophylla*	24°03'50"S	52°33'52"W	KY559607
Pernambuco (PE)						
2009						
-	BR-Car2-09	Caruaru	*Macroptilium atropurpureum*	08º02'23"S	34º54'55"W	JN419001
Rio Grande do Sul (RS)						
2009						

3	BR-Ats3-09	Almirante Tamandaré	*E. heterophylla*	28°06'35"S	52°54'29"W	KY559608
	BR-Ats3.1-09					KY559609
	BR-Ats3.2-09					KY559610
	(BR-Ats3.3-09)					KY559611
14	BR-Cha14-09	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559612
51	BR-Sta51-09	Santo Ângelo	*E. heterophylla*	28°22'54"S	54°18'17"W	KY559613
61	BR-Smm61-09	São Miguel das Missões	*E. heterophylla*	28°29'35"S	54°33'37"W	KY559614
2010						
504	BR-Ats504-10	Almirante Tamandaré	*E. heterophylla*	28°06'35"S	52°54'29"W	KY559615
	BR-Ats504.1-10					KY559616
510	BR-Cha510-10	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559617
	BR-Cha510.1-10					KY559618
	BR-Cha510.2-10					KY559619
536	BR-Str536-10	Santa Rosa	*E. heterophylla*	27°52'52"S	54°26'02"W	KY559620
2011						
920	BR-Cha920-11	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559621
	BR-Cha920.1-11					KY559622
925	BR-Cha925-11	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559623
933	BR-Plm933-11	Palmeira das Missões	*E. heterophylla*	27°45'57"S	53°27'29"W	KY559624
937	BR-Sau937-11	Santo Augusto	*E. heterophylla*	27°44'12"S	53°51'27"W	KY559625
939	BR-Trm939-11	Três de Maio	*E. heterophylla*	27°45'39"S	54°15'42"W	KY559626
942	BR-Str942-11	Santa Rosa	*E. heterophylla*	27°45'39"S	54°15'42"W	KY559627
946	BR-Sta946-11	Santo Ângelo	*E. heterophylla*	28°22'54"S	54°18'17"W	KY559628
984	BR-Iju984-11	Ijuí	*E. heterophylla*	28°23'00"S	54°02'56W	KY559629
	BR-Iju984-1-11					KY559630
1000	BR-Pon1000-11	Pontão	*E. heterophylla*	27°58'09"S	52°43'47"W	KY559631
1003	BR-Non1003-11	Nonoai	*E. heterophylla*	27°29'49"S	52°54'07"W	KY559632

2014						
1179	BR-Cha1179-14	Chapada	*E. heterophylla*	28°02'39"S	53°04'53"W	KY559633
	BR-Cha1179.1-14					KY559634
	BR-Cha1179.2-14					KY559635
	(BR-Cha1179.3-14)					KY559636

Supplementary Table S3. Putative recombination events detected by RDP in the DNA-B of *Euphorbia yellow mosaic virus* (EuYMV) isolates.

Event	Recombinant	Recombination breakpoints*		Parents		Methods†	*P*-value‡
		Initial	Final	Major	Minor		
1	BRFrb1019.1:11	827	2568	BR:Gua179:09	BR:Str942:11	RMCS**T**	3.7×10^{-13}
	BRFrb1019:11	827	30	BR:Gua179:09	BR:Str942:11		
2	BR:Sag44:12	1146	2262	BR:Sag43:12	Unknown	GBMC**S**	5.9×10^{-5}
	BR:Sag:33:12	1230	2338	BR:Sag43:12	Unknown		
3	BR:Ats3.2:09	861	1694	BR:Caf777:10	BR:GO:Ita5082:07	RMCS**T**	1.5×10^{-3}

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

† R, RDP; B, Bootscan; M, MaxChi; C, Chimera; S, SisScan; T, 3SEQ.

‡ The reported *P* values is for the programs indicated in bold and is the lowest *P*-value calculated for the event in question.

Supplementary Table S4. Analysis of molecular variance (AMOVA) performed on *Euphorbia yellow mosaic virus* (EuYMV) subpopulations.

Analysis	Source of variation	D.f	Square sum	Variance components	% of variation
DNA-A					
year*	Among populations	5	1182.995	8.25430 Va	20.91
	Within populations	152	4745.999	31.22368 Vb	79.09
	Total	157	5928.994	39.47797	
	Fst : 0.209				
DNA-A (hierarchical)					
Cluster I-II	Among groups	4	1441.738	9.14395 Va	19.01
Pop 1-6	Among populations, within groups	1	1558.254	19.67925 Vb	40.92
	Within populations	152	2929.002	19.26975 Vc	40.07
	Total	157	5928.994	48.09295	
	Fsc : 0.505				
	Fst : 0.599				
	Fct : 0.190				

* considering as populations samples collected in (1) 2002 to 2008, (2) 2009, (3) 2010, (4) 2011, (5) 2013 to 2013, and (6) 2014.

Supplementary Table S5. Comparison between Spatial Analysis of Molecular Variance (SAMOVA) and Discriminant Analysis of Principal Components (DAPC) inferred subpopulations **(k=6).**

Isolate name	SAMOVA	DAPC
BR:DF:Taq3:02	group3	pop5
BR:DF:Taq8880:09	group3	pop5
BR:Bos957:11	group1	pop1
BR:Sab966:11	group1	pop2
BR:Smm61:09	group1	pop1
BR:Smm61.1:09	group1	pop1
BR:Slg958:11	group1	pop2
BR:Iju984:11	group1	pop2
BR:Iju350:09	group1	pop1
BR:Sta51:09	group1	pop1
BR:Sta946:11	group1	pop2
BR:Pan114:09	group1	pop1
BR:Pan992:11	group1	pop1
BR:Mda1234:14	group1	pop1
BR:Mda1228:14	group1	pop1
BR:Mda1226:14	group1	pop1
BR:Mda1243:14	group1	pop1
BR:Mda1231:14	group1	pop1
BR:Mda1230:14	group1	pop1
BR:Mda1237:14	group1	pop1
BR:Mda1242:14	group1	pop1
BR:Mda1233:14	group1	pop2
BR:Mda1245:14	group1	pop2
BR:Ats916:11	group1	pop1
BR:Ats3:09	group1	pop2
BR:Ats504:10	group1	pop2
BR:Cha14:09	group1	pop1
BR:Cha510.1:10	group1	pop1
BR:Cha517:10	group1	pop1
BR:Cha510:10	group1	pop1
BR:Cha1181:14	group1	pop1
BR:Cha1184:14	group1	pop1
BR:Cha1185:14	group1	pop1
BR:Cha1196:14	group1	pop1
BR:Cha1193:14	group1	pop1
BR:Cha1190:14	group1	pop1
BR:Cha1194:14	group1	pop1
BR:Cha1188:14	group1	pop1
BR:Cha1199:14	group1	pop1
BR:Cha1187:14	group1	pop1
BR:Cha1187:14	group1	pop1
BR:Cha1179.4:14	group1	pop1
BR:Cha1179.1:14	group1	pop1
BR:Cha1179.2:14	group1	pop1

Suppl. Table S5 (cont.)

Isolate name	SAMOVA	DAPC
BR:Cha1179.3:14	group1	pop1
BR:Cha1179:14	group1	pop1
BR:Cha920:11	group1	pop2
BR:Cha920.1:11	group1	pop2
BR:Cha925:11	group1	pop2
BR:Cha925.1:11	group1	pop2
BR:Cha1182:14	group1	pop2
BR:Cha1189:14	group1	pop2
BR:Cha1197:14	group1	pop2
BR:Pon1000:11	group1	pop1
BR:Str42:09	group1	pop1
BR:Str536:10	group1	pop1
BR:Plm25:09	group1	pop1
BR:Plm25.1:09	group1	pop1
BR:Plm933:11	group1	pop1
BR:Str942:11	group1	pop1
BR:Trm939:11	group1	pop2
BR:Sau529:10	group1	pop1
BR:Sau937:11	group1	pop1
BR:Non122.1:09	group1	pop1
BR:Pon119:09	group1	pop1
BR:Non122:09	group1	pop1
BR:Non122.2:09	group1	pop1
BR:Non1003:11	group1	pop1
BR:Non1006:11	group1	pop1
BR:Non585:10	group1	pop2
BR:Xax1010:11	group1	pop1
BR:Fos1054:11	group1	pop1
BR:Abl146:09	group1	pop1
BR:Pab:1063:11	group1	pop1
BR:Pab:1016:11	group1	pop1
BR:Frb156:09	group1	pop1
BR:Frb1019:11	group1	pop1
BR:Cov162:09	group1	pop1
BR:Cov1030:11	group1	pop1
BR:Amp658:10	group1	pop1
BR:Amp658.1:10	group1	pop1
BR:Amp1068:11	group1	pop1
BR:Can1033:11	group1	pop1
BR:Gua183:09	group1	pop1
BR:Cas1073:11	group1	pop1
BR:Caf777:10	group1	pop1
BR:Nom682:10	group1	pop1
BR:Gui729:10	group1	pop1

Suppl. Table S5 (cont.)

Isolate name	SAMOVA	DAPC
BR:Ara1133:11	group1	pop1
BR:Pea738:10	group1	pop1
BR:Tac719:10	group5	pop3
BR:Tac719.1:10	group5	pop3
BR:Itq1092:11	group5	pop3
BR:Pop709:10	group5	pop3
BR:Lac705:10	group5	pop3
BR:Vic9:10	group2	pop4
BR:Vic3:14	group2	pop4
BR:Vic7B:14	group2	pop4
BR:Vic5:14	group2	pop4
BR:Vic6A:14	group2	pop4
BR:Vic1:14	group2	pop4
BR:Vic4:14	group2	pop4
BR:Vic18:14	group2	pop4
BR:Vic22:14	group2	pop4
BR:Vic10:14	group2	pop4
BR:Vic6B:14	group2	pop4
BR:Vic13:14	group2	pop4
BR:Vic8:14	group2	pop4
BR:Vic32:14	group2	pop4
BR:Vic7A:14	group2	pop4
BR:Vic20:14	group2	pop4
BR:Vic30:14	group2	pop4
BR:Vic24:14	group2	pop4
BR:Vic36:14	group2	pop4
BR:MS:Mir2:07	group5	pop3
BR:MS:Mir1:07	group5	pop3
BR:Flo166:08	group3	pop4
BR:GO:Mor9027:09	group3	pop5
BR:GO:Acr3539:05	group3	pop5
BR:GO:Acr3539:09	group3	pop5
BR:GO:Smp404:02	group3	pop5
BR:GO:Smp403:02	group3	pop5
BR:GO:LEA:08	group3	pop5
BR:Sag59.1:12	group3	pop5
BR:Sag54:12	group3	pop5
BR:Sag59:12	group3	pop5
BR:Sag58:12	group3	pop5
BR:Sag50:12	group3	pop5
BR:Sag51:12	group3	pop5
BR:Sag32:12	group3	pop5
BR:Sag56:12	group3	pop5
BR:Sag44:12	group3	pop5

Suppl. Table S5 (cont.)

Isolate name	SAMOVA	DAPC
BR:Sag35:12	group3	pop5
BR:Sag42:12	group3	pop5
BR:Sag43:12	group3	pop5
BR:Sag41:12	group3	pop5
BR:Sag52:12	group3	pop5
BR:Sag61:12	group3	pop5
BR:Sag53:12	group3	pop5
BR:GO:Luz768b:03	group3	pop5
BR:GO:Luz768a:03	group3	pop5
BR:GO:Luz768.3b:08	group3	pop5
BR:GO:Luz8675:09	group3	pop5
BR:DF:Pla5818:07	group3	pop5
BR:GO:Ita5082:07	group3	pop5
BR:Car2:09	group4	pop6
BR:Boq23:14	group4	pop6
BR:Boq8:14	group4	pop6
BR:Boq3:14	group4	pop6
BR:Boq20:14	group4	pop6
BR:Boq5:14	group4	pop6
BR:Boq9:14	group4	pop6
BR:Boq10:14	group4	pop6
BR:Boq7:14	group4	pop6
BR:Man2:13	group6	pop4

Supplementary Table S6. Results of the different neutrality tests and mean ratios of non-synonymous to synonymous substitution (d_N/d_S) for each open reading frame in the DNA-A of each subpopulation of *Euphorbia yellow mosaic virus* (EuYMV).

Population	ORF	d_N/d_S	Tajima's D	Fu and Li's D*	Fu and Li's F*
pop1	CP	0.064	-2.101*	-4.020†	-3.891†
	Rep	0.168	-2.105*	-4.463†	-4.197†
	Trap	0.688	-2.346‡	-3.427†	-3.598†
	Ren	0.235	-2.426‡	-4.082†	-4.107†
pop2	CP	0.051	-1.238	-1.574	-1.711
	Rep	0.178	-0.705	-1.678	-1.619
	Trap	0.497	-0.044	-0.367	-0.319
	Ren	0.157	-0.497	-0.945	-0.945
pop3	CP	0.026	-0.697	-0.614	-0.699
	Rep	0.251	-0.564	-0.461	-0.536
	Trap	0.367	-0.998	-0.904	-1.016
	Ren	0.259	-1.271	-1.310	-1.432
pop4	CP	0.147	-2.004*	-2.544*	-2.779*
	Rep	0.226	-1.338	-1.302	-1.534
	Trap	0.527	-1.956*	-2.009	-2.320
	Ren	0.146	-2.009*	-2.313	-2.588
pop5	CP	0.119	-2.096*	-2.959*	-3.160*
	Rep	0.330	-1.928*	-3.065*	-3.176*
	Trap	0.936	-0.748	-1.965	-1.850
	Ren	0.201	-1.151	-2.900*	-2.752*
pop6	CP	0.630	-1.812*	-1.994*	-2.184*
	Rep	0.201	-1.911*	-2.172†	-2.361†
	Trap	1.264	-1.678	-1.881	-2.039
	Ren	0.303	-1.610	-1.799	-1.948
Total	CP	0.098	-2.041*	-4.713†	-4.134†
	Rep	0.231	-1.913*	-4.856†	-4.077†
	Trap	0.697	-2.004*	-4.146†	-3.805†
	Ren	0.224	-2.328‡	-5.629†	-4.879†

Significant values that reject the null hypothesis of selective neutrality at: *, $P<0.05$; †, $P<0.02$, and ‡, $P<0.01$.

Supplementary Table S7. Number of positively and negatively selected amino acid sites in the *CP*, *Rep*, *Trap* and *Ren* genes of the six *Euphorbia yellow mosaic virus* (EuYMV) subpopulations.

Gene	Population	Negative selection*					Positive selection*				
		SLAC	FEL	IFEL	REL	TOTAL	SLAC	FEL	IFEL	REL	TOTAL
CP	pop1	26	56	6	54	62	0	0	0	0	0
	pop2	2	17	2	†	17	0	0	0	0	0
	pop3	1	17	4	37	37	0	0	0	0	0
	pop4	0	19	1	33	33	0	0	0	2	2
	pop5	3	20	3	0	23	0	0	0	0	0
	pop6	0	4	0	0	4	0	0	0	2	2
	TOTAL	54	85	42	-	86	0	0	1	-	1
Rep	pop1	35	59	27	26	60	1	1	2	12	12
	pop2	2	19	5	33	34	0	0	0	0	0
	pop3	1	15	3	36	36	0	0	1	0	1
	pop4	1	24	2	0	24	0	0	1	5	5
	pop5	8	25	7	9	26	1	3	2	4	5
	pop6	0	7	0	†	7	0	0	0	0	0
	TOTAL	63	85	57	-	93	2	6	7	-	9
Trap	pop1	1	11	2	0	11	1	1	1	0	1‡
	pop2	0	4	3	†	4	0	0	0	0	0
	pop3	0	4	0	0	4	0	0	0	1	1
	pop4	0	5	1	0	5	0	0	1	4	5
	pop5	0	3	2	0	3	0	0	0	1	1
	pop6	0	0	0	1	1	0	0	0	0	0
	TOTAL	8	19	6	-	19	1	5	5	-	6
Ren	pop1	5	21	3	37	37	0	0	1	0	1
	pop2	0	7	0	16	16	0	0	0	0	0
	pop3	0	2	0	7	7	0	0	0	0	0
	pop4	0	5	1	†	5	0	0	0	0	0
	pop5	2	5	1	1	5	0	0	0	1	1
	pop6	0	0	2	0	2	0	0	0	1	1
	TOTAL	14	26	11	-	28	0	0	1	-	1

* negatively and positively selected sites with 0.1 significance level

† no rates with dN>dS were inferred, suggesting that all sites are under purifying selection

‡ PARRIS evidence of positive selection at $p < 0.1$, with dN/dS ratio $\omega = 0.559$ and p-value = 0.0923

- analysis not performed due to alignment size restriction

Supplementary Table S8. Most contributing sites mapped in the genome of *Euphorbia yellow mosaic virus* (EuYMV) isolates, compared to changes at the amino acid level and codon selection analysis.

Discriminant function	Site	Protein	Codon	Position	Site	RC	Amino acid	Selection						
								pop1	pop2	pop3	pop4	pop5	pop6	TOTAL
DF-1	381	CP	57	3	GT**A** GT**T** GT**C**		Valine (V) Valine (V) Valine (V)	FR*	R	FR				SFI
DF-5	507	CP	99	3	GG**C** GG**T**		Glycine (G) Glycine (G)	SFIR	R	FR	R			SFI
DF-5	624	CP	138	3	CC**A** CC**T** CC**G**		Proline(P) Proline(P) Proline(P)	FR	R	FIR		F		SFI
DF-5	792	CP	194	3	CG**G** CT**G** CG**C**		Arginine (R) Arginine (R) Arginine (R)	SFR	R	R				SFI
DF-5	834	CP	208	3	GA**G** GA**A**		Glut. acid (E) Glut. acid (E)	FR	R					F
DF-3	879	CP	223	3	TA**T** TA**C** CA**T**		Tyrosine (Y) Tyrosine (Y) Histidine (H)[¥]	SFR	R	R		SF		SFI
DF-3	1056	Ren	100	3	**T**AC **T**AT **G**AC	GT**A** AT**A** GT**C**	Valine (V) Isoleucine (I) Valine (V)		R		R			S
DF-3	1254	Ren	34	3	**C**AC **G**AC **T**AC	GT**G** GT**C** GT**A**	Valine (V) Valine (V) Valine (V)	FR			FR			SF

DF-3 DF-5	1254	Trap	80	1	TC**C** **G**GA TC**G** **C**GA TC**T** **A**GA	Glycine (G) Arginine (R) Arginine (R)	P(+)†	R		R(+)			F(+)
DF-3 DF-5	1259	Ren	33	1	CC**C** **G**GG CC**T** **A**GG	Glycine (G) Arginine (R)	R			R			
DF-3	1259	Trap	78	2	C**C**G C**G**G C**T**G C**A**G	Arginine (R) Glutamine (Q)		R		I(+)			
DF-5	1309	Ren	16	2	A**T**T A**A**T A**C**T A**G**T	Asparagine (N) Serine (S)				R			
DF-5	1309	Trap	61	3	**T**TC GA**A** **C**TC GA**G**	Glut. acid (E) Glut. acid (E)	F	R	F	F	FI		SFI
DF-3	1670	Rep	271	3	**G**TG CA**C** **A**TG CA**T**	Histidine (H) Histidine (H)	SFIR	FR	FR		SFIR	R	SFI
DF-2	1697	Rep	262	3	**G**AC GT**C** **T**AC GT**A** **A**AC GT**T**	Valine (V) Valine (V) Valine (V)	SFIR	R		F		R	SFI
DF-3	1718	Rep	255	3	**C**TC GA**G** **A**TC GA**T** **T**TC GA**A** **G**TC GA**C** **C**TT AA**G**	Glut. acid (E) Aspartic acid (D) Glut. acid (E) Aspartic acid (D) Lysine (K)	SFIR	R	FR	F	I	R	SFI
DF-5	1940	Rep	181	3	**T**GG CC**A** **A**GG CC**T**	Proline (P) Proline (P)			R	F	SFR	R	SFI
DF-3 DF-1	1943 1945	Rep Rep	180 180	3 1	AG**G** **C**CT AG**A** **T**CT AC**A** **T**GT GG**G** **C**CC	Proline(P) Serine (S) Cysteine (C) Proline (P)	R(+)				SFR	R	S

DF-5 DF-2	1967	Rep	172	3	**CAT** **ATG** **ATG** **CAT**	Metionine (M) Histidine (H)		R			R		
DF-1	1968	Rep	172	2	**GTG** **CAC**	Histidine (H)							
DF-1 DF-2	1969	Rep	172	1	**AAG** **CTT** **TAG** **CTA** **CAG** **CTG** **AAT** **ATT** **GAT** **ATC**	Leucine (L) Leucine (L) Leucine (L) Isoleucine (I) Isoleucine (I)							
DF-5	1997	Rep	162	3	**G**TT AA**C** **A**TT AA**T**	Asparagine (N) Asparagine (N)			F	SFIR	R	SFI	
DF-1	2007	Rep	160	2	A**T**T A**A**T A**C**T A**G**T	Asparagine (N) Serine (S)					R		
DF-1	2037	Rep	149	2	C**C**T A**G**G C**T**T A**A**G T**T**T A**A**A T**C**T A**G**A	Arginine (R) Lysine (K) Lysine (K) Arginine (R)	SFI		R(+)	SFIR	R	SFI	
DF-2	2244	Rep	80	3	**A**AG CT**T** **G**AG CT**C** **G**AA TT**C** **G**AT AT**C**	Leucine (L) Leucine (L) Phenylalanine (F) Isoleucine (I)	S	R		F	R	S	

* Negatively selected site with 0.1 significance level with SLAC (S), FEL (F), IFEL (I), and REL (R) methods

† Positively selected site with 0.1 significance level with SLAC (S+) FEL (F+) IFEL (I+) REL (R+) and PARRIS (P+) methods

¥ We observed a predominance of Y in all subpopulations, except pop5 in which H was present instead

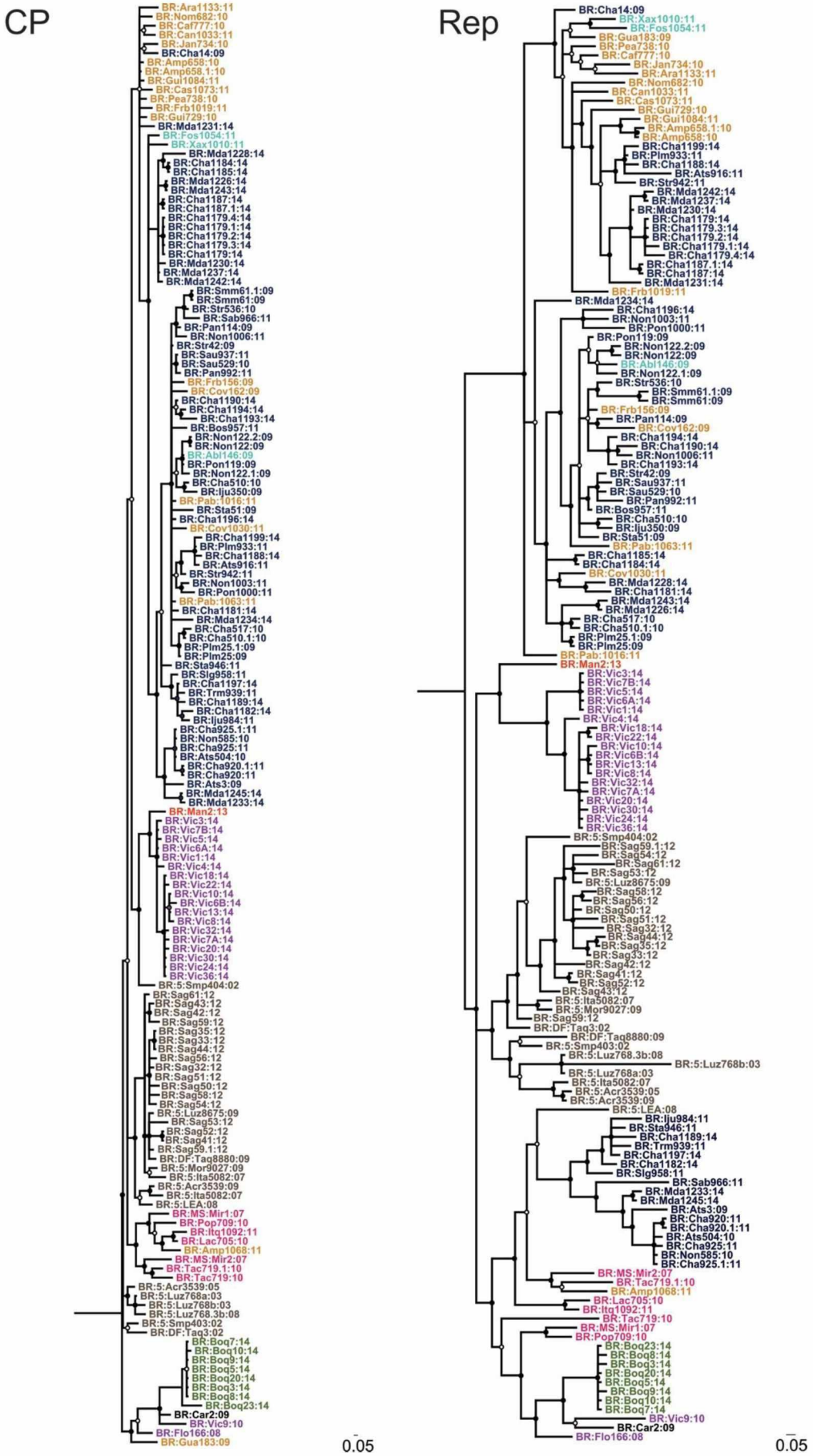
Supplementary figure legends

Supplementary Fig. S1. Midpointed-rooted Bayesian inference trees based on the nucleotide sequences of the *CP*, *Rep* (**a**), *MP* and *NSP* (**b**) genes of EuYMV isolates. Nodes to the right of branches with posterior probabilities equal to or higher than 0.8 are indicated by filled circles and those with values lower than 0.8 and higher than 0.5 by empty circles. Isolate colors based on the geographical region of origin: AM, Amazonas; GO, Goiás; MG, Minas Gerais; MS, Mato Grosso do Sul; PB, Paraíba; PE, Pernambuco; PR, Paraná; RS, Rio Grande do Sul; SC, Santa Catarina. Recombinant isolates according to RDP analysis (see Suppl. Table S3) are indicated by arrows.

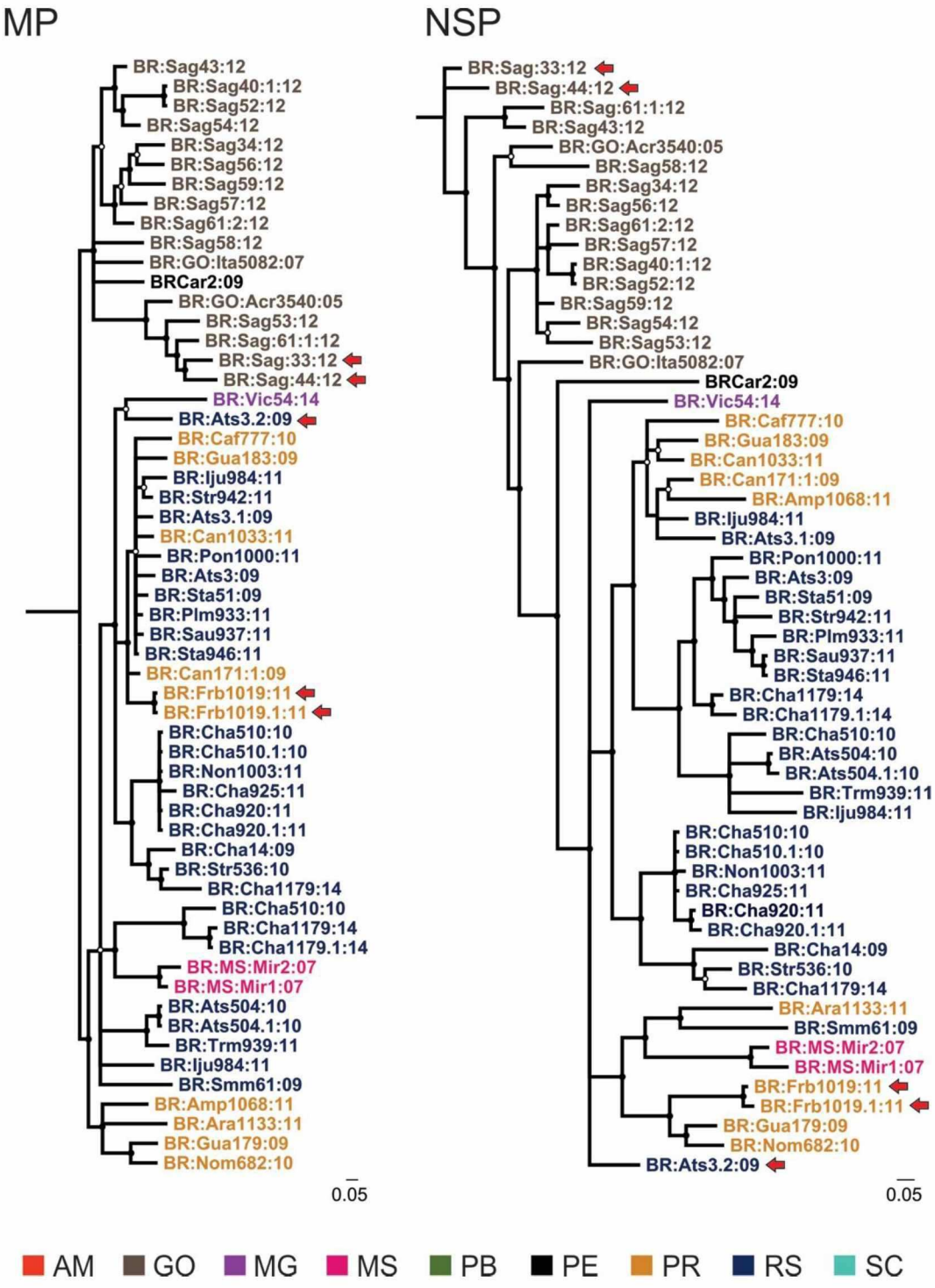
Supplementary Fig. S2. Bayesian clustering analysis of population subdivision using STRUCTURE, based on (**a**) EuYMV DNA-A and (**b**) DNA-B. Each individual is represented by a horizontal line divided into cluster I and cluster II (K=2). Admixing percentages for each isolate were calculated and the average admixing percentage for isolates from the same sampling location are represented on the map as pie charts. The size of the pie chart is proportional to the number of isolates from each sampling location.

Supplementary Fig. S3. Multivariate statistical clustering analysis based on EuYMV DNA-A. **(a)** The slot indicates the proportions of successful reassignment (based on the discriminant functions) of isolates to subpopulations. Heat colors represent membership probabilities (red=1, white=0) and blue crosses represent the prior cluster provided to DAPC. **(b)** Distribution of the SNPs extracted for DAPC analysis throughout the DNA-A. **(c)** a-score analysis, displaying the proportion of successful reassignment corrected for the number of retained PCs. **(d)** Cross-validation analysis. Individual replicates appear as points, and the density of those points in different regions of the plot is displayed in blue.

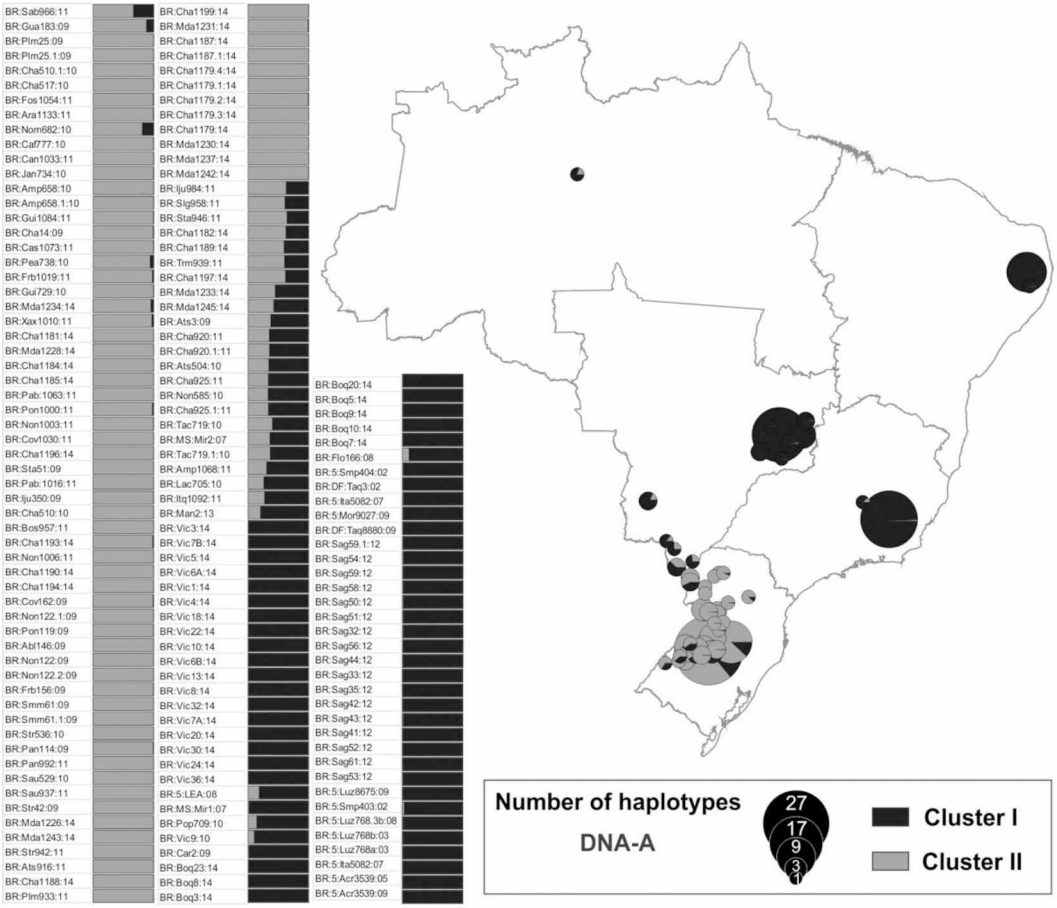
Supplementary Fig. S1a



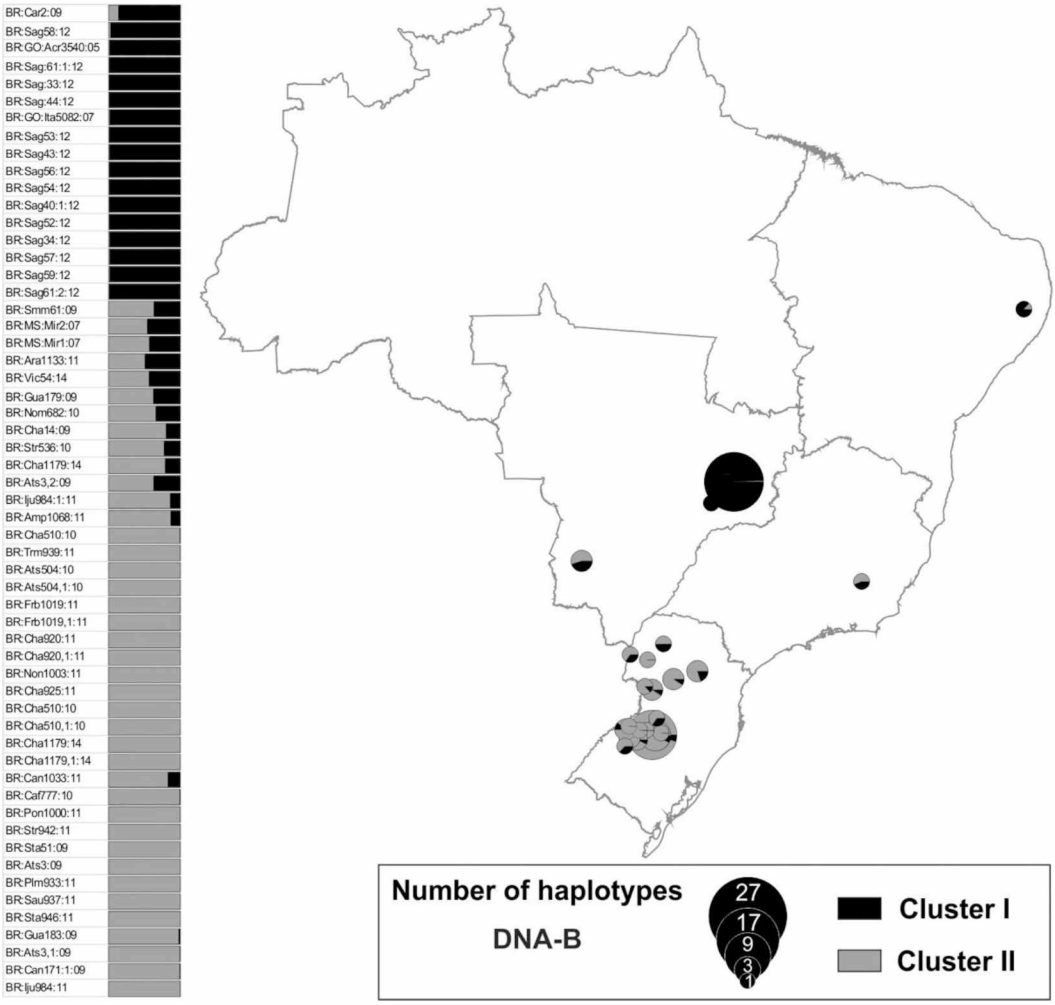
Supplementary Fig. S1b



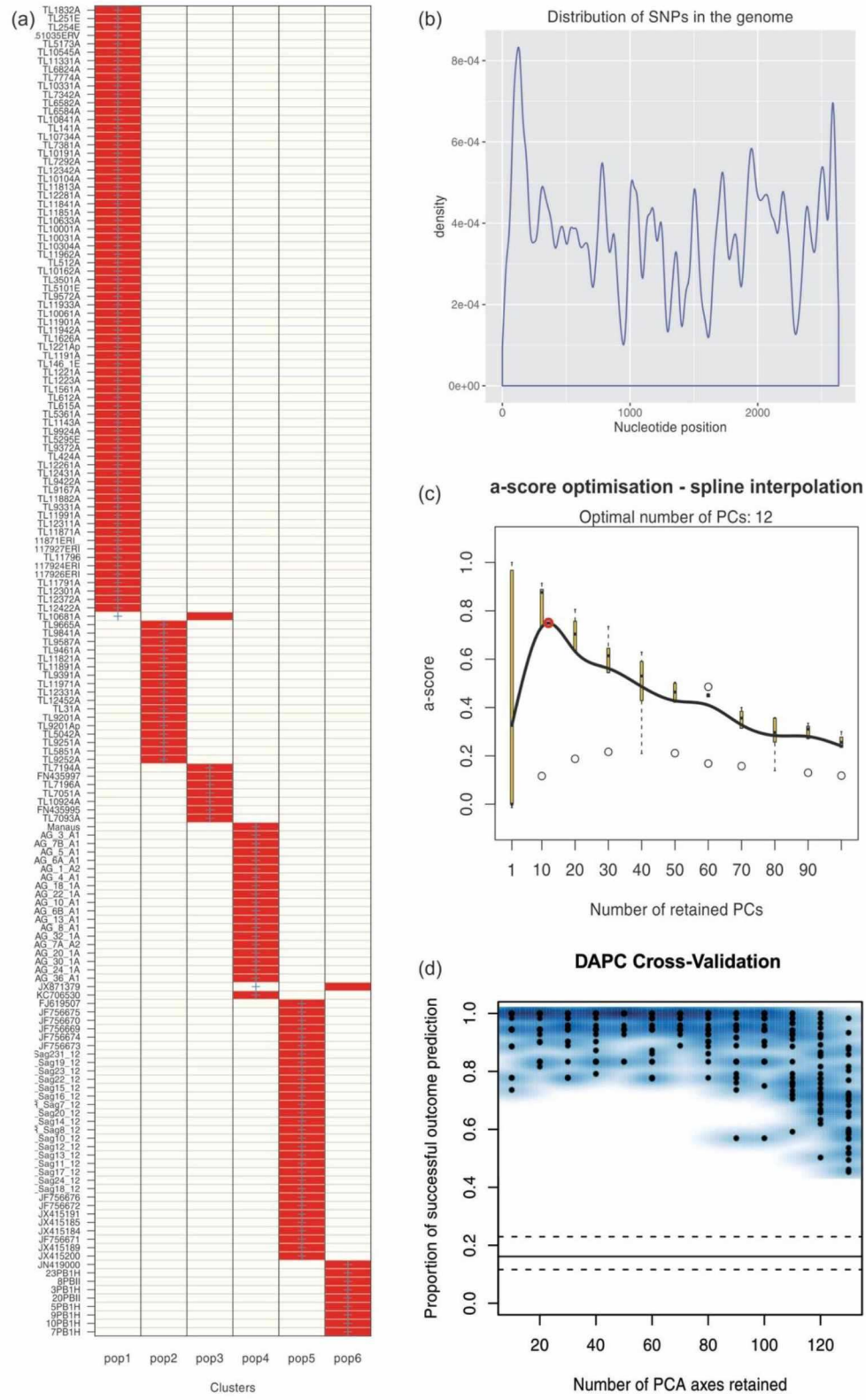
Supplementary Fig. S2a



Supplementary Fig. S2b



Supplementary Fig. S3



CHAPTER 2

INTERACTION BETWEEN THE NEW WORLD BEGOMOVIRUS *Euphorbia yellow mosaic virus* AND ITS ASSOCIATED ALPHASATELLITE: EFFECTS ON INFECTION AND TRANSMISSION BY THE WHITEFLY *Bemisia tabaci*

Mar, T.B., Mendes, I.R., Lau, D., Fiallo-Olivé, E., Navas-Castillo, J., Alves, M.S., Zerbini, F.M. 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, *submitted*.

INTERACTION BETWEEN THE NEW WORLD BEGOMOVIRUS *Euphorbia yellow mosaic virus* AND ITS ASSOCIATED ALPHASATELLITE: EFFECTS ON INFECTION AND TRANSMISSION BY THE WHITEFLY *Bemisia tabaci*

Talita B. Mar[1,2], Igor R. Mendes[1,2], Douglas Lau[3], Elvira Fiallo-Olivé[2,4], Jesús Navas-Castillo[2,4], Murilo S. Alves[1,2], F. Murilo Zerbini[1,2*]

[1]Dep. de Fitopatologia/BIOAGRO and [2]National Research Institute for Plant-Pest Interactions, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil
[3]Embrapa Trigo, Rodovia BR-285, 3081, Passo Fundo, RS, 99001-970, Brazil
[4]Instituto de Hortofruticultura Subtropical y Mediterránea 'La Mayora', Universidad de Málaga - Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), 29750 Algarrobo-Costa, Málaga, Spain

Running title: Alphasatellite modulates infection by a begomovirus

*Corresponding author:
Telephone: (+55-31) 3899-2935
E-mail: zerbini@ufv.br

Words in abstract: 229
Word in main text: 4,612
Number of tables: 2
Number of figures: 5

The sequences reported in this paper have been deposited in GenBank with access numbers KY559637-KY559646.

Abstract

The majority of Old World monopartite begomoviruses (family *Geminiviridae*) are associated with satellite DNAs. Alphasatellites are capable of autonomous replication, but depend on the helper virus for movement, encapsidation and transmission by the insect vector. Recently, Euphorbia yellow mosaic alphasatellite (EuYMA) was found in association with *Euphorbia yellow mosaic virus* (EuYMV) infecting *Euphorbia heterophylla* plants in Brazil. We assessed the geographical range of EuYMA in a representative sampling of *E. heterophylla* plants collected in several states of Brazil from 2009 to 2014. Infectious clones were generated and used to assess the phenotype of viral infection in the presence or absence of the alphasatellite in tomato, *E. heterophylla*, *Nicotiana benthamiana, Arabidopsis thaliana* and *Crotalaria juncea*. Phenotypic differences of EuYMV infection in the presence or absence of EuYMA were observed in *A. thaliana*, *N. benthamiana* and *E. heterophylla*. Symptoms were more severe when EuYMV was inoculated in combination with EuYMA in *N. benthamiana* and *E. heterophylla*, and the presence of the alphasatellite was determinant for symptom development in *A. thaliana*. The quantification of EuYMV and EuYMA indicated that EuYMA affects the accumulation of EuYMV during infection on a host-dependent basis. Transmission assays indicated that EuYMA negatively affects the transmission of EuYMV by *Bemisia tabaci* MEAM1. Together, these results indicate that EuYMA is capable of modulating symptoms, viral accumulation and whitefly transmission of EuYMV, potentially interfering with virus dissemination in the field.

Keywords: alphasatellite, begomovirus, geminivirus, non-cultivated hosts, whitefly

Introduction

The family *Geminiviridae* consists of plant viruses with genomes comprised of one or two molecules of circular, single-stranded (ss) DNA encapsidated by a single structural protein into twinned, quasi-icosahedral particles. The family includes nine genera based on the type of insect vector, host range, genome organization and phylogenetic relationships [1-3]. Viruses in the genus *Begomovirus* are transmitted by whiteflies of the *Bemisia tabaci* cryptic species complex (Hemiptera: Aleyrodidae) and infect dicot plants. Most begomoviruses fit into two major groups, Old World (OW; Europe, Africa, Asia and Australasia) and New World (NW; the Americas) based on genome features, phylogeny and geographical distribution [4, 5]. The vast majority of NW begomoviruses possess a bipartite genome comprised of two ssDNA molecules named DNA-A and DNA-B [4, 6]. OW begomoviruses can be either mono- or bipartite.

Monopartite OW begomoviruses are frequently associated with circular ssDNA satellites or satellite-like DNA molecules of half the size of begomovirus genome components, named betasatellites and alphasatellites [7]. Recently, a third class of DNA satellites, the deltasatellites, has been described in association with begomoviruses [8]. DNA satellites require a helper begomovirus for replication (except in the case of alphasatellites), systemic spread in the plant and insect transmission.

Betasatellites (previously known as DNA-β) have a highly conserved genome organization, including a conserved region among all betasatellites, named satellite conserved region (SCR), an A-rich region and a single gene in the complementary-sense encoding the betaC1 protein [9]. BetaC1 is a pathogenicity determinant [10, 11], and acts as a transcriptional and post-transcriptional suppressor of gene silencing [12, 13]. The presence of betasatellites is usually determinant for the induction of symptoms by the helper begomovirus [7, 14]. Betasatellites have never been reported in association with NW begomoviruses.

Alphasatellites (previously known as DNA-1) are frequently associated with OW monopartite begomovirus/betasatellite complexes [15]. The genome of alphasatellites contains a single gene in the virion sense, encoding a replication-associated protein named alpha-Rep. Alpha-Rep is similar to the master-Rep

protein encoded by the DNA-R component of nanoviruses (family *Nanoviridae*). Therefore, alphasatellites can replicate autonomously in plant cells. The stem-loop structure predicted for alphasatellites contains the nonanucleotide sequence TAGTATTAC, identical to that of nanoviruses [7].

Although alphasatellites were discovered more than 15 years ago, little is known about their effects on begomovirus infection. They are generally described as having no obvious effect on symptoms induced by begomoviruses or by begomovirus/betasatellite complexes, and may affect betasatellite replication, but not replication of the helper begomovirus [7, 16]. However, there is evidence that alphasatellites may be involved in pathogenicity. The alpha-Rep of two alphasatellite isolates can suppress post-transcriptional gene silencing [17]. Also, an alphasatellite isolate from Oman was shown to attenuate symptoms by reducing betasatellite accumulation [18]. This alphasatellite from Oman is more closely related to an unusual (DNA-2-type) alphasatellite reported in Singapore, and both are phylogenetically divergent from typical (DNA-1-type) alphasatellites [18, 19].

Recently, alphasatellites have been found in association with NW begomoviruses infecting non-cultivated plants in Brazil [20] and Cuba [21], and watermelon in Venezuela [22]. Alphasatellites were also detected by vector-enabled metagenomics (VEM) in samples from Guatemala and Puerto Rico [23]. The NW alphasatellites are more closely related to DNA-2-type alphasatellites and were classified as a third phylogenetically distinct group (DNA-3-type) by Rosario *et al.* [23].

In Brazil, DNA-3-type alphasatellites were found in *Euphorbia heterophylla* (Euphorbiaceae) and *Cleome affinis* (Cleomaceae) plants infected by bipartite begomoviruses [20]. Interestingly, and in contrast to the DNA-2-type alphasatellite from Oman, whose presence attenuated symptoms induced by the helper begomovirus [18], the DNA-3-type alphasatellites described in Brazil were reported to increase symptom severity [20]. The DNA-3-type alphasatellites had 51%-55% amino acid identity with DNA-2-type alphasatellites from Singapore and Oman [23].

Understanding the dynamics of the interaction between begomoviruses and DNA satellites in non-cultivated plants is important as they can be transferred to cultivated plants. Although sap transmission of a begomovirus/alphasatellite

complex was demonstrated for the DNA-3-type alphasatellite described in Venezuela [24], transmission studies by *B. tabaci* are lacking in the literature. In free-choice tests, adult preference and oviposition indicated that *E. heterophylla* was the most suitable host for *B. tabaci* among seven non-cultivated plants tested [25]. Given the polyphagous habit and preference of *B. tabaci* for *E. heterophylla*, the acquisition of DNA satellites associated with bipartite begomoviruses could be facilitated and may result in the emergence and spread of new disease complexes [7, 25, 26].

The objectives of this study were to assess the occurrence and prevalence of satellites in *E. heterophyll*a plants infected by begomoviruses, and to investigate their influence in the infection phenotype, viral accumulation and transmission. We assessed the geographical range of Euphorbia yellow mosaic alphasatellite (EuYMA) in Brazil and described its phylogenetic relationship with typical (DNA-1-type) and unusual (DNA-2- and DNA-3-type) alphasatellites. Dimeric constructs of *Euphorbia yellow mosaic virus* (EuYMV) genomic components (DNA-A and DNA-B) and EuYMA were generated to perform the biological characterization. Phenotypic differences of EuYMV infection in the presence of EuYMA were found in *Arabidopsis thaliana*, *Nicotiana benthamiana* and *E. heterophylla*. Whitefly transmission assays indicated that EuYMA negatively affects the transmission of EuYMV, potentially affecting the spread of the virus in the field.

Results

Geographical range and phylogeny of EuYMV and EuYMA

We investigated 165 *E. heterophylla* samples, collected over a period of six years in locations throughout the states of Amazonas, Goiás, Mato Grosso do Sul, Minas Gerais, Paraná, Pernambuco, Rio Grande do Sul and Santa Catarina, for the presence of begomoviruses and associated DNA satellites. While EuYMV DNA-A and DNA-B were cloned from 129 and 41 samples, respectively (Mar *et al. - accompanying manuscript*), EuYMA was detected in only six samples collected in the states of Rio Grande do Sul and Paraná (Fig. 1a; Suppl. Table S1). Pairwise sequence comparisons indicated ≥93% identity amongst EuYMA sequences (Suppl. Fig. S1; Suppl. Table S1).

A Bayesian phylogenetic tree based on complete nucleotide sequences separated EuYMA isolates according to sampling location (Fig. 1a,b). A tree based on nucleotide sequences of the *alpha-Rep* gene showed two major clades, confirming the phylogenetic distinction between typical (DNA-1-type) and unusual (DNA-2-type and DNA-3-type) alphasatellites (Fig. 1c). A tree based on deduced amino acid sequences had equivalent topology (Suppl. Fig. S2). The *alpha-Rep* genes of DNA-2-type and DNA-3-type alphasatellites share ≥58% pairwise identity amongst themselves. Besides Croton yellow vein mosaic alphasatellite, Tomato leaf curl New Delhi alphasatellite and NW alphasatellites (DNA-3-type) [23], other molecules from India, Pakistan, Sudan, Cameroon and Guatemala were also included in the "unusual alphasatellites" clade (Fig. 1c).

The EuYMA *alpha-Rep* gene is closely related to those from other NW alphasatellites (Fig. 1c). These NW alphasatellites share ≥79% pairwise identity amongst themselves and are phylogenetically distinct from OW alphasatellites. Only two sequences sampled in the Americas are not placed in the NW alphasatellite clade. Both were detected in insects, alphasatellite-1 reported from whiteflies feeding on tomato plants in Guatemala, and dragonfly-associated alphasatellite detected in dragonflies from Puerto Rico [23, 27].

The presence of EuYMA influences the EuYMV infection phenotype in selected hosts

The influence of EuYMA on the EuYMV infection phenotype was investigated in five hosts: *Solanum lycopersicum* (tomato), *E. heterophylla*, *N. benthamiana, A. thaliana* and *Crotalaria juncea*. Tomato and *C. juncea* showed no symptoms of viral infection (Table 1). Nevertheless, both EuYMV infection alone (6/45 tomato and 3/45 *C. juncea*) and with the presence of EuYMA (4/45 tomato and 4/45 *C. juncea*) were detected by PCR at 14 and 28 dpi (Table 1; Suppl. Table S2). On average, the number of infected tomato and *C. juncea* plants (approximately 10% of total inoculated plants) was much smaller compared with *E. heterophylla*, *N. benthamiana* and *A. thaliana* in both treatments (Fig. 2a).

A. thaliana showed no symptoms of EuYMV infection (Table 1). Nevertheless, the virus was detected by PCR in 31 out of 45 plants infected with EuYMV alone at 28 dpi (Fig. 2a; Suppl. Table S2). Strikingly, severe symptoms were observed upon virus infection in the presence of EuYMA (Table 1; Fig. 3).

The presence of the alphasatellite was determinant to the development of severe curling in *A. thaliana* leaves (Fig. 3; Suppl. Fig. S3). The first symptoms were observed at 10 dpi, and leaf curling was observed in 96.5% of the infected plants at 28 dpi (Table 1; Fig. 2b). The number of EuYMV-infected plants in the presence of EuYMA was statistically different comparing 14 and 28 dpi (17 and 29 out of 45 inoculated plants, respectively), suggesting that the presence of EuYMA may delay infection development in *A. thaliana* (Table 1; Fig. 2a,b).

The first symptomatic *N. benthamiana* plants were observed at 5 dpi. At the end of the experiment (28 dpi) EuYMV was detected in 87% of the inoculated plants (39 out of 45 plants), while only 40% were infected with EuYMV and EuYMA (18 out of 45) (Table 1; Fig. 2a). At both time points, a lower number of plants were infected with EuYMV in the presence of EuYMA compared to EuYMV alone (Table 1; Fig. 2a; Suppl. Table S2). Also, from a total of 45 *N. benthamiana* inoculated with both virus and alphasatellite, EuYMV was detected alone in six plants. Symptoms of vein chlorosis and leaf deformation were observed in approx. 90% of the plants infected with EuYMV (Table 1; Fig. 2b). In general, the severity of symptoms increased in the presence of EuYMA (Fig. 3; Suppl. Fig. S4). Moreover, in the absence of EuYMA, some plants recovered from the symptoms (Suppl. Fig. S3).

The same number of infected *E. heterophylla* plants was detected by PCR at either 14 or 28 dpi. When EuYMV was inoculated alone, infection was detected in 40% of the plants (18/45), while inoculation of the two agents lead to infection in approx. 30% of the plants (13/45) (Fig. 2a; Table 1; Suppl. Table S2). Similar to *N. benthamiana,* from a total of 45 *E. heterophylla* inoculated with both virus and alphasatellite, EuYMV alone was detected in six plants (data not shown). The first symptoms were observed at 4 dpi and by 28 dpi all infected plants showed symptoms of mosaic and vein chlorosis (Fig. 2b; Table 1). A significant difference in symptoms was observed in the presence of EuYMA. In addition to mosaic and vein chlorosis, leaf curling was also observed in *E. heterophylla* infected with both EuYMV and EuYMA (Fig. 3; Suppl. Fig. S3).

The presence of EuYMA influences EuYMV accumulation in three hosts

Considering the influence of EuYMA in the EuYMV infection phenotype in *A. thaliana*, *E. heterophylla* and *N. benthamiana,* we quantified the

accumulation of each viral genomic component (DNA-A and DNA-B) at 14 and 28 dpi to compare viral accumulation in the presence or absence of EuYMA.

The accumulation of EuYMV DNA-A increased in *E. heterophylla* and *N. benthamiana* in the presence of EuYMA, compared with plants infected with the virus alone at both 14 and 28 dpi (Fig. 4; Suppl Fig. S5). Viral accumulation varied widely among plants of *E. heterophylla*. Also, the number of *E. heterophylla* infected plants was lower compared with *A. thaliana* and *N. benthamiana*, impairing the statistical analysis (Suppl. Fig. S5). Interestingly, and unlike what was observed in *E. heterophylla* and (more convincingly) in *N. benthamiana,* the presence of EuYMA did not cause an increase in the accumulation of EuYMV DNA-A in *A. thaliana* (Fig. 4). In fact, the results point to a reduction, albeit with weak statistical support (Fig. 4; Suppl. Fig. S5).

The results did not indicate significant differences in the accumulation of EuYMV DNA-B in the presence or absence of EuYMA in *N. benthamiana* and *A. thaliana* (Fig. 4; Suppl. Fig. S5). However, comparisons in *E.heterophylla* pointed to an increase of EuYMV DNA-B accumulation in the presence of EuYMA at both 14 and 28 dai (Fig. 4).

EuYMA accumulated at high levels in the systemically infected leaves of *A. thaliana*, with a significantly higher accumulation compared to both EuYMV components at 28 dpi (Suppl. Fig. S6). In *N. benthamiana* and *E. heterophylla*, EuYMA accumulated at higher levels than EuYMV DNA-B (Suppl. Fig. S6).

EuYMA negatively affects the transmission of EuYMV by the insect vector

We compared the transmission of EuYMV by *B. tabaci* MEAM1 in the presence or absence of EuYMA. The plants were PCR-tested for the presence of the EuYMV genomic components and for EuYMA. In both replications of the experiment (for a total of 40 inoculated plants in each treatment), EuYMV was transmitted more efficiently when the source plants were infected with EuYMV alone (Table 2). This result indicates that EuYMA could negatively affect the vector transmission of EuYMV.

Similarly to what was observed in the host range assay (when the plants were inoculated by biolistics), *E. heterophylla* plants infected with EuYMV alone showed symptoms of interveinal chlorosis and yellow mosaic (Fig. 5a). When the virus was inoculated in combination with EuYMA the symptoms were more

severe, characterized by a more intense yellow mosaic, leaf shriveling and stunting (Fig. 5b,c). Interestingly, after 28 dpi, some plants recovered from the symptoms (EuYMV alone and EuYMV+EuYMA; nine and five plants, respectively) (Fig. 5d). The presence of all molecules was confirmed by PCR, with oligonucleotides for the *CP* and *MP* genes of EuYMV and for the *alpha-Rep* of EuYMA (Fig. 5f).

Interestingly, in the transmission assay, only the EuYMV DNA-A was detected in some *E. heterophylla* plants, at both time points (in one plant the DNA-A was detected at 14 dpi but no genomic components were detected at 28 dpi). This was observed in three plants in each treatment, in the two experiments (3 out of 40 plants inoculated with EuYMV alone and 3 out of 40 plants in which EuYMV was inoculated together with EuYMA). Mild mosaic symptoms were observed in these plants (Fig. 5e). For the purpose of statistical analysis, these plants were considered negative for the presence of the virus.

Discussion

In a representative sampling encompassing six years and locations throughout eight Brazilian states, EuYMA was detected in only six samples collected in the states of Rio Grande do Sul and Paraná. It seems clear that the association of EuYMV with EuYMA is not common. Nevertheless, EuYMA was detected in four additional locations besides the previously reported one (in the state of Mato Grosso do Sul; [20]), in agreement with the large geographical range of alphasatellites in South America. A recent study corroborates these conclusions. Ferro *et al.* [28] detected EuYMA in one out of 43 samples of *Sida* spp. collected in an area overlapping the one covered in our study. Although the associated begomovirus was not identified at the species level, EuYMV was detected in two other *Sida* spp. samples, indicating that EuYMV can infect *Sida* spp. Together, these results suggest that the association between the EuYMV and EuYMA could be specific.

Pairwise sequence comparisons indicated that alphasatellite sequences were more variable than the helper begomovirus components (Mar *et al.* - *accompanying manuscript*). The EuYMA *alpha-Rep* gene is most closely related

to those from alphasatellites found in association with bipartite begomoviruses in the Americas (DNA-3-type), corroborating the analysis of Rosario *et al.* [23].

The presence of EuYMA correlated with an increase of symptom severity in *N. benthamiana* and *E. heterophylla*, and was determinant for the development of severe leaf curling in *A. thaliana*. Also, the EuYMV/EuYMA complex can systemically infect *C. juncea* and tomato. Besides the characteristic mosaic symptoms and interveinal chlorosis induced by EuYMV in *E. heterophylla*, leaf shriveling, downward leaf rolling and stunting symptoms were observed when EuYMA was present. This is the opposite of what was observed for the DNA-2-type Ageratum yellow vein alphasatellite from Oman, which is associated with begomovirus/betasatellite complexes. In this case, the presence of the alphasatellite leads to attenuated symptoms in *N. benthamiana* and substantially reduced betasatellite accumulation [18]. Symptom attenuation is likely due to reduced betasatellite titer, leading to lower accumulation of the betaC1 protein, a suppressor of gene silencing [12, 18].

One possible explanation for the symptoms of EuYMV being more severe in combination with EuYMA is that the alpha-Rep protein of EuYMA may possess silencing suppressor activity, as shown by Nawaz-ul-Rehman *et al.* [17] for alpha-Rep proteins encoded by two alphasatellites from Pakistan. However, preliminary experiments did not support a role as *p*ost-transcriptional gene silencing suppressor for the alpha-Rep protein of EuYMA in *N. benthamiana* leaves (*data not shown*). It is possible that it may suppress transcriptional silencing, as recently reported for the alpha-Rep of Cotton leaf curl Multan alphasatellite [29].

The presence of EuYMA affects the accumulation of EuYMV in *E. heterophylla* and *N. benthamiana*. Considering all classes of DNA satellites which can be found in association with begomoviruses, an increase in helper begomovirus accumulation is known to occur only in the presence of betasatellites [30]. The recently described deltasatellites have been shown to decrease accumulation of viral genome components in agroinoculated *N. benthamiana* plants [31, 32], and some of the unusual alphasatellites can affect only betasatellite replication without affecting the helper virus [17, 18]. Therefore, at this time, the pathosystem *E. heterophylla*/EuYMV/EuYMA is the only one in which the presence of an alphasatellite increases symptom severity and DNA

accumulation of the helper begomovirus. The mechanism(s) by which alphasatellites act during begomovirus infection remains unknown [7]. For EuYMV/EuYMA, both symptom modulation and changes in DNA accumulation were observed in the three hosts, emphasizing the need for further studies to determine the role of alphasatellites in begomovirus infections.

We attempted to clarify the role of alphasatellites in the transmission of begomoviruses by their insect vector, and found that the presence of EuYMA negatively affects the transmission of EuYMV by *B. tabaci* MEAM1 (the prevalent species in Brazil; [33]). Alphasatellites are capable of autonomous replication in plant cells as they encode their own, nanovirus-like, Rep. However, they do not encode a coat protein gene, and thus depend on their helper begomovirus for insect transmission [7]. It is possible that alphasatellites compete with begomoviruses for encapsidation of the genome. Thus, depending on the relative amount of alphasatellite and begomovirus DNA, many particles may contain alphasatellite molecules instead of viral genomic components, reducing the chances of virus particles being acquired and transmitted by the vector. In agreement with this hypothesis, we observed a high level of EuYMA accumulation in the upper leaves of *A. thaliana*, *N. benthamiana* and *E. heterophylla*. Transmission was favored in this experiment by using thirty whiteflies per plant, since the number of infected plants increases as whitefly density increases [34]. We obtained a statistically significant difference between the number of infected plants in the presence and absence of EuYMA. Studies using gradually lower densities of whiteflies and also a higher number of plants could lead to even greater differences, providing more conclusive results.

EuYMA was efficiently transmitted by *B. tabaci* MEAM1, being detected in all plants in which EuYMV (DNA-A+DNA-B) was also detected. Thus, the reason why it is not widespread in the field is probably not due to low efficiency in the transmission by the insect vector. The encapsidation of EuYMA probably did not alter the structural properties of the CP (the only essential protein for vector transmission), and thus did not affect the CP interaction with GroEL chaperones produced by endosymbionts, which is indispensable for particle stability during their passage through the gut of whiteflies [35-38].

We detected a small number of *E. heterophylla* plants infected by EuYMV DNA-A alone, displaying mild mosaic symptoms. Systemic infection by

begomovirus DNA-A alone is not without precedent. Previous reports indicate that, in exceptional cases, the DNA-A is capable of infecting and moving in plants in the absence of the DNA-B [39-41]. Weigel *et al.* [42] demonstrated that the DNA-A can reach the cell nucleus of upper, systemically infected leaves and replicate therein, independently of the DNA-B. However, such infection is usually asymptomatic. The occurrence of mild symptoms in *E. heterophylla* plants infected by EuYMV DNA-A alone suggests that the virus is very well adapted to this host, while at the same time reinforcing the role of the DNA-B on symptom induction by bipartite begomoviruses.

Our results indicate that the association of the alphasatellite EuYMA with the bipartite begomovirus EuYMV is not common, although their geographical range is large. We also provide evidence that alphasatellites can increase the symptom severity of bipartite begomovirus infections and promote changes in DNA accumulation, although the mechanisms have to be elucidated. The presence of EuYMA negatively interfered with transmission by the vector, which can negatively affect the spread of the virus. EuYMV seems to be well adapted to *E. heterophylla*, and can also infect a number of other non-cultivated hosts. This is an interesting system to study the interaction between begomovirus and DNA satellites in non-cultivated hosts, and to assess the possibility and consequences of them being transferred to cultivated plants.

Methods

Cloning and sequencing of DNA satellite genomes

The same samples of *E. heterophylla* described for the genetic variability study of EuYMV (Mar *et al.*, *accompanying paper*) were used for the detection and cloning of DNA satellites. Circular DNA molecules were amplified using rolling-circle amplification (RCA) with phi29 DNA polymerase according to Inoue-Nagata *et al.* [43]. Restriction fragments of approx. 1,300 nucleotides (nt), corresponding to one genomic copy of satellites DNAs, were ligated to the pBLUESCRIPT-KS+ (Stratagene) plasmid vector. Inserts of selected clones were completely sequenced at Macrogen, Inc. (Seoul, South Korea). Pairwise sequence comparisons were performed with Sequence Demarcation Tool (SDT) v. 1.2 [44].

Phylogenetic analysis

Multiple sequence alignments were performed using the MUSCLE algorithm [45] implemented in MEGA 6 [46]. Phylogenetic reconstruction using Bayesian inference was performed with MrBayes 3.2 [47] available at the CIPRES Science Gateway [48], with the models selected by MrModeltest2.2 [49] in the Akaike Information Criterion (AIC). Two independent analyses were conducted, each one running at least 10,000,000 generations. Phylogenetic trees were visualized using FigTree 1.3 (tree.bio.ed.ac.uk/software/figtree).

Construction of infectious clones and host range assay

Tandem dimeric constructs of each genomic component of EuYMV-[BR-Cha510-10] (Mar *et al.*, *accompanying paper*) and of the associated EuYMA (Suppl. Table S1) were obtained by partial restriction digestion [50] and were used in host range assays using biolistics *[51]*. Fifteen plants each of tomato (cv. Santa Clara), *E. heterophylla*, *N. benthamiana, A. thaliana* and *C. juncea* were inoculated with EuYMV alone (DNA-A+DNA-B), and 15 were inoculated with EuYMV and EuYMA. Three independent replications of the experiment were conducted, for a total of 45 plants of each species/treatment. The youngest upper leaves were collected at 14 and 28 days post-inoculation (dpi) for total DNA extraction [52]. Plants were also visually evaluated for symptoms and PCR-tested for the presence of virus and satellite at both time points.

Virus and alphasatellite titers were determined by quantitative PCR. Total DNA was quantified using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific). A standard curve was obtained by dilution of known quantities of plasmids containing one copy of the EuYMV DNA-A and DNA-B and of the EuYMA genome. Primers were designed with Primer Express Software 3.0.1 (Applied-Biosystems), based on the sequences of the EuYMV *CP* gene (to estimate the DNA-A titer; 5'-AAG GCC TCT TCA TGG GTG AA-3' and 5'-TCC GGT ACA TCT GGG CCT CTA-3'), the EuYMV *MP* gene (to estimate the DNA-B titer; 5'-TTG GCG CTC TGA AAG GGA TA-3' and 5'-GAC ACA TTT CGC GAA GTT CAA-3') and the EuYMA *alpha-Rep* gene (to estimate the alphasatellite titer; 5'-ACA AGA GGA AGA CGA TGG AAC GT-3' and 5'-GCA GCG ACG ATA CAG CTT AGG-3'). Quantitative PCR was carried out using Fast SYBR Green Master Mix in a StepOnePlus Real-Time PCR System (Applied

Biosystems). Each sample was analyzed in triplicate, in a reaction containing 0.1 ng of total DNA. Data analysis was performed with GraphPad Prism Software 5 [53] and the significance between means was calculated with Student's *t* test.

Whitefly transmission assay

Seedlings of *E. heterophylla* (2-3 days post-germination) were biolistically inoculated with EuYMV (DNA-A+DNA-B) alone or with EuYMV and EuYMA. Total DNA was extracted 14 days post-inoculation [52]. The presence of the two agents was assayed by conventional PCR with oligonucleotides for the *CP* and *MP* genes of EuYMV and the alpha-*Rep* gene of EuYMA. One set of plants infected by the virus alone and one set infected by the virus and the alphasatellite were used as sources for acquisition by *B. tabaci* Middle East-Asia Minor 1 (MEAM1), as described by Polston and Capobianco [54]. Non-viruliferous whitefly adults (maintained in cabbage plants) were released in insect-proof cages containing the source plants. After an acquisition access period (AAP) of 48 hours, the whiteflies were transferred to healthy *E. heterophylla* plants with the first pair of true leaves fully expanded. Each plant was maintained in an insect-proof cage with 30 whiteflies per cage for an inoculation access period (IAP) of 48 hours. Twenty plants were used in each treatment (EuYMV transmission in the presence or absence of EuYMA), with three healthy plants as negative controls. Two independent replications of the experiment were performed.

Total DNA was extracted at 14 and 28 days post whitefly transmission and the presence of the two agents was confirmed by PCR as described above. Transmission rates were compared using a 2×2 contingency table constructed with counting data obtained by qualitative evaluation of plants (PCR-positive or PCR-negative). To infer if the variation found in the data is dependent or not of the tested factor (in this case the presence or not of EuYMA in co-infection with EuYMV) the independence between rows and columns were tested using an unconditional test, a more powerful alternative than a conditional test. As recommended by Mato and Andrés [55], independence test statistics were performed with the *exact.test* from the R package, through Barnard's unconditional *t*est [34, 56].

Funding information

This work was funded by the Science Without Borders (SWB) program of CNPq (grant 401838/2013-7 to FMZ and JNC), by FAPEMIG (grant CAG-APQ-02037-13 to FMZ) and by Embrapa (grant SEG 02.08.01.008.00.00).

Conflicts of interest

The authors declare that there are no conflicts of interest.

Acknowledgements

JNC was the recipient of a CNPq Special Visiting Scientist fellowship under the SWB program. EFO was the recipient of a visiting scientist fellowship from Fapemig (BPV-00060-13). TBM was the recipient of a CNPq doctoral fellowship.

Ethical statement

The research reported in this paper did not involve animals or humans.

References

1. Varsani A, Roumagnac P, Fuchs M, Navas-Castillo J, Moriones E, Idris A, et al. *Capulavirus* and *Grablovirus*: two new genera in the family *Geminiviridae*. Arch Virol. 2017;162:doi:10.1007/s00705-017-3268-6.
2. Varsani A, Navas-Castillo J, Moriones E, Hernández-Zepeda C, Idris A, Brown JK, et al. Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. Arch Virol. 2014;159:2193-203.
3. Brown JK, Zerbini FM, Navas-Castillo J, Moriones E, Ramos-Sobrinho R, Silva JC, et al. Revision of *Begomovirus* taxonomy based on pairwise sequence comparisons. Arch Virol. 2015;160:1593-619.
4. Harrison BD, Robinson DJ. Natural genomic and antigenic variation in white-fly transmitted geminiviruses (begomoviruses). Annu Rev Phytopathol. 1999;39:369-98.
5. Rojas MR, Hagen C, Lucas WJ, Gilbertson RL. Exploiting chinks in the plant's armor: Evolution and emergence of geminiviruses. Annu Rev Phytopathol. 2005;43:361-94.
6. Briddon RW, Patil BL, Bagewadi B, Nawaz-ul-Rehman MS, Fauquet CM. Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. BMC Evol Biol. 2010;10:97.
7. Zhou X. Advances in understanding begomovirus satellites. Annu Rev Phytopathol. 2013;51:357-81.
8. Lozano G, Trenado HP, Fiallo-Olive E, Chirinos D, Geraud-Pouey F, Briddon RW, et al. Characterization of non-coding DNA satellites associated with sweepoviruses (genus *Begomovirus*, *Geminiviridae*) -

definition of a distinct class of begomovirus-associated satellites. Front Microbiol. 2016;7:162.
9. Briddon RW, Brown JK, Moriones E, Stanley J, Zerbini FM, Zhou X, et al. Recommendations for the classification and nomenclature of the DNA-beta satellites of begomoviruses. Arch Virol. 2008;153:763-81.
10. Cui XF, Tao XR, Xie Y, Fauquet CM, Zhou XP. A DNA beta associated with *Tomato yellow leaf curl China virus* is required for symptom induction. J Virol. 2004;78:13966-74.
11. Bhattacharyya D, Gnanasekaran P, Kumar RK, Kushwaha NK, Sharma VK, Yusuf MA, et al. A geminivirus betasatellite damages the structural and functional integrity of chloroplasts leading to symptom formation and inhibition of photosynthesis. J Exp Bot. 2015;66:5881-95.
12. Yang XL, Xie Y, Raja P, Li SZ, Wolf JN, Shen QT, et al. Suppression of methylation-mediated transcriptional gene silencing by betaC1-SAHH protein interaction during geminivirus-betasatellite infection. PLoS Pathog. 2011;7:e1002329.
13. Li F, Huang C, Li Z, Zhou X. Suppression of RNA silencing by a plant DNA virus satellite requires a host calmodulin-like protein to repress *RDR6* expression. PLoS Pathog. 2014;10:e1003921.
14. Kumar J, Singh SP, Kumar A, Khan JA, Tuli R. Detection and characterization of a new betasatellite: variation in disease symptoms of tomato leaf curl Pakistan virus-India due to associated betasatellite. Arch Virol. 2013;158:257-61.
15. Leke WN, Mignouna DB, Brown JK, Kvarnheden A. Begomovirus disease complex: Emerging threat to vegetable production systems of West and Central Africa. Agric Food Sec. 2015;4:1.
16. Saunders K, Stanley J. A nanovirus-like DNA component associated with yellow vein disease of *Ageratum conyzoides*: Evidence for interfamilial recombination between plant DNA viruses. Virology. 1999;264:142-52.
17. Nawaz-ul-Rehman MS, Nahid N, Mansoor S, Briddon RW, Fauquet CM. Post-transcriptional gene silencing suppressor activity of two non-pathogenic alphasatellites associated with a begomovirus. Virology. 2010;405:300-8.
18. Idris AM, Shahid MS, Briddon RW, Khan AJ, Zhu JK, Brown JK. An unusual alphasatellite associated with monopartite begomoviruses attenuates symptoms and reduces betasatellite accumulation. J Gen Virol. 2011;92:706-17.
19. Saunders K, Bedford ID, Stanley J. Adaptation from whitefly to leafhopper transmission of an autonomously replicating nanovirus-like DNA component associated with ageratum yellow vein disease. J Gen Virol. 2002;83:907-13.
20. Paprotka T, Metzler V, Jeske H. The first DNA 1-like alphasatellites in association with New World begomoviruses in natural infections. Virology. 2010;404:148-57.
21. Jeske H, Kober S, Schäfer B, Strohmeier S. Circomics of Cuban geminiviruses reveals the first alpha-satellite DNA in the Caribbean. Virus Genes. 2014;49:312-24.
22. Romay G, Chirinos D, Geraud-Pouey F, Desbiez C. Association of an atypical alphasatellite with a bipartite New World begomovirus. Arch Virol. 2010;155:1843-7.

23. Rosario K, Marr C, Varsani A, Kraberger S, Stainton D, Moriones E, et al. Begomovirus-associated satellite DNA diversity captured through vector-enabled metagenomic (VEM) surveys using whiteflies (Aleyrodidae). Viruses. 2016;8:36.
24. Romay G, Lecoq H, Desbiez C. Melon chlorotic mosaic virus and associated alphasatellite from Venezuela: genetic variation and sap transmission of a begomovirus–satellite complex. Plant Pathol. 2015;64:1224-34.
25. Sottoriva LDM, Lourenção AL, Colombo CA. Performance of *Bemisia tabaci* (Genn.) biotype B (Hemiptera: Aleyrodidae) on weeds. Neotrop Entomol. 2014;43:574-81.
26. 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. Annu Rev Virol. 2015;2:67-93.
27. Rosario K, Padilla-Rodriguez M, Kraberger S, Stainton D, Martin DP, Breitbart M, et al. Discovery of a novel mastrevirus and alphasatellite-like circular DNA in dragonflies (Epiprocta) from Puerto Rico. Virus Res. 2013;171:231-7.
28. Ferro CG, Silva JP, Xavier CAD, Godinho MT, Lima ATM, Mar TB, et al. The ever increasing diversity of begomoviruses infecting non-cultivated hosts: new species from *Sida* spp. and *Leonurus sibiricus*, plus two New World alphasatellites. Ann Appl Biol. 2017:10.1111/aab.12329.
29. Abbas Q, Amin I, Mansoor S, Wassenegger M, Briddon RW. Geminivirus-associated alphasatellites suppress transcriptional not post-transcriptional gene silencing. VirusDis. 2016;27:419.
30. Saunders K, Bedford ID, Briddon RW, Markham PG, Wong SM, Stanley J. A unique virus complex causes *Ageratum* yellow vein disease. Proc Natl Acad Sci USA. 2000;97:6890-5.
31. Fiallo-Olivé E, Tovar R, Navas-Castillo J. Deciphering the biology of deltasatellites from the New World: maintenance by New World begomoviruses and whitefly transmission. New Phytol. 2016;212:680-92.
32. Hassan I, Orilio AF, Fiallo-Olive E, Briddon RW, Navas-Castillo J. Infectivity, effects on helper viruses and whitefly transmission of the deltasatellites associated with sweepoviruses (genus Begomovirus, family Geminiviridae). Sci Rep. 2016;6:30204.
33. Marubayashi JM, Yuki VA, Rocha KCG, Mituti T, Pelegrinotti FM, Ferreira FZ, et al. At least two indigenous species of the *Bemisia tabaci* complex are present in Brazil. J Appl Entomol. 2013;137:113-21.
34. Guo T, Guo Q, Cui X-Y, Liu Y-Q, Hu J, Liu S-S. Comparison of transmission of *Papaya leaf curl China virus* among four cryptic species of the whitefly *Bemisia tabaci* complex. Sci Rep. 2015;5:15432.
35. Azzam O, Frazer J, De La Rosa D, Beaver JS, Ahlquist PG, Maxwell DP. Whitefly transmission and efficient ssDNA accumulation of bean golden mosaic geminivirus require functional coat protein. Virology. 1994;204:289-96.
36. Hofer P, Bedford ID, Markham PG, Jeske H, Frischmuth T. Coat protein gene replacement results in whitefly transmission of an insect nontransmissible geminivirus isolate. Virology. 1997;236:288-95.
37. Morin S, Ghanim M, Zeidan M, Czosnek H, Verbeek M, van den Heuvel JF. A GroEL homologue from endosymbiotic bacteria of the whitefly

B*emisia tabaci* is implicated in the circulative transmission of tomato yellow leaf curl virus. Virology. 1999;256:75-84.
38. Morin S, Ghanim M, Sobol I, Czosnek H. The GroEL protein of the whitefly *Bemisia tabaci* interacts with the coat protein of transmissible and nontransmissible begomoviruses in the yeast two-hybrid system. Virology. 2000;276:404-16.
39. Klinkenberg FA, Stanley J. Encapsidation and spread of *African cassava mosaic virus* DNA-A in the absence of DNA-B when agroinoculated to *Nicotiana benthamiana*. J Gen Virol. 1990;71:1409-12.
40. Evans D, Jeske H. DNA B facilitates, but is not essential for, the spread of *Abutilon mosaic virus* in agroinoculated *Nicotiana benthamiana*. Virology. 1993;194:752-7.
41. Galvão RM, Mariano AC, Luz DF, Alfenas PF, Andrade EC, Zerbini FM, et al. A naturally occurring recombinant DNA-A of a typical bipartite begomovirus does not require the cognate DNA-B to infect *Nicotiana benthamiana* systemically. J Gen Virol. 2003;84:715-26.
42. Weigel K, Pohl JO, Wege C, Jeske H. A population genetics perspective on geminivirus infection. J Virol. 2015;89:11926-34.
43. Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T. A simple method for cloning the complete begomovirus genome using the bacteriophage phi29 DNA polymerase. J Virol Met. 2004;116:209-11.
44. Muhire BM, Varsani A, Martin DP. SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLoS ONE. 2014;9:e108277.
45. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinf. 2004;5:1-19.
46. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30:2725-9.
47. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61:539-42.
48. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proc Gateway Comp Environ Work (GCE). 2010:1-8.
49. Nylander JAA. MrModeltest v2. Program distributed by the author Evolutionary Biology Centre, Uppsala University2004.
50. Wu C-Y, Lai Y-C, Lin N-S, Hsu Y-H, Tsai H-T, Liao J-Y, et al. A simplified method of constructing infectious clones of begomovirus employing limited restriction enzyme digestion of products of rolling circle amplification. J Virol Met. 2008;147:355-9.
51. Aragão FJL, Barros LMG, Brasileiro ACM, Ribeiro SG, Smith FD, Sanford JC, et al. Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris* L.) co-transformed via particle bombardment. Theor Appl Genet. 1996;93:142-50.
52. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem Bull. 1987;19:11-5.
53. Swift ML. GraphPad Prism, data analysis, and scientific graphing. J Chem Inf Comp Sci. 1997;37:411-2.

54. Polston JE, Capobianco H. Transmitting plant viruses using whiteflies. J Vis Exp. 2013:e4332.
55. Mato AS, Andrés AM. Simplifying the calculation of the P-value for Barnard's test and its derivatives. Stat Comp. 1997;7:137-43.
56. Barnard G. A new test for 2× 2 tables. Nature. 1945;156:177.

Table 1. Experimental host range of *Euphorbia yellow mosaic virus* (EuYMV) and its associated alphasatellite, Euphorbia yellow mosaic alphasatellite (EuYMA).

	Arabidopsis thaliana				*Nicotiana benthamiana*				*Euphorbia heterophylla*				*Solanum lycopersicum*				*Crotalaria juncea*			
	14 dpi*		28 dpi		14 dpi		28 dpi		14 dpi		28 dpi		14 dpi		28 dpi		14 dpi		28 dpi	
Alphasatellite	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
Symptoms†	Ns	Cr	Ns	Cr	Vc, Ld	Vc, Ld, Mc	Vc, Ld	Vc, Ld, Mc	M, Vc	Cr, M, Vc	M, Vc	Cr, M, Vc	Ns	Ns	Ns	Ns	Ns	Ns	Ns	Ns
Virus detection‡	0/21	3/17	0/31	28/29	34/38	16/16	31/39	16/18	16/18	13/13	18/18	13/13	0/6	0/4	0/6	0/4	0/3	0/4	0/3	0/4

* Days post-biolistic inoculation

† Ns: no symptoms; Cr: curling; Vc: vein chlorosis; Ld: leaf deformation; Mc: mild chlorosis; M: mosaic

‡ No. of symptomatic plants / No. of PCR-positive plants at 14 and 28 dpi, in a total of 45 plants (15 plants of each of three independent replications)

Table 2. Transmission of *Euphorbia yellow mosaic virus* (EuYMV) to *Euphorbia heterophylla* plants, alone or in the presence of Euphorbia yellow mosaic alphasatellite (EuYMA), by *Bemisia tabaci* Middle East-Asia Minor 1 (MEAM1).

Experiment	No of infected plants/ No of inoculated plants		
	EuYMV[§]	EuYMV and EuYMA[†]	Negative control[‡]
1	17/20	15/20*	0/3
2	18/20	14/20*	0/3

[§]PCR detection of EuYMV DNA-A and DNA-B in plants inoculated with EuYMV alone

[†]PCR detection of EuYMV DNA-A and DNA-B and EuYMA in plants inoculated with EuYMV and EuYMA

[‡]Transmission with aviruliferous whiteflies

*Statistical significance ($p \leq 0.05$) according to Barnard's unconditional test.

Figure legends

Fig. 1. (a) Map of Brazil indicating the locations where Euphorbia yellow mosaic alphasatellite has been detected in the states of Rio Grande do Sul, Paraná (light and medium grey, respectively; this study) and Mato Grosso do Sul (dark grey; Paprotka et al., 2010). **(b, c)** Midpointed-rooted Bayesian phylogenetic trees based on full-length nucleotide sequences of EuYMA **(b)** and *alpha-Rep* gene nucleotide sequences **(c)**. Nodes to the right of branches with posterior probabilities equal to or higher than 0.8 are indicated by filled circles and those with values lower than 0.8 and higher than 0.5 by empty circles. In **(b)**, the vertical bar at the right indicate the geographical regions represented in the map. In **(c)**, the vertical bar indicates Old World and New World begomoviruses, and the classification proposed by Rosario *et al.* [23] is shown in brackets.

Fig. 2. Experimental host range of *Euphorbia yellow mosaic virus* (EuYMV) and its associated alphasatellite, Euphorbia yellow mosaic alphasatellite (EuYMA). **(a)** The total number of infected plants (detected by PCR) with standard errors is indicated for each host at 14 and 28 days post-inoculation (dpi) in the presence (+) or absence (-) of EuYMA (three independent replications, 15 plants per replication). Statistically significant differences (Student's *t* test) are indicated with vertical bars (*, $p \leq 0.05$). **(b)** Percentage of PCR-positive plants showing symptoms at 14 and 28 dpi.

Fig. 3. Symptoms induced by *Euphorbia yellow mosaic virus* (EuYMV) alone or in the presence of Euphorbia yellow mosaic alphasatellite (EuYMA) at 14 and 28 days post-inoculation (dpi) by biolistics.

Fig. 4. Absolute quantification by quantitative PCR of *Euphorbia yellow mosaic virus* (EuYMV) DNA-A and DNA-B and of Euphorbia yellow mosaic alphasatellite (EuYMA) in different hosts. Viral accumulation is presented as the log of the number of molecules, and values below the detection threshold were considered as background. Each circle represents one infected plant. (-) represents infection by EuYMV alone; (+) represents infection by EuYMV together with EuYMA. Data refers to all plants of three independent replications, evaluated at 14 and 28 days post-inoculation (dpi) and is shown as means with standard errors.

Significance of comparisons between the log of the number of molecules is indicated with vertical bars (ns, $p > 0.05$; *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$).

Fig. 5. Whitefly transmission of *Euphorbia yellow mosaic virus* (EuYMV) to *E. heterophylla* plants, alone or in the presence of Euphorbia yellow mosaic alphasatellite (EuYMA). **(a)** Yellow mosaic and vein chlorosis induced by EuYMV alone at 14 days post-inoculation (dpi). **(b, c)** Severe mosaic, leaf shriveling and stunting induced by EuYMV in the presence of EuYMA at 14 dpi. **(d)** Recovering symptoms at 28 dpi in plants infected with EuYMV alone, similar to plants infected with EuYMV and EuYMA. **(e)** Symptoms caused by infection with EuYMV DNA-A only. The arrow points to mild mosaic in plants infected with the DNA-A alone. **(f)** PCR-based amplification of fragments of the *CP* (1), *MP* (2) and *alpha-Rep* (3) genes for detection of EuYMV DNA-A and DNA-B and of EuYMA, respectively. M: Size marker (1Kb Plus DNA Ladder).

Figure 1

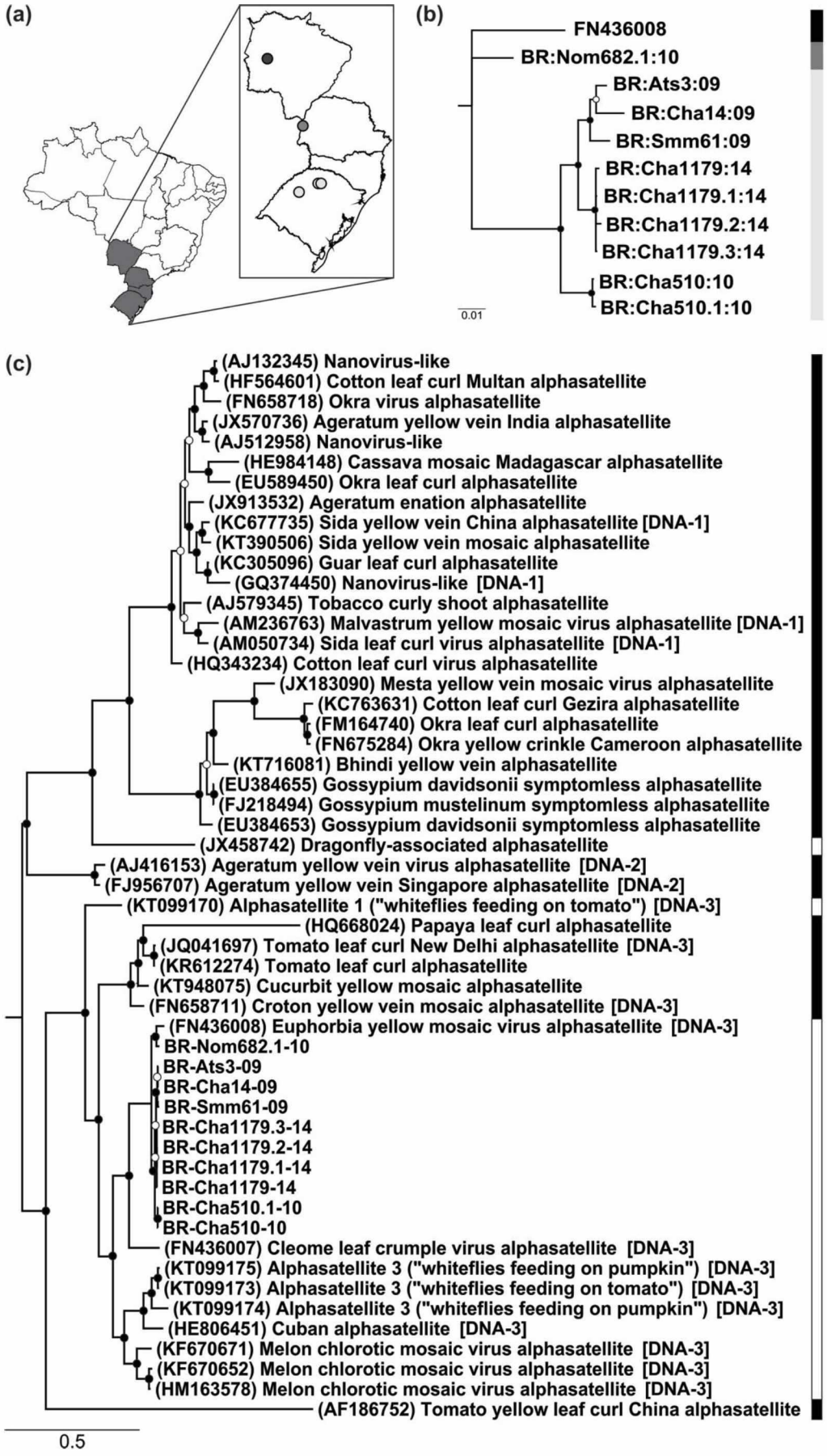

(a)
(b)
FN436008
BR:Nom682.1:10
BR:Ats3:09
BR:Cha14:09
BR:Smm61:09
BR:Cha1179:14
BR:Cha1179.1:14
BR:Cha1179.2:14
BR:Cha1179.3:14
BR:Cha510:10
BR:Cha510.1:10
0.01
(c)
(AJ132345) Nanovirus-like
(HF564601) Cotton leaf curl Multan alphasatellite
(FN658718) Okra virus alphasatellite
(JX570736) Ageratum yellow vein India alphasatellite
(AJ512958) Nanovirus-like
(HE984148) Cassava mosaic Madagascar alphasatellite
(EU589450) Okra leaf curl alphasatellite
(JX913532) Ageratum enation alphasatellite
(KC677735) Sida yellow vein China alphasatellite [DNA-1]
(KT390506) Sida yellow vein mosaic alphasatellite
(KC305096) Guar leaf curl alphasatellite
(GQ374450) Nanovirus-like [DNA-1]
(AJ579345) Tobacco curly shoot alphasatellite
(AM236763) Malvastrum yellow mosaic virus alphasatellite [DNA-1]
(AM050734) Sida leaf curl virus alphasatellite [DNA-1]
(HQ343234) Cotton leaf curl virus alphasatellite
(JX183090) Mesta yellow vein mosaic virus alphasatellite
(KC763631) Cotton leaf curl Gezira alphasatellite
(FM164740) Okra leaf curl alphasatellite
(FN675284) Okra yellow crinkle Cameroon alphasatellite
(KT716081) Bhindi yellow vein alphasatellite
(EU384655) Gossypium davidsonii symptomless alphasatellite
(FJ218494) Gossypium mustelinum symptomless alphasatellite
(EU384653) Gossypium davidsonii symptomless alphasatellite
(JX458742) Dragonfly-associated alphasatellite
(AJ416153) Ageratum yellow vein virus alphasatellite [DNA-2]
(FJ956707) Ageratum yellow vein Singapore alphasatellite [DNA-2]
(KT099170) Alphasatellite 1 ("whiteflies feeding on tomato") [DNA-3]
(HQ668024) Papaya leaf curl alphasatellite
(JQ041697) Tomato leaf curl New Delhi alphasatellite [DNA-3]
(KR612274) Tomato leaf curl alphasatellite
(KT948075) Cucurbit yellow mosaic alphasatellite
(FN658711) Croton yellow vein mosaic alphasatellite [DNA-3]
(FN436008) Euphorbia yellow mosaic virus alphasatellite [DNA-3]
BR-Nom682.1-10
BR-Ats3-09
BR-Cha14-09
BR-Smm61-09
BR-Cha1179.3-14
BR-Cha1179.2-14
BR-Cha1179.1-14
BR-Cha1179-14
BR-Cha510.1-10
BR-Cha510-10
(FN436007) Cleome leaf crumple virus alphasatellite [DNA-3]
(KT099175) Alphasatellite 3 ("whiteflies feeding on pumpkin") [DNA-3]
(KT099173) Alphasatellite 3 ("whiteflies feeding on tomato") [DNA-3]
(KT099174) Alphasatellite 3 ("whiteflies feeding on pumpkin") [DNA-3]
(HE806451) Cuban alphasatellite [DNA-3]
(KF670671) Melon chlorotic mosaic virus alphasatellite [DNA-3]
(KF670652) Melon chlorotic mosaic virus alphasatellite [DNA-3]
(HM163578) Melon chlorotic mosaic virus alphasatellite [DNA-3]
(AF186752) Tomato yellow leaf curl China alphasatellite
0.5
Old World
New World

Figure 2

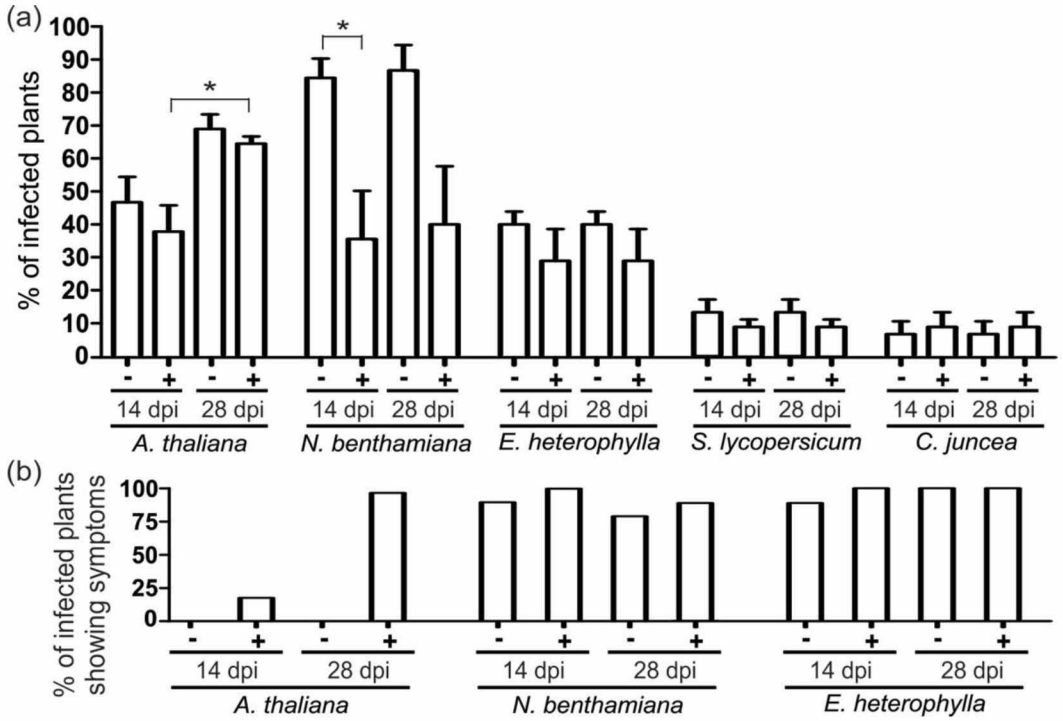

(a)
% of infected plants
100
90
80
70
60
50
40
30
20
10
0
*
*
- + - + - + - + - + - + - + - + - + - +
14 dpi 28 dpi 14 dpi 28 dpi 14 dpi 28 dpi 14 dpi 28 dpi 14 dpi 28 dpi
A. thaliana
N. benthamiana
E. heterophylla
S. lycopersicum
C. juncea
(b)
% of infected plants showing symptoms
100
75
50
25
0
- + - + - + - + - + - +
14 dpi 28 dpi 14 dpi 28 dpi 14 dpi 28 dpi
A. thaliana
N. benthamiana
E. heterophylla

Figure 3

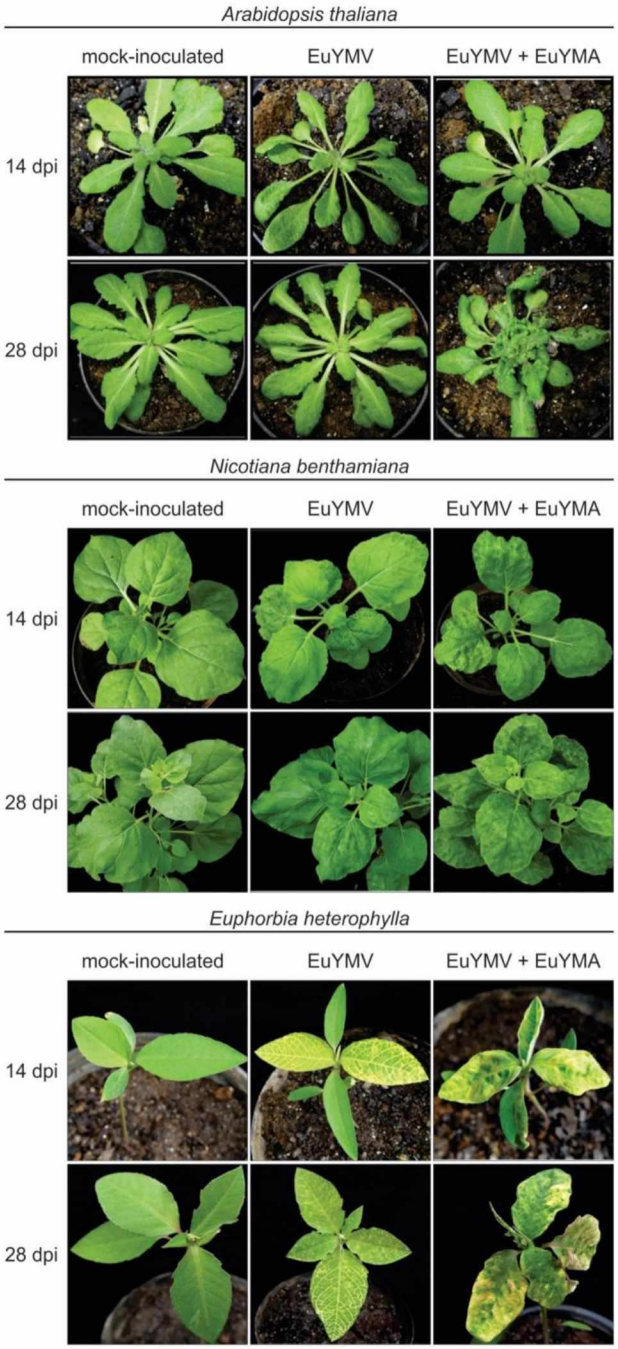

Arabidopsis thaliana
mock-inoculated
EuYMV
EuYMV + EuYMA
14 dpi
28 dpi
Nicotiana benthamiana
mock-inoculated
EuYMV
EuYMV + EuYMA
14 dpi
28 dpi
Euphorbia heterophylla
mock-inoculated
EuYMV
EuYMV + EuYMA
14 dpi
28 dpi

Figure 4

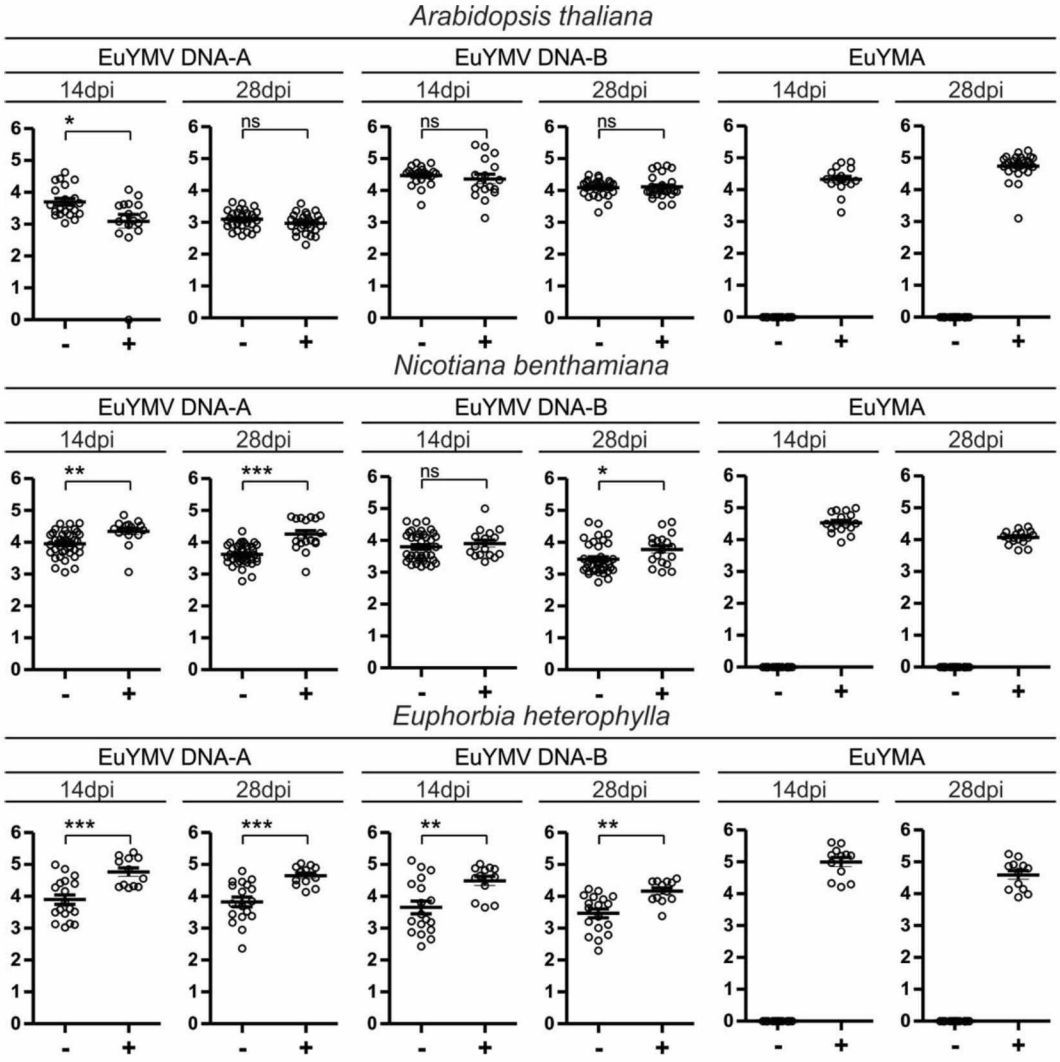
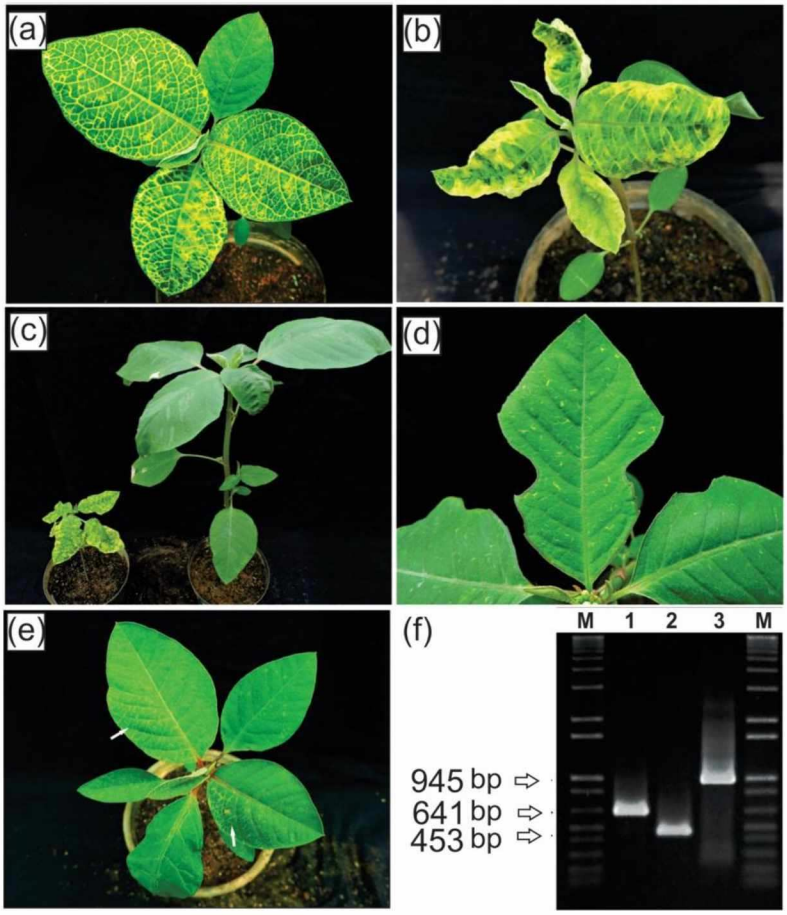

Arabidopsis thaliana
EuYMV DNA-A
EuYMV DNA-B
EuYMA
14dpi
28dpi
*
ns
ns
ns
Nicotiana benthamiana
**
***
ns
*
Euphorbia heterophylla
***
***
**
**
- +

Figure 5



Supplementary material

Supplementary Table S1. Sequences of Euphorbia yellow mosaic alphasatellite (EuYMA) described in this study. The isolate used for the construction of infectious clones is in bold.

Code	Location	Geographical coordinates		Year	Isolate	GenBank accession number**
3	Almirante Tamandaré - RS	28°06'35" S	52°54'29" W	2009	BR-Ats3-09	KY559637
14	Chapada - RS	28°02'39" S	53°04'53" W	2009	BR-Cha14-09	KY559638
61	São Miguel das Missões - RS	28°29'35" S	54°33'37" W	2009	BR-Smm61-09	KY559639
510	**Chapada - RS**	**28°02'39" S**	**53°04'53" W**	**2010**	**BR-Cha510-10**	**KY559640**
					BR-Cha510.1-10	KY559641
682	Mercedes - PR	24°30'54" S	54°07'00" W	2010	BR-Nom682.1-10	KY559642
1179	Chapada - RS	28°02'39" S	53°04'53" W	2014	BR-Cha1179-14	KY559643
					BR-Cha1179.1-14	KY559644
					BR-Cha1179.2-14	KY559645
					BR-Cha1179.3-14	KY559646

Supplementary Table S2. Results of the host range assay. No. of PCR-positive plants out of 15 plants inoculated at 14 and 28 days post-inoculation (dpi) with *Euphorbia yellow mosaic virus* (EuYMV) and its associated alphasatellite Euphorbia yellow mosaic alphasatellite (EuYMA) in tree independent experiments.

	EuYMV								EuYMV + EuYMA							
	14 dpi				28 dpi				14 dpi				28 dpi			
	Exp 1	Exp 2	Exp 3	Total	Exp 1	Exp 2	Exp 3	Total	Exp 1	Exp 2	Exp 3	Total	Exp 1	Exp 2	Exp 3	Total
Arabidopsis thaliana	5	9	7	**21**	9	11	11	**31**	5	4	8	**17**	9	10	10	**29**
Nicotiana benthamiana	11	14	13	**38**	11	15	13	**39**	1	8	7	**16**	1	10	7	**18**
Euphorbia heterophylla	5	7	6	**18**	5	7	6	**18**	4	2	7	**13**	4	2	7	**13**
Solanum lycopersicum	3	2	1	**6**	3	2	1	**6**	1	2	1	**4**	1	2	1	**4**
Crotalaria juncea	0	1	2	**3**	0	1	2	**3**	2	2	0	**4**	2	2	0	**4**

Supplementary figure legends

Supplementary Fig. S1. Pairwise sequence comparisons between isolates of Euphorbia yellow mosaic alphasatellite (EuYMA). Identities were calculated with SDT v.1.2.

Supplementary Fig. S2. Midpointed-rooted Bayesian phylogenetic trees based on *alpha-Rep* gene deduced amino acid sequences. Nodes to the right of branches with posterior probabilities equal to or higher than 0.8 are indicated by filled circles and those with values lower than 0.8 and higher than 0.5 by empty circles. The vertical bar indicates Old World and New World begomoviruses.

Supplementary Fig. S3. Symptoms of *Euphorbia yellow mosaic virus* (EuYMV) in the presence and absence of Euphorbia yellow mosaic alphasatellite (EuYMA) at 14 (left) and 28 (right) days post-inoculation. Images are coded as follows: Uppercase A, *A. thaliana*; N, *N. benthamiana*; E, *E. heterophylla*; First number 1 to 15, EuYMV in the absence of EuYMA; 15 to 30, EuYMV in the presence of EuYMA; Second number 1, first; 2, second; 3, third replications. *plants which were PCR negative at 14 dai but PCR positive at 28 dai. The *E. heterophylla* panel displays plants at 14 dai only for the first replication, as the symptoms were essentially identical at 14 and 28 dai.

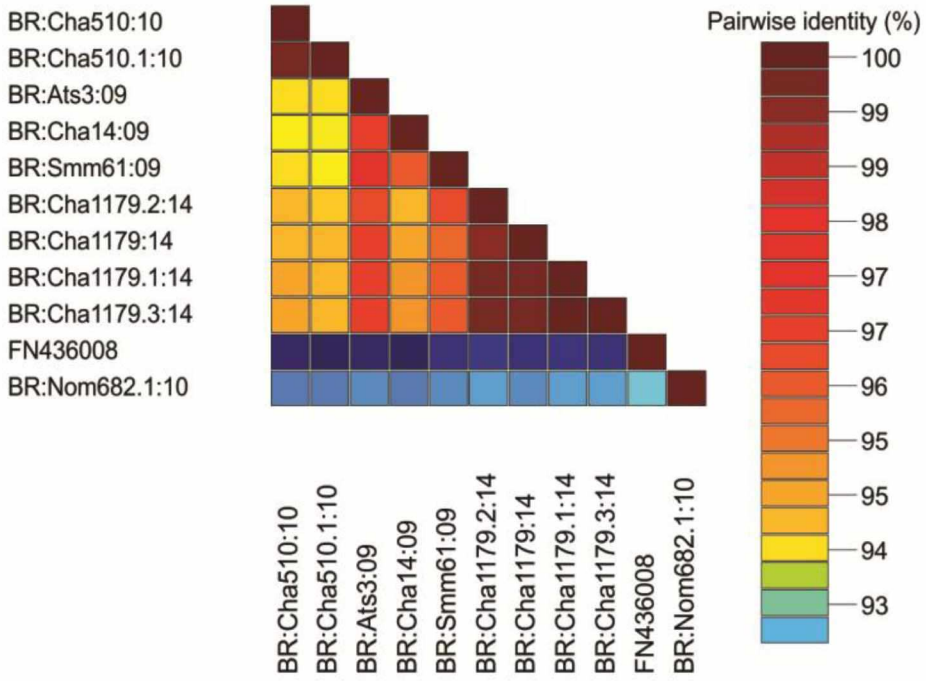
Supplementary Fig. S4. Symptoms in young (left) and old (right) leaves of *N. bethamiana* infected with *Euphorbia yellow mosaic virus* (EuYMV) in the absence (top) and presence of Euphorbia yellow mosaic alphasatellite (EuYMA) (bottom) at 28 days post-inoculation (dpi).

Supplementary Fig. S5. Absolute quantification by quantitative PCR of *Euphorbia yellow mosaic virus* (EuYMV) DNA-A and DNA-B and of Euphorbia yellow mosaic alphasatellite (EuYMA) in different hosts. Viral accumulation is presented as the *log* of the number of molecules. Each circle represents one infected plant. (-) represents infection by EuYMV alone; (+) represents infection by EuYMV together with EuYMA. Data is presented separately for each of three independent replications, evaluated at 14 and 28 days post-inoculation (dpi) and is shown as means with standard errors. Significance between comparisons of the

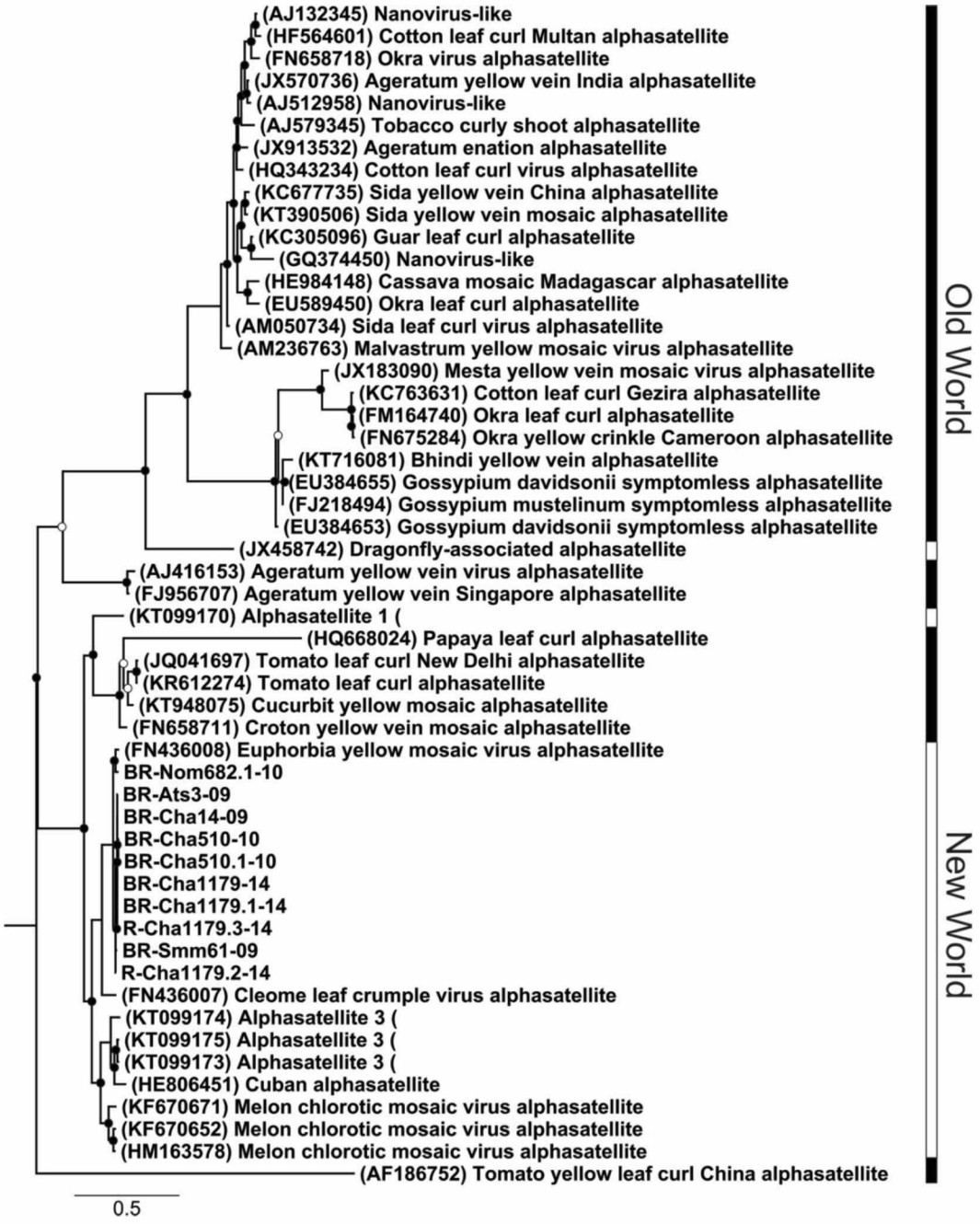
log of the number of molecules is indicated with vertical bars (ns, $p>0.05$; *, $p\leq0.05$; **, $p\leq0.01$; ***, $p\leq0.001$).

Supplementary Fig. S6. Comparisons between the number of *Euphorbia yellow mosaic virus* (EuYMV) genomic components and Euphorbia yellow mosaic alphasatellite (EuYMA) molecules in plants infected with EuYMV alone or in the presence of EuYMA. **(a)** Data refers to all plants of three independent replications, evaluated at 14 and 28 days post-inoculation (dpi), and is shown as means with standard errors. Significance is indicated by vertical bars (ns, $p > 0.05$; *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$). **(b)** Data is presented separately for each of three independent replications, evaluated at 14 and 28 days post-inoculation (dpi), and is shown as means with standard errors. Significance is indicated with vertical bars (ns, $p>0.05$; *, $p\leq0.05$; **, $p\leq0.01$; ***, $p\leq0.001$).

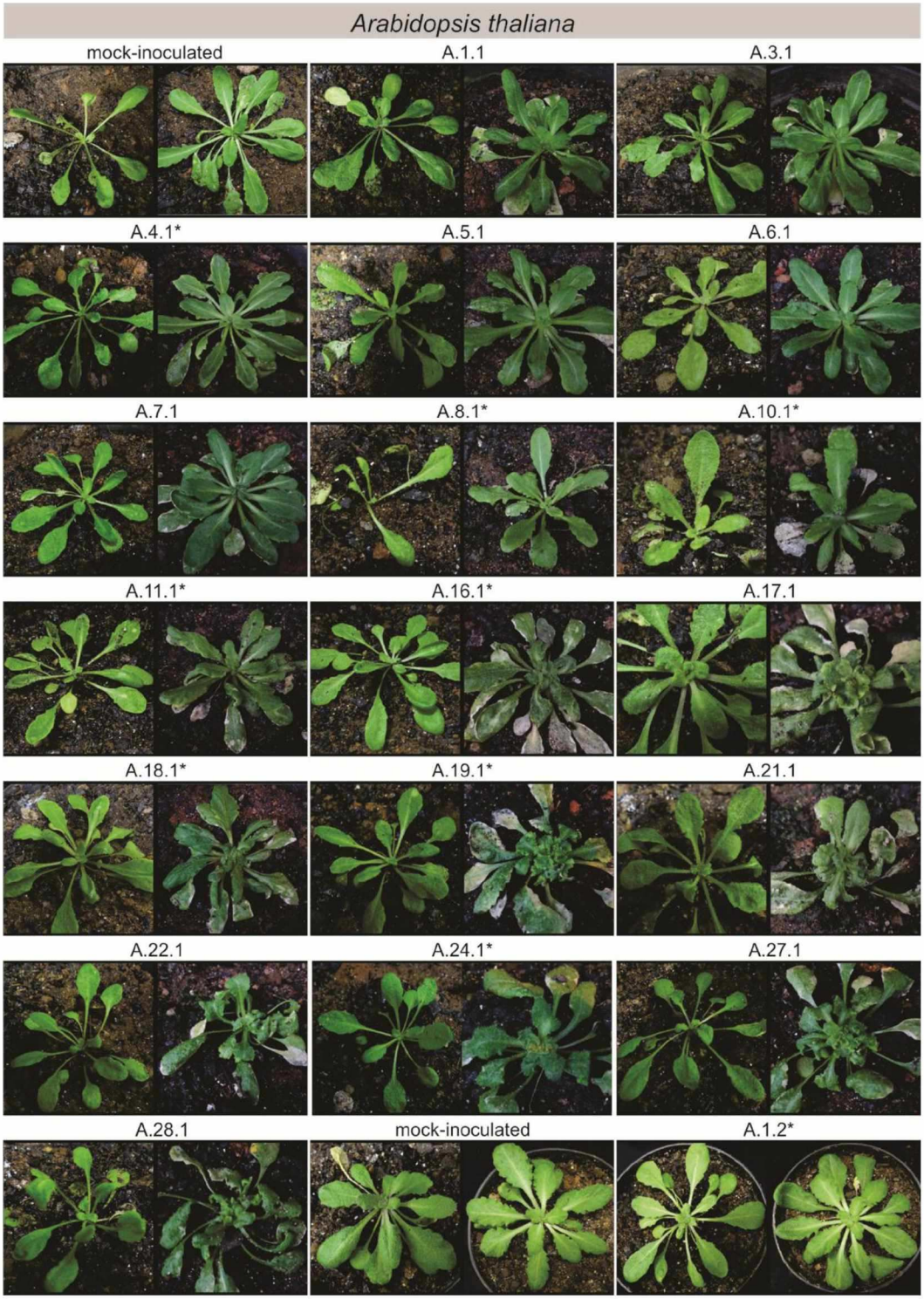
Supplementary Fig. S1



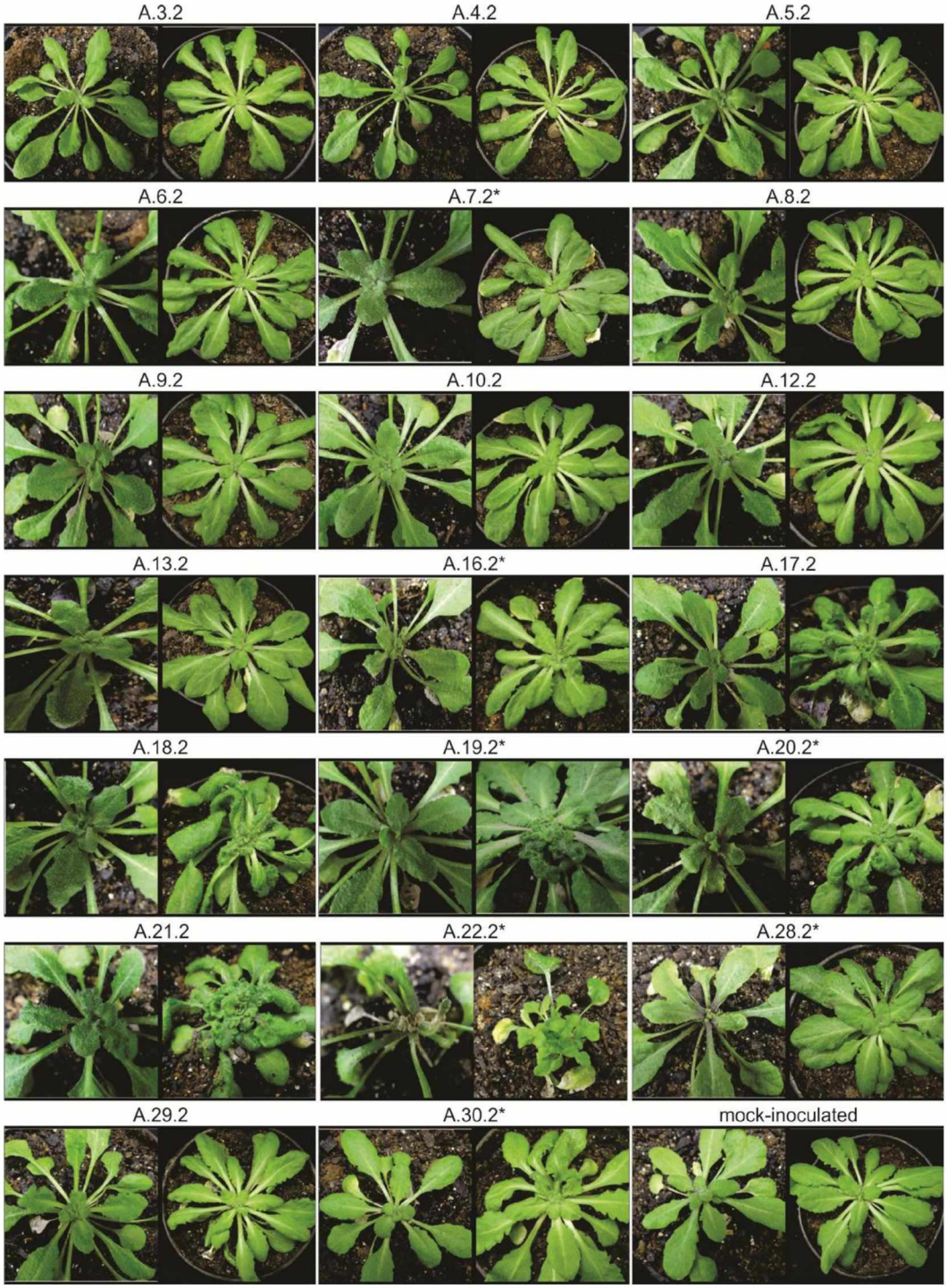
Supplementary Fig. S2



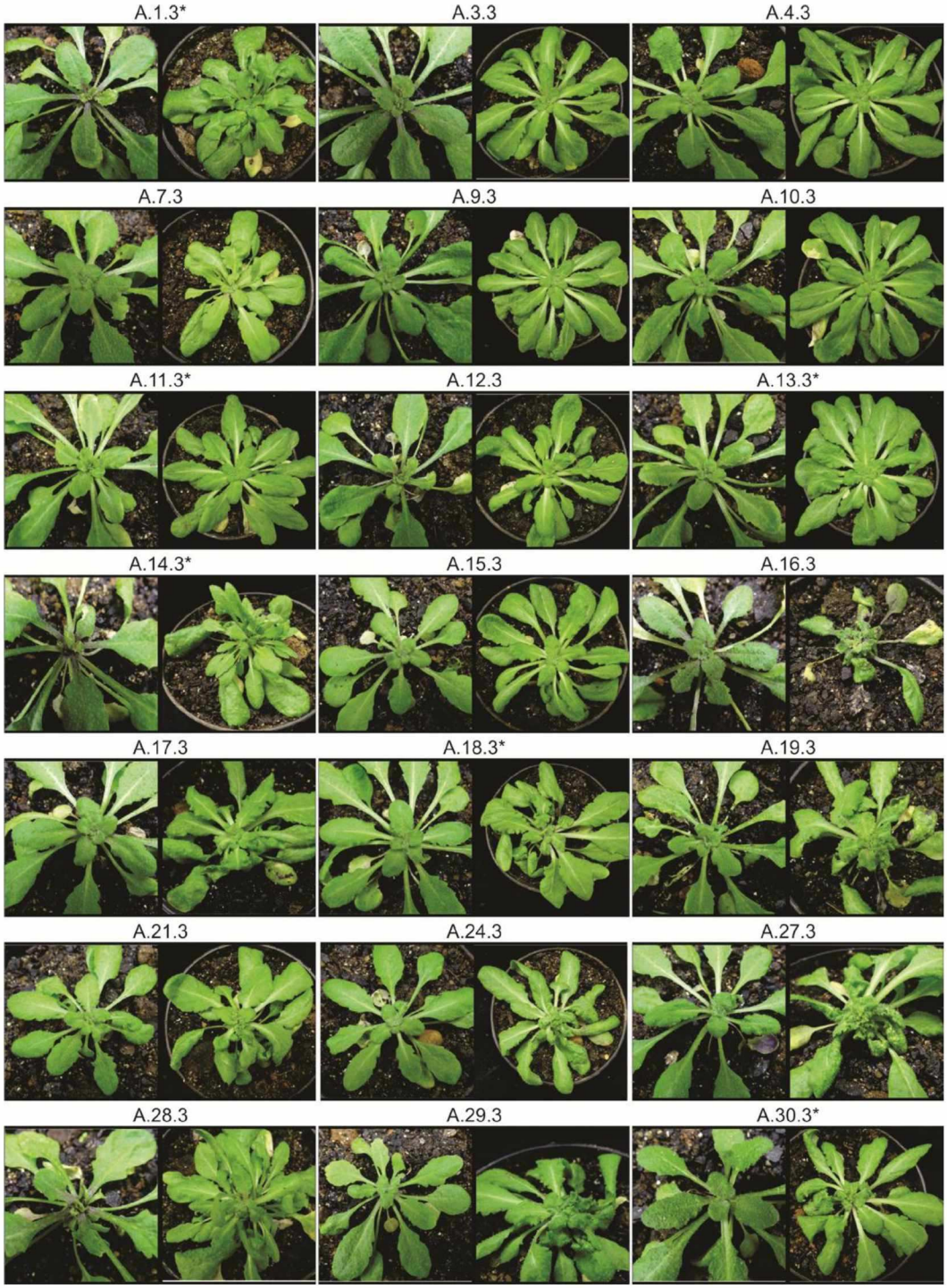
Supplementary Fig. S3



Supplementary Fig. S3 (cont.)

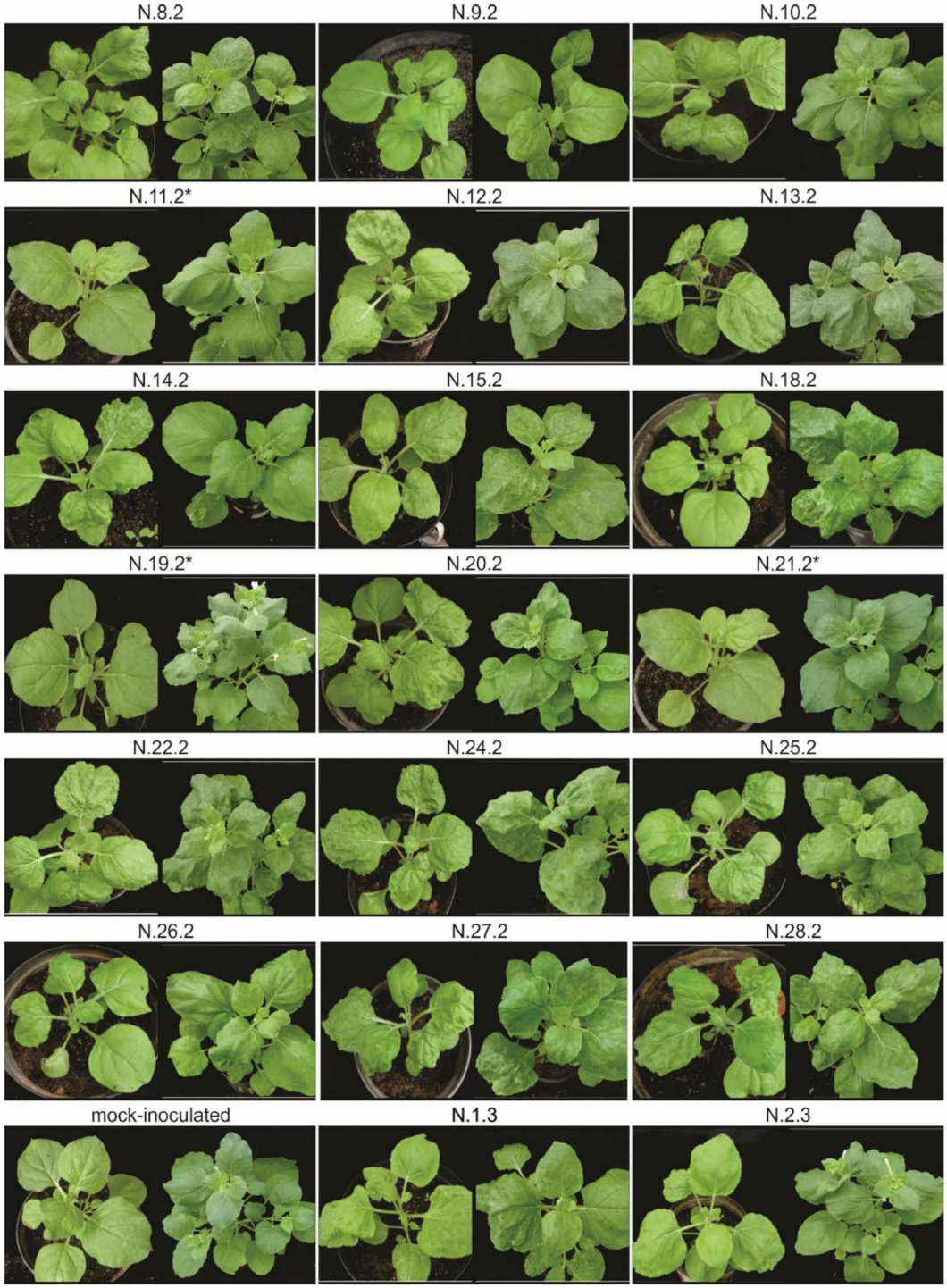


Supplementary Fig. S3 (cont.)

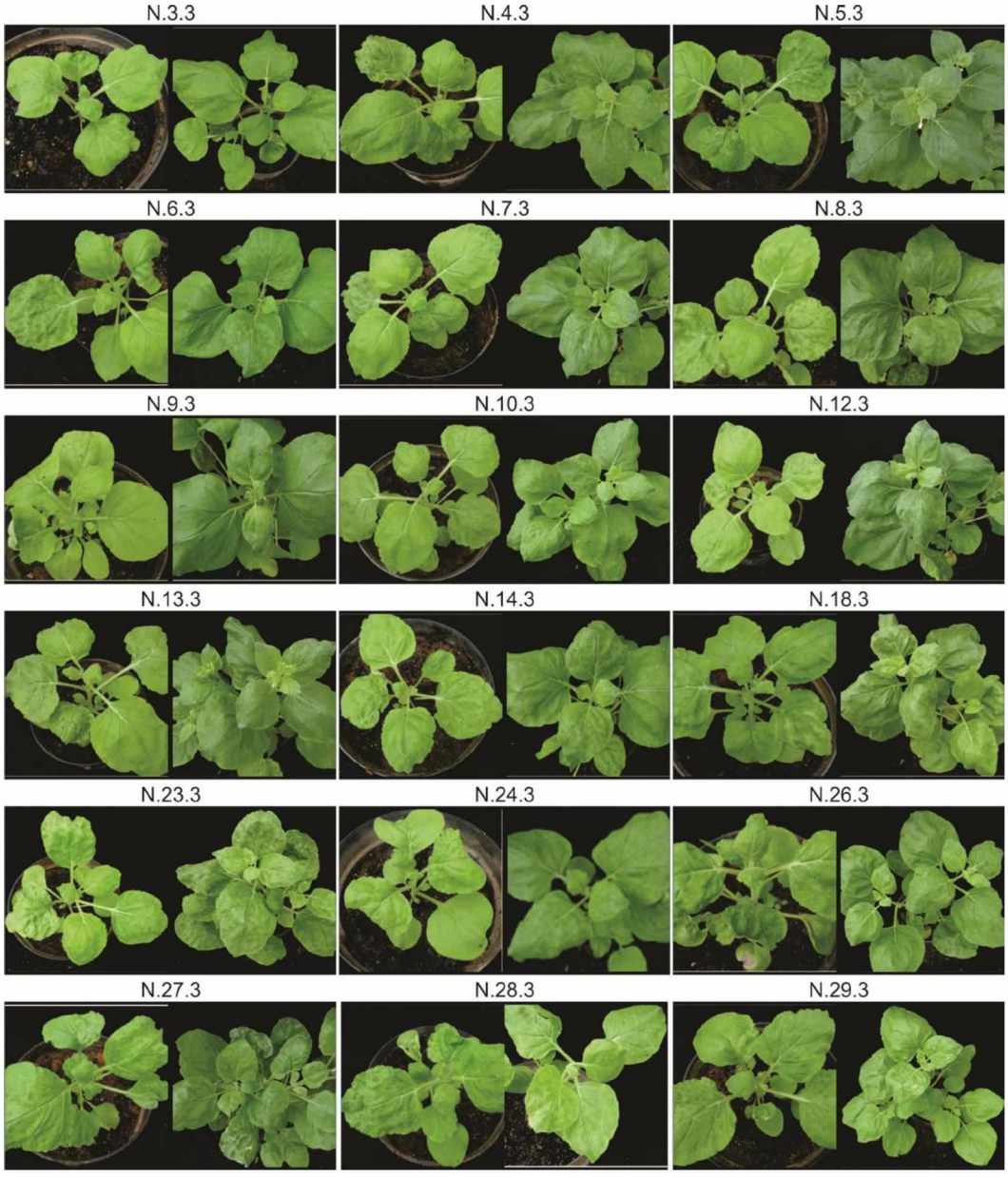


Supplementary Fig. S3 (cont.)

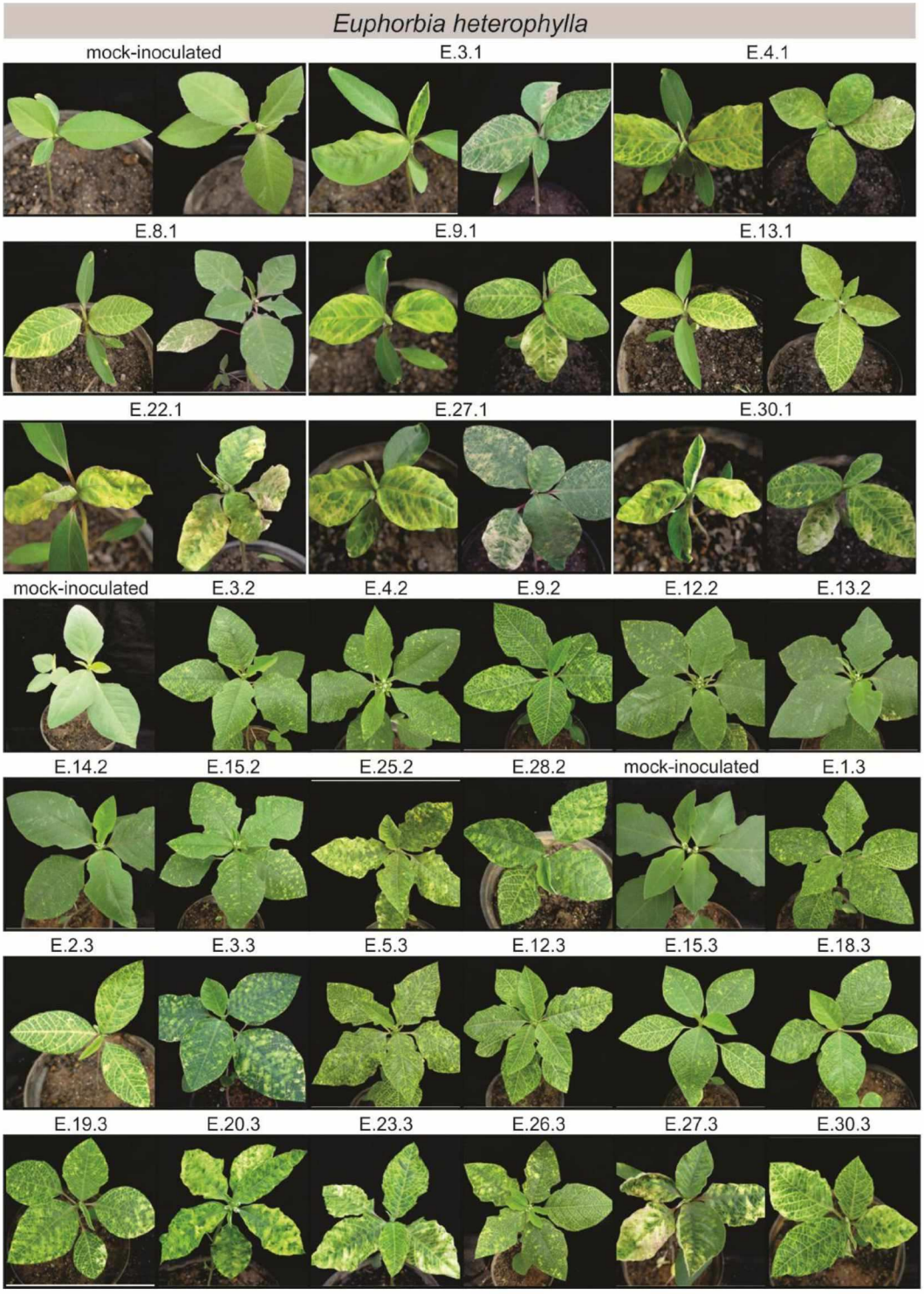
Supplementary Fig. S3 (cont.)



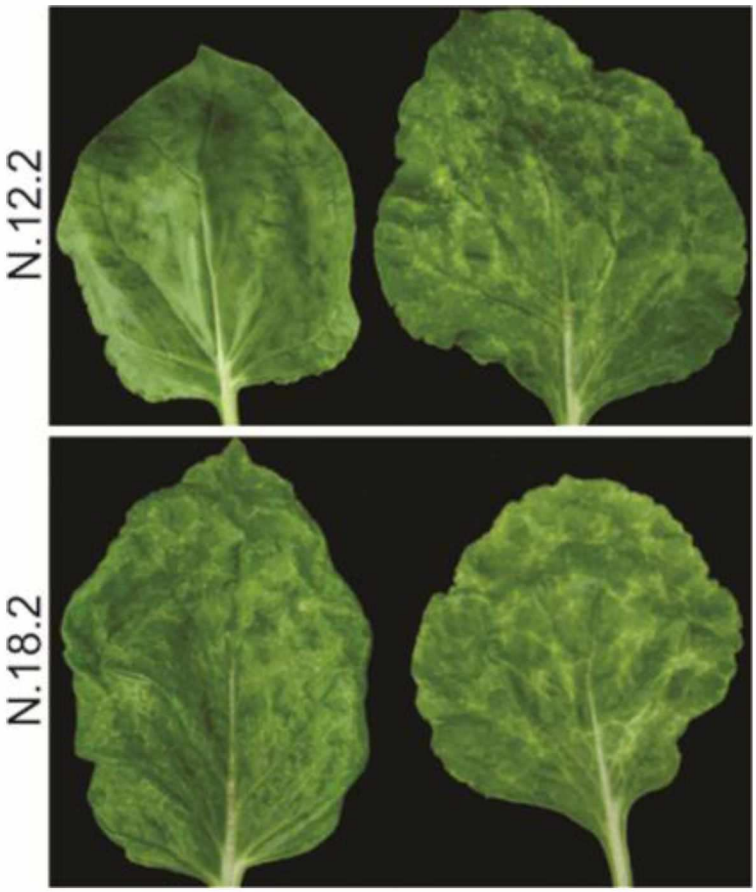
Supplementary Fig. S3 (cont.)



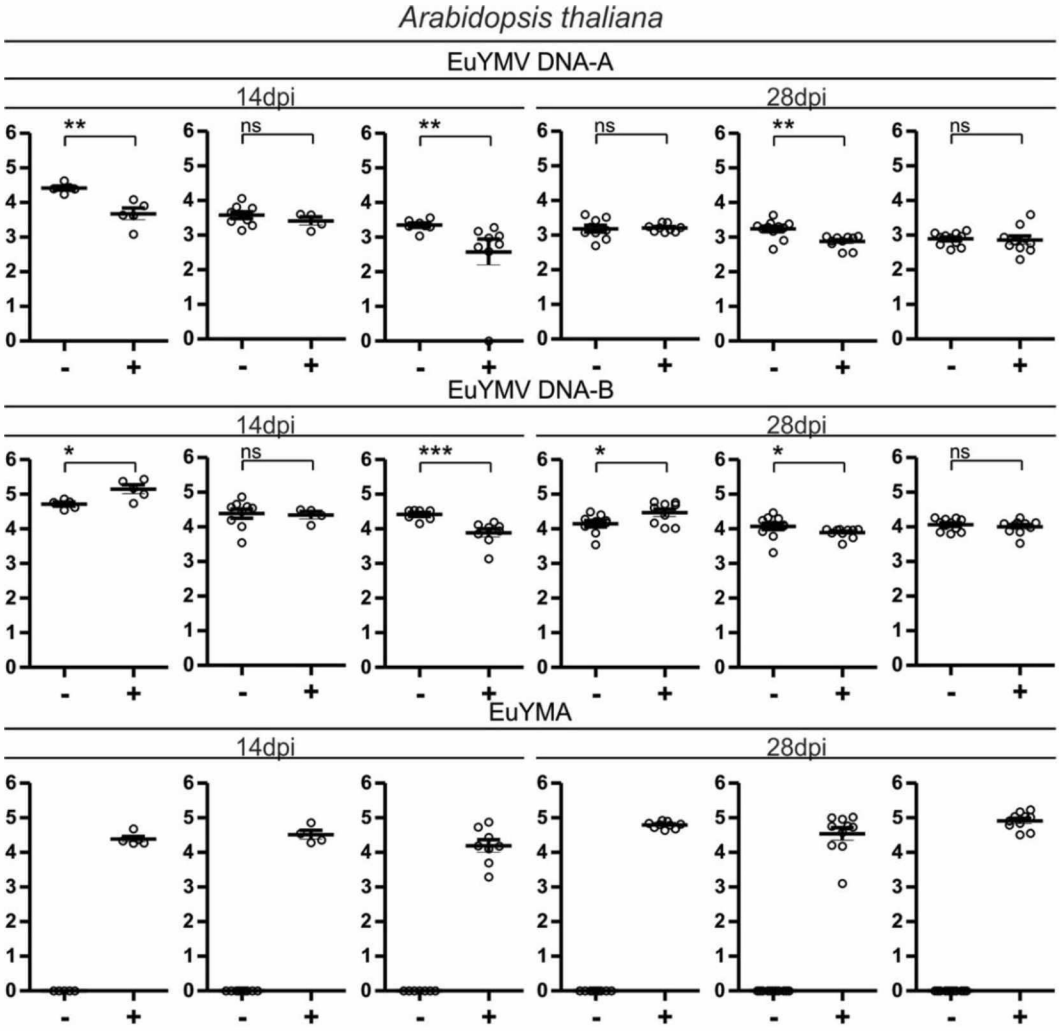
Supplementary Fig. S3 (cont.)



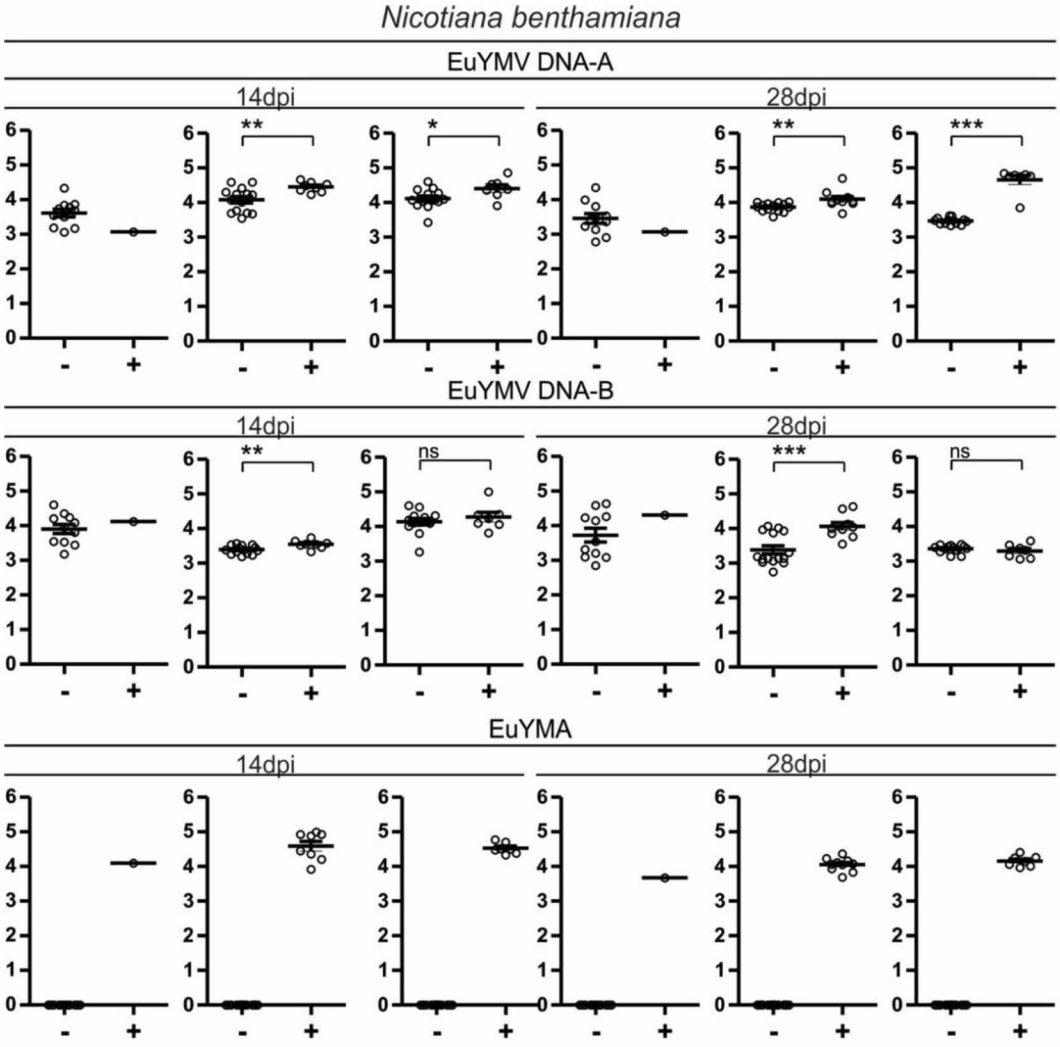
Supplementary Fig. S4



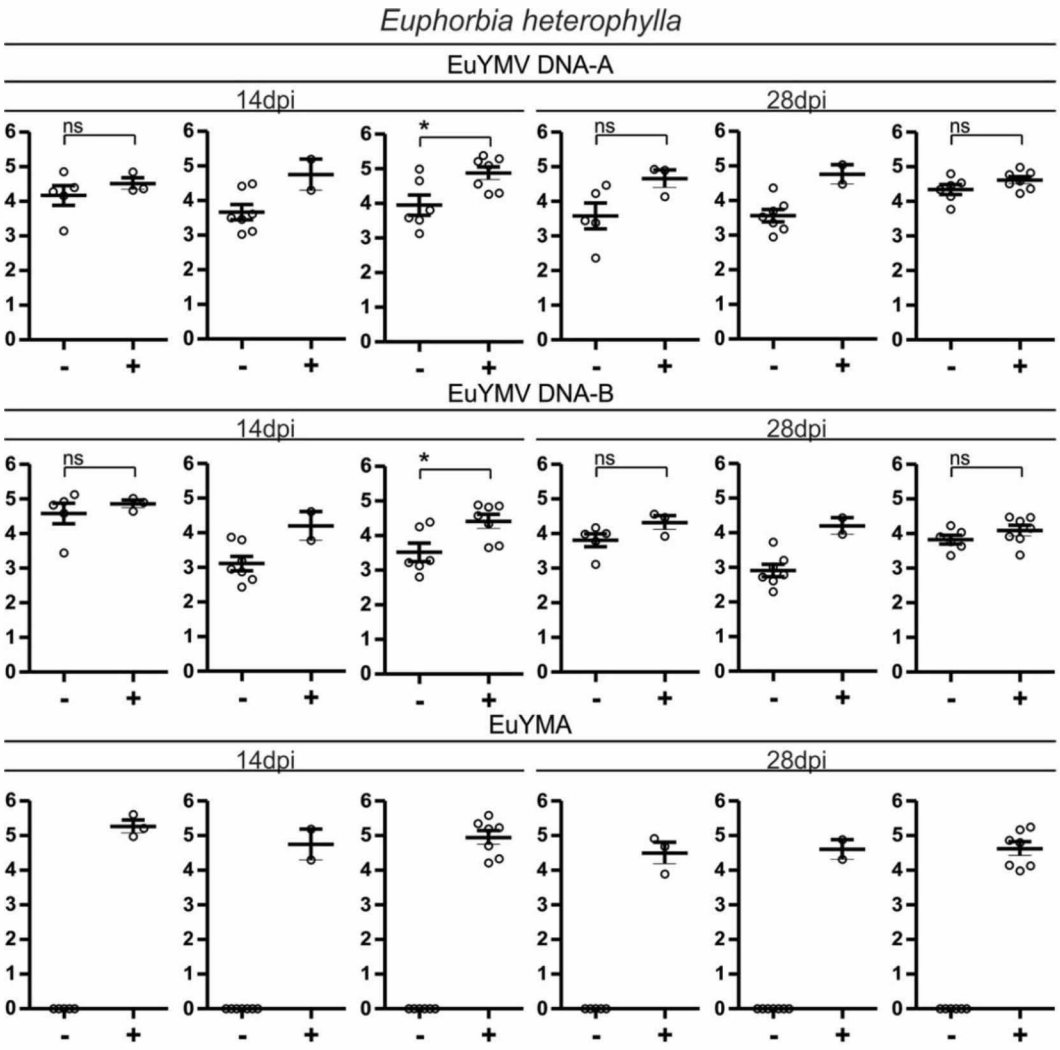
Supplementary Fig. S5



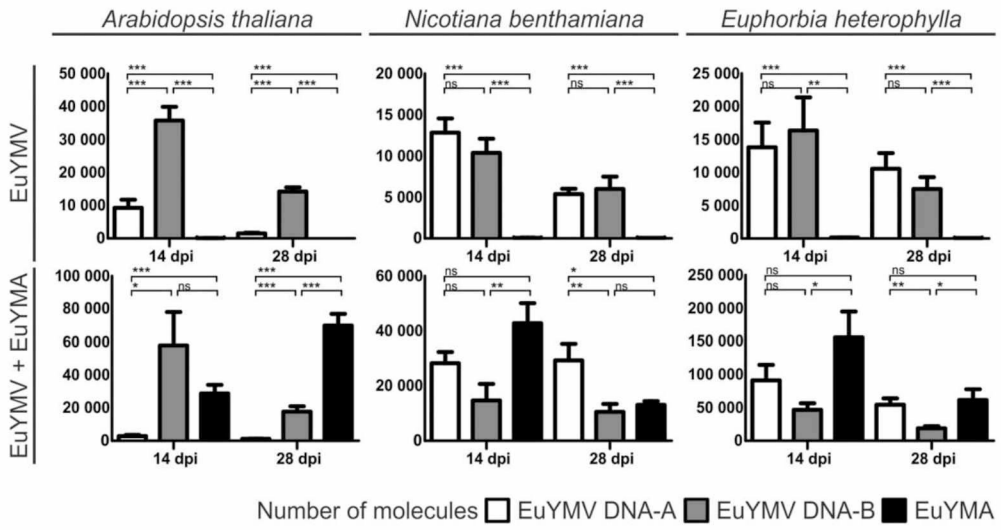
Supplementary Fig. S5 (cont.)



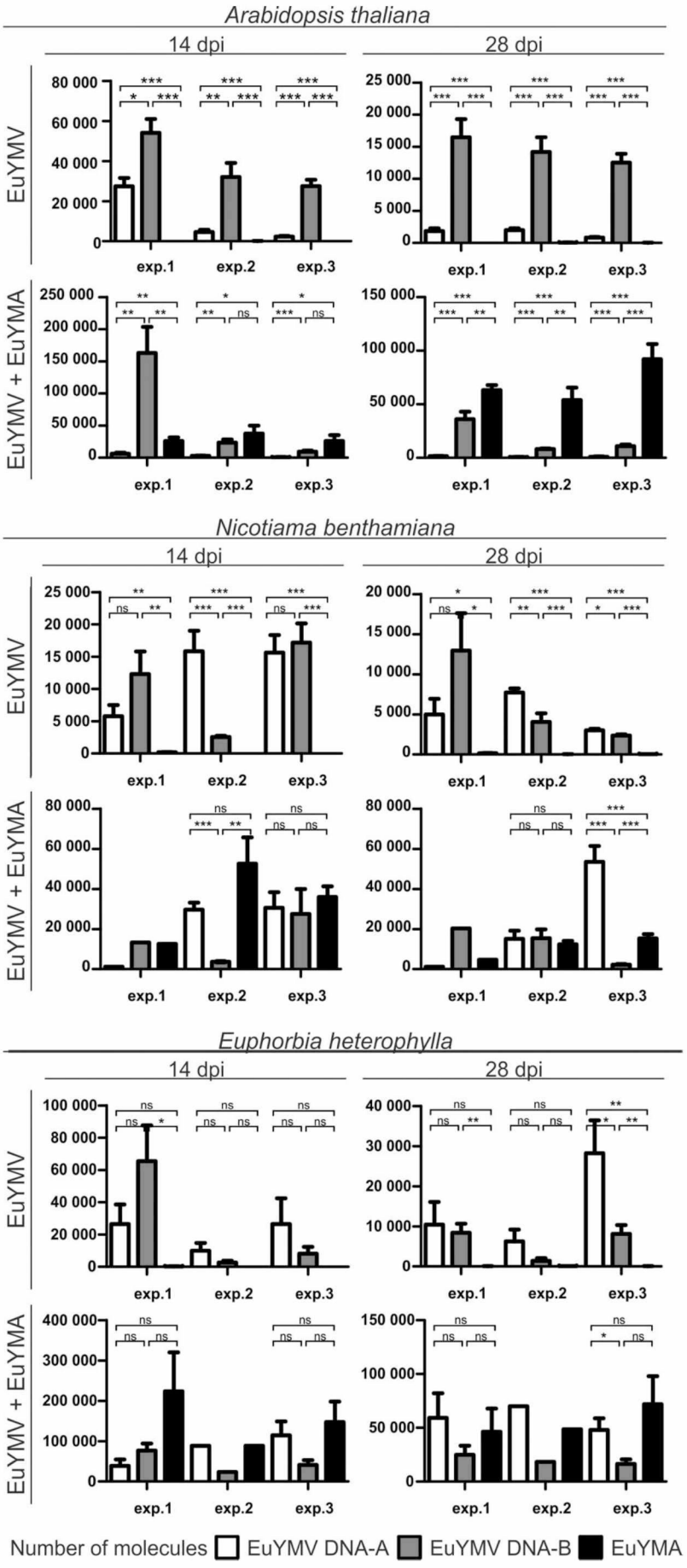
Supplementary Fig. S5 (cont.)



Supplementary Fig. S6a



Supplementary Fig. S6b



GENERAL CONCLUSIONS

- EuYMV has a lower degree of genetic variability compared with other begomovirus populations infecting cultivted and non-cultivated plants, and only a few intraspecific recombination events (restricted to the DNA-B component) were detected;

- EuYMV displays a different pattern from that commonly observed in begomovirus populations (high variability and frequent recombination events), but nevertheless, similarly to other begomoviruses, segregates according to sampling location;

- The association of the alphasatellite EuYMA with the bipartite begomovirus EuYMV is rare, although their geographical range is large;

- EuYMA can increase the symptom severity of EuYMV infection and promotes changes in DNA accumulation;

- The presence of EuYMA negatively interferes with transmission by the whitefly vector, which can negatively affect the spread of the virus in the field.