UNIVERSIDADE FEDERAL DE VIÇOSA

BEGOMOVIRUSES IN NON-CULTIVATED PLANTS: CHARACTERIZATION AND GENETIC VARIABILITY

Marcelo Henrique Oliveira Gonçalves
Magister Scientiae

VIÇOSA - MINAS GERAIS
2024

MARCELO HENRIQUE OLIVEIRA GONÇALVES

BEGOMOVIRUSES IN NON-CULTIVATED PLANTS: CHARACTERIZATION AND GENETIC VARIABILITY

Dissertation submitted to the Plant Pathology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

Adviser: Francisco M Zerbini Junior

VIÇOSA - MINAS GERAIS
2024

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
G635b
2024

Gonçalves, Marcelo Henrique Oliveira, 1999-
Begomoviruses in non-cultivated plants: characterization and genetic variability / Marcelo Henrique Oliveira Gonçalves. – Viçosa, MG, 2024.
1 dissertação eletrônica (65 f.): il. (algumas color.).

Texto em inglês.
Orientador: Francisco Murilo Zerbini Júnior.
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Fitopatologia, 2024.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2024.761
Modo de acesso: World Wide Web.

1. Begomovirus. 2. Plantas - Doenças e pragas. 3. Vírus - Genética. 4. Recombinação (Genética). I. Zerbini Júnior, Francisco Murilo, 1966-. II. Universidade Federal de Viçosa. Departamento de Fitopatologia. Programa de Pós-Graduação em Fitopatologia. III. Título.

CDD 22. ed. 632.8

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317

MARCELO HENRIQUE OLIVEIRA GONÇALVES

BEGOMOVIRUSES IN NON-CULTIVATED PLANTS: CHARACTERIZATION AND GENETIC VARIABILITY

Dissertation submitted to the Plant Pathology Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Magister Scientiae*.

APPROVED: August 22, 2024.

Assent:

Marcelo Henrique Oliveira Gonçalves
Author

Francisco Murilo Zerbini Junior
Adviser

Essa dissertação foi assinada digitalmente pelo autor em 19/11/2024 às 14:16:27 e pelo orientador em 19/11/2024 às 15:56:05. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse https://siadoc.ufv.br/validar-documento. No campo 'Código de registro', informe o código **9QJB.32ZF.YZB3** e clique no botão 'Validar documento'.

À minha mãe e à minha avó,
com gratidão pelo amor e suporte que sempre me ofereceram,
dedico.

ACKNOWLEDGMENTS

Agradeço a toda minha família, especialmente à minha mãe e à minha avó, a quem devo ser quem sou hoje.

Ao meu orientador, Prof. Francisco Murilo Zerbini, pela parceria e pelos ensinamentos valiosos.

Agradeço à Prof. Rosana Blawid e à Prof. Sarah Jacqueline Cavalcanti da Silva por disponibilizarem o material de estudo e pela colaboração essencial para o desenvolvimento deste trabalho.

Meu sincero agradecimento a todos os membros do Laboratório de Ecologia e Evolução de Vírus pelo apoio e colaboração indispensáveis ao longo desta jornada: Aline, Anelise, Franciely, João Paulo, Júlia, Lorenna, Maraísa, Narla, Pâmella, Patrícia, Ruither, Temmy e Valdemir. Mas em especial, agradeço a Ayane Quadros por toda a ajuda e conselhos.

Agradeço à Universidade Federal de Viçosa e ao Departamento de Fitopatologia, onde fiz grandes amigos: Débora, Dyênici, Eduardo, Henara, Isabella e Monalisa.

Ao Roger, pela paciência, amor e constante apoio em todos os momentos desta jornada. Sua presença foi essencial para a realização deste trabalho.

Aos amigos que fiz, cuja amizade me acolheu em Viçosa: Bárbara, Eliel, Fábulo, Jackson, João Paulo, Juniel, Luíza e Rafael.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

“Ao meu passado, eu devo o meu saber e a minha ignorância, as minhas necessidades, as minhas relações, a minha cultura e o meu corpo. Que espaço meu passado deixa para minha liberdade hoje?”

— Parafraseando Simone de Beauvoir,
na canção "Amarelo, azul e branco" por ANAVITÓRIA e Rita Lee.

ABSTRACT

GONÇALVES, Marcelo Henrique Oliveira, M.Sc., Universidade Federal de Viçosa, August, 2024. **BEGOMOVIRUSES IN NON-CULTIVATED PLANTS: CHARACTERIZATION AND GENETIC VARIABILITY**. Adviser: Francisco Murilo Zerbini Junior.

Viruses in the genus Begomovirus (family *Geminiviridae*) are characterized by geminate particles and a single-stranded, circular DNA genome (ssDNA), which can be mono- or bipartite. These viruses are transmitted by whiteflies of the *Bemisia tabaci* species complex and often cause significant economic losses in agricultural crops. Many begomoviruses are indigenous to Brazil and infect non-cultivated plants. Evidence indicates that some of these viruses have spilled-over and adapted to cultivated plants. Thus, the study of begomoviruses in the natural environment is important to understand the factors that favour spillover and the emergence of viruses in crops. The first aim of this study was to characterize the isolate GS-20 of bean bushy stunt virus (*Begomovirus phaseoliretorridi*, BBSV) found in *Neustanthus phaseoloides* (tropical kudzu) in Brazil. The isolate was classified as a new strain of BBSV. Unusual for begomoviruses, the CP gene was the most diverse compared to other genes. A recombination event was detected between BBSV and soybean chlorotic spot virus (SoCSV). The isolate was unable to infect soybean but showed a low infection rate in two common bean cultivars. The second aim of this study was to characterize the genetic structure and variability of Blainvillea yellow spot virus (*B. blainvilleae*, BlYSV), a virus with a restricted host range but high genetic variability. Twenty-three DNA-A clones of BlYSV were obtained from samples collected in the states of Minas Gerais, Rio Grande do Norte, and Alagoas. Four variants (A-D), with genetic identity over 96%, were identified, with variant A classified as a distinct strain. Five recombination events were detected among the isolates. The relative distribution of variants in Viçosa and Coimbra in 2022-2023 was markedly different from 2010-2014. A correlation was observed between the year of collection and the root-to-tip distance in the phylogenetic tree. Variant A showed lower nucleotide diversity compared to the other variants, and four isolates from this variant had a substitution (A2667G) within the nonanucleotide motif at the origin of replication (from 5'-TAATATTAC-3' to 5'-TAATGTTAC-3'). Despite strong negative selection, two variants showed positive selection in the AC4 gene.

Keywords: geminivirus; *begomovirus phaseoliretorridi*; *begomovirus blainvilleae*; *virus evolution*; *population structure*

RESUMO

GONÇALVES, Marcelo Henrique Oliveira, M.Sc., Universidade Federal de Viçosa, agosto de 2024. **BEGOMOVÍRUS EM PLANTAS NÃO CULTIVADAS: CARACTERIZAÇÃO E VARIABILIDADE GENÉTICA**. Orientador: Francisco Murilo Zerbini Junior.

Os vírus classificados no gênero Begomovirus (família *Geminiviridae*) são caracterizados por possuírem partículas geminadas e um genoma de DNA de fita simples circular (ssDNA), que pode ser mono- ou bipartido. Esses vírus são transmitidos por moscas-brancas do complexo de espécies *Bemisia tabaci*, e frequentemente causam perdas econômicas significativas em culturas agrícolas. Evidências sugerem que diversos begomovírus são nativos do Brasil, infectando plantas não cultivadas, e que alguns desses vírus passaram por eventos de spillover e se adaptaram a plantas cultivadas. O primeiro objetivo deste trabalho foi caracterizar o isolado GS-20 do bean bushy stunt virus (*Begomovirus phaseoliretorridi*, BBSV), encontrado na planta não-cultivada *Neustanthus phaseoloides*. O isolado GS-20 foi classificado como uma nova estirpe do BBSV. O gene CP apresentou a maior diversidade em relação aos outros genes. Um evento de recombinação foi detectado entre o BBSV e o soybean chlorotic spot virus (SoCSV). O isolado não foi capaz de infectar soja, mas apresentou uma baixa taxa de infecção em duas cultivares de feijão comum. O segundo objetivo deste estudo foi caracterizar a estrutura e variabilidade genética do Blainvillea yellow spot virus (*B. blainvilleae*, BlYSV), um vírus com uma gama de hospedeiros restrita, mas com alta variabilidade genética. Foram obtidos 23 clones do DNA-A do BlYSV a partir de amostras coletadas nos estados de Minas Gerais, Rio Grande do Norte e Alagoas. Quatro variantes (A-D), com identidade genética superior a 96%, foram identificadas. A variante A foi classificada como uma estirpe distinta. Foram detectados cinco eventos de recombinação. A distribuição das variantes em Viçosa e Coimbra em 2022-2023 foi significativamente diferente de 2010-2014. Observou-se também uma correlação entre o ano de coleta dos isolados e a distância da ponta até a raiz da árvore filogenética. A variante A apresentou menor diversidade nucleotídica em comparação às outras variantes, e quatro isolados dessa variante apresentaram uma substituição (A2667G) dentro do nonanucleotídeo localizado na origem de replicação (de 5'-TAATATTAC-3' para 5'-

TAATGTTAC-3'). Embora sob forte seleção negativa, duas variantes possuem o gene AC4 sob seleção positiva.

Palavras-chave: geminivirus; begomovirus phaseoliretorridi; begomovirus blainvilleae; evolução viral; estrutura de população

SUMMARY

General Introduction 11

References 14

CHAPTER 1. Biological and molecular characterization of a Brazilian isolate of bean bushy stunt virus (*Begomovirus phaseoliretorridi*) 17

Abstract 18

Introduction 19

Methods 20

Results and Discussion 21

References 25

Supplementary Material 28

CHAPTER 2. Genetic variability of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) reveals recombinant variants and a distinct nonanucleotide motif. 33

Abstract 34

Introduction 35

Methods 36

Results 38

Discussion 45

Conclusions 47

References 47

Supplementary Material 51

GENERAL INTRODUCTION

The *Geminiviridae* family comprises a group of viruses with a genome consisting of one or two circular, single-stranded DNA (ssDNA), encapsulated in a geminate icosahedral particle approximately 22×38 nm in size, formed by 110 subunits of a single capsid protein (CP) organized into 22 capsomers in a T=1 triangulation (Zhang *et al.*, 2001). The family is divided into 15 genera (*Becurtovirus*, *Begomovirus*, *Capulavirus*, *Citlodavirus*, *Curtovirus*, *Eragrovirus*, *Grablovirus*, *Maldovirus*, *Mastrevirus*, *Mulcrilevirus*, *Opunvirus*, *Topilevirus*, *Topocuvirus*, *Turncurtovirus, Welwivirus*) based on host range, type of insect vector, and phylogenetic relationship (Fiallo-Olive *et al.*, 2021; Debat & Bejerman, 2022).

The genus *Begomovirus*, with 445 species, is the largest and most economically important (Fiallo-Olive *et al.*, 2021; Rojas *et al.*, 2018). Begomoviruses can be divided into two major groups: those with monopartite genomes, prevalent in Europe, Africa, Asia, and Oceania (EAAO), and those with bipartite genomes, which have two genomic components of approximately 2.6 kb named DNA-A and DNA-B, prevalent in the Americas (AM) (Rybicki, 1994). The genome of monopartite begomoviruses is homologous to the DNA-A of bipartite ones (Brown *et al.*, 2015).

The two genomic components of bipartite viruses have no sequence similarity, except for an intergenic region of approximately 200 nt that shows high identity among isolates of the same species but is highly divergent among different species. This region is known as the common region (CR). Within the CR, the formation of a stem-loop is observed, containing a conserved nonanucleotide motif (5'-TAATATTAC-3') which forms the loop of the structure and contains the viral origin of replication (ori) (Fontes *et al.*, 1992).

The DNA-A of AM begomoviruses encodes five canonical genes associated with processes of replication, transcription and viral particle formation. Four of these genes (*Rep*, *REn*, *TrAP* and *AC4*) are located on the complementary strand, while the *CP* gene is located on the viral strand. The DNA-B encodes two genes, *MP* on the complementary strand and *NSP* on the viral strand, responsible for systemic movement and nuclear-cytoplasmic transport, respectively (Hanley-Bowdoin *et al.*, 1999).

These viruses use a rolling-circle replication mechanism, with the REP protein playing a crucial role. REP initiates replication by nicking the nonanucleotide sequence at the last thymine (Fontes *et al.*, 1992; Fontes, Eagle*, et al.*, 1994), recruits the host replication complex, thereby initiating the synthesis of a concatenated viral molecule, cleaves this molecule into unit-

length viral genomes, and then ligates them into circular molecules (Laufs *et al.*, 1995). Although not essential, the REn protein stabilizes the interaction of the replication complex with the stem-loop secondary structure, thereby increasing replication efficiency (Morris *et al.*, 1991; Settlage *et al.*, 2005).

Begomoviruses are transmitted in a circulative persistent manner by whiteflies of the *Bemisia tabaci* cryptic species complex (Hemiptera: Aleyrodidae). Although morphologically similar, these whitefly species exhibit variations in host preference, adaptability, and efficiency in transmitting viruses. Species demarcation within the *B. tabaci* complex is typically achieved through genetic analysis of the mitochondrial cytochrome oxidase subunit 1 (mtCOI) gene (De Barro *et al.*, 2011).

The whitefly species *B. tabaci* New World (NW, previously referred to as biotype A) predominated in Brazilian agricultural areas until the mid-1900s (Lourenção & Nagai, 1994). Following the introduction of *B. tabaci* Middle-East-Asia Minor 1 (MEAM1, previously referred to as biotype B), a rapid displacement occurred, leading to the predominance of *B. tabaci* MEAM1 throughout the country. This concurred with the emergence of diverse begomoviruses, primarily affecting solanaceous crops (Ribeiro *et al.*, 1998). Despite reports of the presence of the species *B. tabaci* Mediterranean (MED, previously referred to as biotype Q) in the country since 2014 (Barbosa *et al.*, 2015), *B. tabaci* MEAM1 remained predominant in open fields until recently (Xavier *et al.*, 2021).

Brazil is considered a center of begomovirus diversity. It is believed that the epidemics observed in agricultural crops in recent decades resulted from host jumps (spillover) from indigenous begomoviruses previously restricted to non-cultivated plants (Rocha *et al.*, 2013). These plants act as sources of inoculum, as virus reservoirs, and contribute to greater genetic variability, providing an environment where viruses can evolve and adapt (García-Arenal & Zerbini, 2019).

At the genetic level, evolution can be understood as the variation in allele frequencies over generations (Escriu, 2017). The main mechanisms generating variability are mutation, recombination, and pseudorecombination, and through evolutionary forces such as selection and genetic drift, they modify the genetic structure and variability of a population (García-Arenal *et al.*, 2001).

Mutations are the primary source of variability. A mutation is a change in the DNA sequence that can occur naturally, resulting from an error in replication, or be induced by various factors. The mutation rate refers to the frequency at which these changes are produced in the genome at each replication cycle. Many of these mutations are eliminated over time, with

only a few becoming fixed in subsequent generations (García-Arenal *et al.*, 2001). This is expressed in the nucleotide substitution rate, or evolutionary rate, measured by substitutions per site per year (Duffy *et al.*, 2008). RNA viruses have high mutation and evolutionary rates because their RNA-dependent RNA polymerases lack error correction mechanisms (Duffy *et al.*, 2008). Interestingly, and despite using the host's DNA polymerase during replication (which has error correction capacity), begomoviruses exhibit high evolutionary rates similar to those of RNA viruses (Duffy & Holmes, 2008; Duffy & Holmes, 2009).

It is well established that recombination is a major mechanism driving the evolution of geminiviruses (Lefeuvre & Moriones, 2015; Martin *et al.*, 2011; Padidam *et al.*, 1999). Recombination leads to the exchange of DNA fragments between different species or isolates. For reasons not yet fully elucidated, the *Rep* and CR regions are recombination hotspots in the begomovirus genome (Lefeuvre & Moriones, 2015). Recombination usually results in defective offspring, but occasionally it produces favorable combinations with greater fitness than the parents, which can lead to increased host range (Monci *et al.*, 2002) or changes in vector specificity (Fan *et al.*, 2023).

Bipartite begomoviruses are subject to a third mechanism of genetic variability called pseudorecombination, or reassortment, where entire genome components are exchanged between different isolates. This mechanism is more common among isolates of the same species or genetically related species (Gilbertson *et al.*, 1993). The specificity of the Rep protein binding to short direct repeat sequences in the common region is considered the main determinant for the formation of viable pseudorecombinants between begomoviruses (Arguello-Astorga *et al.*, 1994; Arguello-Astorga & Ruiz-Medrano, 2001; Fontes, Gladfelter, *et al.*, 1994).

Selection is an evolutionary mechanism whereby organisms possessing superior fitness levels are more likely to survive and reproduce, thereby transmitting their advantageous genetic traits to subsequent generations (García-Arenal *et al.*, 2001). Viruses face selective pressures throughout their entire life cycle, for example during interactions with their vectors (Dolan *et al.*, 2018). Thus, deleterious mutations are eliminated through negative or purifying selection. Conversely, when a mutation increases viral fitness, it is eventually fixed in the population through positive or diversifying selection (García-Arenal *et al.*, 2001).

Based on the neutral theory of evolution, most observed mutations are deleterious, and therefore there is a bias towards mutations that are fixed at the third position of the genetic codon and do not lead to amino acid changes as they are selectively neutral. As a result, they are maintained or removed from the population in a stochastic manner, a process known as

genetic drift (Kimura, 1983). Events that drastically reduce the size of a population, known as genetic bottlenecks, usually result in a significant loss of genetic diversity, increasing the influence of genetic drift (Escriu, 2017). However, it is difficult to separate the independent effects of neutral selection and genetic drift, since they often act together. Systemic movement, particle acquisition by the vector and spillover events are strong genetic bottlenecks exerted on viral populations, leading to reduced variability due to founder effects (García-Andrés *et al.*, 2007). The combined effects of these evolutionary mechanisms suggest to be the reason why begomovirus populations are often structured based on geographical location rather than host (Mar *et al.*, 2017; Rocha *et al.*, 2013). However, this hypothesis remains to be confirmed experimentally.

REFERENCES

ARGUELLO-ASTORGA, G.R.; GUEVARA-GONZÁLEZ, R.G.; HERRERA-ESTRELLA, L.R.; RIVERA-BUSTAMANTE, R.F. Geminivirus replication origins have a group-specific organization of interative elements: a model for replication. **Virology,** v. 203, p. 90-100, 1994.

ARGUELLO-ASTORGA, G.R.; RUIZ-MEDRANO, R. An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. **Archives of Virology,** v. 146, p. 1465-1485, 2001.

BARBOSA, L.F.; YUKI, V.A.; MARUBAYASHI, J.M.; DE MARCHI, B.R.; PERINI, F.L.; PAVAN, M.A.; BARROS, D.R.; GHANIM, M.; MORIONES, E.; NAVAS-CASTILLO, J.; KRAUSE-SAKATE, R. First report of *Bemisia tabaci* Mediterranean (Q biotype) species in Brazil. **Pest Management Science,** v. 71, p. 501-504, 2015.

BROWN, J.K.; ZERBINI, F.M.; NAVAS-CASTILLO, J.; MORIONES, E.; RAMOS-SOBRINHO, R.; SILVA, J.C.; FIALLO-OLIVE, E.; BRIDDON, R.W.; HERNANDEZ-ZEPEDA, C.; IDRIS, A.; MALATHI, V.G.; MARTIN, D.P.; RIVERA-BUSTAMANTE, R.; UEDA, S.; VARSANI, A. Revision of *Begomovirus* taxonomy based on pairwise sequence comparisons. **Archives of Virology,** v. 160, p. 1593-1619, 2015.

DE BARRO, P.J.; LIU, S.S.; BOYKIN, L.M.; DINSDALE, A.B. *Bemisia tabaci*: A statement of species status. **Annual Review of Entomology,** v. 56, p. 1-19, 2011.

DEBAT, H.; BEJERMAN, N. A glimpse into the DNA virome of the unique "living fossil" *Welwitschia mirabilis*. **Gene**, v. 843, p. 146806, 2022.

DOLAN, P.T.; WHITFIELD, Z.J.; ANDINO, R. Mechanisms and concepts in RNA virus population dynamics and evolution. **Annual Review of Virology,** v. 5, p. 69-92, 2018.

DUFFY, S.; HOLMES, E.C. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus *Tomato yellow leaf curl virus*. **Journal of Virology,** v. 82, p. 957-965, 2008.

DUFFY, S.; SHACKELTON, L.A.; HOLMES, E.C. Rates of evolutionary change in viruses: Patterns and determinants. **Nature Reviews Genetics,** v. 9, p. 267-276, 2008.

DUFFY, S.; HOLMES, E.C. Validation of high rates of nucleotide substitution in geminiviruses: Phylogenetic evidence from East African cassava mosaic viruses. **Journal of General Virology,** v. 90, p. 1539-1547, 2009.

ESCRIU, F. Diversity of plant virus populations: a valuable tool for epidemiological studies. In: BITZ, L. (Ed.). **Genetic Diversity**. London: InTechOpen, 2017. p.3-18.

FAN, Y.Y.; ZHONG, Y.W.; PAN, L.L.; WANG, X.W.; DING, M.; LIU, S.S. A shift of vector specificity acquired by a begomovirus through natural homologous recombination. **Molecular Plant Pathology,** v. 24, p. 882-895, 2023.

FIALLO-OLIVE, E.; LETT, J.M.; MARTIN, D.P.; ROUMAGNAC, P.; VARSANI, A.; ZERBINI, F.M.; NAVAS-CASTILLO, J. ICTV Virus Taxonomy Profile: *Geminiviridae* 2021. **Journal of General Virology,** v. 102, p. 001696, 2021.

FONTES, E.P.B.; LUCKOW, V.A.; HANLEY-BOWDOIN, L. A geminivirus replication protein is a sequence-specific DNA binding protein. **Plant Cell,** v. 4, p. 597-608, 1992.

FONTES, E.P.B.; EAGLE, P.A.; SIPE, P.S.; LUCKOW, V.A.; HANLEY-BOWDOIN, L. Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. **Journal of Biological Chemistry,** v. 269, p. 8459-8465, 1994.

FONTES, E.P.B.; GLADFELTER, H.J.; SCHAFFER, R.L.; PETTY, I.T.D.; HANLEY-BOWDOIN, L. Geminivirus replication origins have a modular organization. **Plant Cell,** v. 6, p. 405-416, 1994.

GARCÍA-ANDRÉS, S.; ACCOTTO, G.P.; NAVAS-CASTILLO, J.; MORIONES, E. Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. **Virology,** v. 359, p. 302-312, 2007.

GARCÍA-ARENAL, F.; FRAILE, A.; MALPICA, J.M. Variability and genetic structure of plant virus populations. **Annual Review of Phytopathology,** v. 39, p. 157-186, 2001.

GARCÍA-ARENAL, F.; ZERBINI, F.M. Life on the edge: geminiviruses at the interface between crops and wild plant hosts. **Annual Review of Virology,** v. 6, p. 411-433, 2019.

GILBERTSON, R.L.; HIDAYAT, S.H.; PAPLOMATAS, E.J.; ROJAS, M.R.; HOU, Y.-H.; MAXWELL, D.P. Pseudorecombination between infectious cloned DNA components of tomato mottle and bean dwarf mosaic geminiviruses. **Journal of General Virology,** v. 74, p. 23-31, 1993.

HANLEY-BOWDOIN, L.; SETTLAGE, S.B.; OROZCO, B.M.; NAGAR, S.; ROBERTSON, D. Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. **Critical Reviews in Plant Sciences,** v. 18, p. 71-106, 1999.

KIMURA, M. **The neutral theory of molecular evolution**. Cambridge: Cambridge University Press, 1983. 367

LAUFS, J.; TRAUT, W.; HEYRAUD, F.; MATZEIT, G.; ROGERS, S.G.; SCHELL, J.; GRONENBORN, B. *In vitro* cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. **Proceedings of the National Academy of Sciences, USA,** v. 92, p. 3879-3883, 1995.

LEFEUVRE, P.; MORIONES, E. Recombination as a motor of host switches and virus emergence: Geminiviruses as case studies. **Current Opinion in Virology,** v. 10, p. 14-19, 2015.

LOURENÇÃO, A.L.; NAGAI, H. Surtos populacionais de *Bemisia tabaci* no Estado de São Paulo. **Bragantia,** v. 53, p. 53-59, 1994.

MAR, T.B.; XAVIER, C.A.D.; LIMA, A.T.M.; NOGUEIRA, A.M.; SILVA, J.C.F.; RAMOS-SOBRINHO, R.; LAU, D.; ZERBINI, F.M. Genetic variability and population structure of the New World begomovirus *Euphorbia yellow mosaic virus*. **Journal of General Virology,** v. 98, p. 1537-1551, 2017.

MARTIN, D.P.; LEFEUVRE, P.; VARSANI, A.; HOAREAU, M.; SEMEGNI, J.Y.; DIJOUX, B.; VINCENT, C.; REYNAUD, B.; LETT, J.M. Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. **PLOS Pathogens,** v. 7, p. e1002203, 2011.

MONCI, F.; SANCHEZ-CAMPOS, S.; NAVAS-CASTILLO, J.; MORIONES, E. A natural recombinant between the geminiviruses tomato yellow leaf curl Sardinia virus and tomato yellow leaf curl virus exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. **Virology,** v. 303, p. 317-326, 2002.

MORRIS, B.; RICHARDSON, K.; EDDY, P.; ZHAN, X.; HALEY, A.; GARDNER, R. Mutagenesis of the AC3 open reading frame of African cassava mosaic virus DNA-A reduces DNA-B replication and ameliorates disease symptoms. **Journal of General Virology,** v. 72, p. 1205-1213, 1991.

PADIDAM, M.; SAWYER, S.; FAUQUET, C.M. Possible emergence of new geminiviruses by frequent recombination. **Virology,** v. 265, p. 218-224, 1999.

RIBEIRO, S.G.; ÁVILA, A.C.; BEZERRA, I.C.; FERNANDES, J.J.; FARIA, J.C.; LIMA, M.F.; GILBERTSON, R.L.; ZAMBOLIM, E.M.; ZERBINI, F.M. Widespread occurrence of tomato geminiviruses in Brazil, associated with the new biotype of the whitefly vector. **Plant Disease,** v. 82, p. 830, 1998.

ROCHA, C.S.; CASTILLO-URQUIZA, G.P.; LIMA, A.T.M.; SILVA, F.N.; XAVIER, C.A.D.; HORA-JUNIOR, B.T.; BESERRA-JUNIOR, J.E.A.; MALTA, A.W.O.; MARTIN, D.P.; VARSANI, A.; ALFENAS-ZERBINI, P.; MIZUBUTI, E.S.G.; ZERBINI, F.M. Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. **Journal of Virology,** v. 87, p. 5784-5799, 2013.

ROJAS, M.R.; MACEDO, M.A.; MALIANO, M.R.; SOTO-AGUILAR, M.; SOUZA, J.O.; BRIDDON, R.W.; KENYON, L.A.; RIVERA-BUSTAMANTE, R.F.; ZERBINI, F.M.; ADKINS, S.; LEGG, J.P.; KVARNHEDEN, A.; WINTERMANTEL, W.M.; SUDARSHANA, M.R.; PETERSCHMITT, M.; LAPIDOT, M.; MARTIN, D.P.; MORIONES, E.; INOUE-NAGATA, A.K.; GILBERTSON, R.L. World management of geminiviruses. **Annual Review of Phytopathology,** v. 56, p. 637-677, 2018.

RYBICKI, E.P. A phylogenetic and evolutionary justification for three genera of *Geminiviridae*. **Archives of Virology,** v. 139, p. 49-77, 1994.

SETTLAGE, S.B.; SEE, R.G.; HANLEY-BOWDOIN, L. Geminivirus C3 protein: Replication enhancement and protein interactions. **Journal of Virology,** v. 79, p. 9885-9895, 2005.

XAVIER, C.A.D.; NOGUEIRA, A.M.; BELLO, V.H.; WATANABE, L.F.M.; BARBOSA, T.M.C.; ALVES-JÚNIOR, M.; BARBOSA, L.; BESERRA-JÚNIOR, J.E.A.; BOARI, A.; CALEGARIO, R.; GORAYEB, E.S.; HONORATO-JÚNIOR, J.; KOCH, G.; LIMA, G.S.A.; LOPES, C.; MELLO, R.N.; PANTOJA, K.; SILVA, F.N.; RAMOS-SOBRINHO, R.; SANTANA, E.N.; SILVA, J.W.P.; KRAUSE-SAKATE, R.; ZERBINI, F.M. Assessing the diversity of whiteflies infesting cassava in Brazil. **PeerJ,** v. 9, p. e11741, 2021.

ZHANG, W.; OLSON, N.H.; BAKER, T.S.; FAULKNER, L.; AGBANDJE-MCKENNA, M.; BOULTON, M.I.; DAVIES, J.W.; MCKENNA, R. Structure of the *Maize streak virus* geminate particle. **Virology,** v. 279, p. 471-477, 2001.

CHAPTER 1

BIOLOGICAL AND MOLECULAR CHARACTERIZATION OF A BRAZILIAN ISOLATE OF BEAN BUSHY STUNT VIRUS (*Begomovirus phaseoliretorridi*)

GONÇALVES, M.H.O., SANTOS, G.A.S., QUADROS, A.F.F., ZERBINI, F.M., BLAWID, R. Biological and molecular characterization of a brazilian isolate of bean bushy stunt virus (*Begomovirus phaseoliretorridi*). Archives of Virology, *in preparation*.

Biological and molecular characterization of a Brazilian isolate of bean bushy stunt virus (*Begomovirus phaseoliretorridi*)

Marcelo Henrique Oliveira Gonçalves[1], Grenville Antonio Souza Dos Santos[2], Ayane Fernanda Ferreira Quadros[1], Francisco Murilo Zerbini[1], Rosana Blawid[2]

[1]Dep. de Fitopatologia/BIOAGRO, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[2]Dep. de Agronomia, Universidade Federal Rural de Pernambuco, Recife, Pernambuco, Brazil

ABSTRACT

Bean bushy stunt virus (*Begomovirus phaseoliretorridi*, BBSV; family *Geminiviridae*) was described infecting common bean in Argentina, and has recently been detected in samples of *Neustanthus phaseoloides* (tropical kudzu) in Acre state, northern Brazil. The aim of this research was to characterize the isolate at the molecular level and to evaluate its pathogenic potential in bean and soybean. The taxonomic classification was reassessed according to the criteria established by the ICTV. The analyses demonstrated that the isolate GS-20, sequenced in a previous study, represents a new strain of *Begomovirus phaseoliretorridi*. An interspecies recombination event was identified in the coat protein (CP) of the virus, with soybean chlorotic spot virus inferred as the minor parent. Total DNA from the original sample was extracted and used for rolling-circle amplification (RCA). The RCA product was used for biolistic inoculation of eight different cultivars of common bean and soybean. Symptoms of leaf curling and foliar deformation were observed in bean plants at a low infectivity rate. Notably, soybean plants could not be infected. These findings provide a basis for discussing the origin of the virus in Brazil.

Key words: Geminivirus; Recombination; Virus evolution

INTRODUCTION

Begomovirus is a genus within the family *Geminiviridae*, whose members are characterized by a circular, single-stranded DNA genome (~2.6 kb or 5.2 kb for viruses with mono- and bipartite genomes, respectively) and a unique morphology of twinned quasi-icosahedral particles [9]. Begomoviruses from the Americas are mostly bipartite, with two componentes named DNA-A and DNA-B. Those from Africa, Asia, Europe and Oceania are mostly monopartite, with their component homologous to the DNA-A of bipartite viruses, and are often accompanied by satellite DNAs [4].

The DNA-A of bipartite begomoviruses contains five canonical genes. Four of these genes (*Rep*, *REn*, *TrAP* and *AC4*), located on the complementary strand, are involved in replication, transcription, and inhibition of RNA silencing. The *CP* gene, located on the viral strand, is associated with particle formation and vector transmission. The DNA-B encodes two genes: *MP* on the complementary strand, responsible for systemic movement, and *NSP* on the viral strand, involved in nucleo-cytoplasmic transport [11].

Begomoviruses are transmitted by whiteflies of the *Bemisia tabaci* cryptic species complex [4], and comprise one of the most economically significant groups of plant viruses owing to their widespread occurrence and severe impact on major crop plants in tropical and subtropical regions worldwide [32].

A great diversity of begomoviruses is found in Brazil. Many of them affect cultivated plants, mostly common bean (*Phaseolus vulgaris*) and tomato (*Solanum lycopersicum*) [14], while others infect non-cultivated plants, mostly in the Euphorbiaceae, Fabaceae, Malvaceae and Solanaceae families [16, 17, 19, 22, 24-27, 35, 36]. Begomoviruses from non-cultivated plants are believed to have originated those in crops, and spillover events from non-cultivated to cultivated plants have been observed [12, 30].

In 2017, a new begomovirus, named bean bushy stunt virus (*Begomovirus phaseoliretorridi*, BBSV), was described causing leaf roll and stunting in common bean plants in Argentina [29]. In December 2020, in the state of Acre, a tropical kudzu plant (*Neustanthus phaseoloides*) exhibiting severe mosaic symptoms was found to be infected with this virus [34]. Considering that this represents the first report of this virus in the country, the aim of this research was to characterize this isolate at the molecular level and to evaluate its pathogenic potential in common bean and soybean (*Glycine max*) crops.

METHODS

Sample

The *N. phaseoloides* sample identified as GS-20 was subjected to prior high-throughput sequencing (HTS) and the full genome of BBSV was assembled from the raw sequence data [34]. Sequence analysis using Sequence Demarcation Tool (SDT) v. 1.3 [21] was conducted to confirm the viral identity and classification. A multiple sequence alignment was performed using the MUSCLE algorithm [7], on two datasets: one for the DNA-A (Suppl. Table S1) and one for the DNA-B (Suppl. Table S2). A threshold of 91% nucleotide sequence identity was used to delineate isolates of the same species, while 94% identity was used to define isolates of the same strain, according to the guidelines of the *Geminiviridae* and *Tolecusatellitidae* Study Group of the International Committee on Taxonomy of Viruses (ICTV) [4].

Phylogenetic and recombination analysis

A multiple sequence alignment was performed using the MAFFT algorithm [15] implemented in the online server (https://mafft.cbrc.jp). The best nucleotide substitution model was determined using MrModelTest v. 2 [23] in the Akaike Information Criterion (AIC) and used for the construction of a Bayesian phylogenetic tree in MrBayes v. 3.2 [33] available in the CIPRES server [20]. This was conducted in two independent analyses, each with 10,000,000 generations and visualized in Figtree (tree.bio.ed.ac.uk/software/figtree/). Recombination events were verified in the RDP5 [18]. Each one of the seven methods implemented in the program was configured with default parameters, and significance was determined with a Bonferroni-corrected *p*-value set at 0.05. Only recombination events identified by a minimum of four methods were deemed reliable.

Pathogenicity test

Total DNA from the GS-20 sample was extracted following the protocol described by Doyle and Doyle [6]. Subsequently, viral DNA was amplified using rolling circle amplification (RCA) with the phi29 DNA polymerase [13]. To verify the amplification of both components, the RCA product was diluted and subjected to PCR amplification using degenerate primers for the detection of DNA-A (PAL1v1978 and PAR1c496) and DNA-B (PBL1v2040 and PCRc1) [31]. The five RCA product (approximately 5µg) was fragmented using a sonicator and associated with gold particles, and then biolistically inoculated into the radicles of newly germinated seedlings [1]. Four common bean cultivars, three from carioca group (BRS Estilo, BRS Pérola, IPR Celeiro) and one from black group (Ouro Negro) and four soybean (*Glycine*

max) cultivars (BRS 133, BRS 213, Cariri and TMG 7063 IPRO) were used. Mock-inoculated plants were bombarded with gold particles without DNA. Plants were grown in pots inside cages with insect-proof mesh. They were evaluated for the presence of symptoms at 14- and 28-days post-inoculation (dpi), and viral detection was performed by PCR amplification using the previously mentioned primers. The experiment was repeated twice.

RESULTS AND DISCUSSION

The viral isolate GS-20, identified in Acre state, Brazil, exhibits a nucleotide sequence identity of 92.9% for the DNA-A (Suppl. Figure S1) and 92.7% for the DNA-B (Suppl. Figure S2) with the Argentine BBSV (GenBank accession numbers MN414067 and MN414068 for the DNA-A and DNA-B, respectively). According to the species and strain demarcation criteria for the genus *Begomovirus* [4], this level of nucleotide identity indicates that isolate GS-20 represents a distinct strain within the species *B. phaseoliretorridi*.

Phylogenetic analysis indicated that the DNA-A sequences from the two BBSV isolates form a distinct clade, with a closer phylogenetic relationship with muntingia yellow spot virus (MuYSV) and pepper leafroll virus (*Begomovirus capsicumcontorsioris*, PepLRV), collectively constituting a sister clade (Figure 1a). However, analysis of the DNA-B indicated that, while the two BBSV isolates also cluster together, they form a sister clade with squash leaf curl virus (*B. cucurbitapeponis*, SLCuV) (Figure 1b). This indicates a distinct evolutionary history between the two genomic components of BBSV, consistent with previously observations within the genus [3, 37].

Nucleotide identities of the viral genes of the two isolates ranged from 84.8% to 98.6% (Table 1). Notably, the *CP* gene showed the lowest identity (84.8%). While most proteins had high amino acid sequence identity, ranging from 92% to 97.9%, the CP exhibited significantly lower identity, at only 84.5%. The result was unexpected as the CP is the most conserved begomovirus gene/protein, supposedly due to being subjected to strong purifying selection [10, 37].

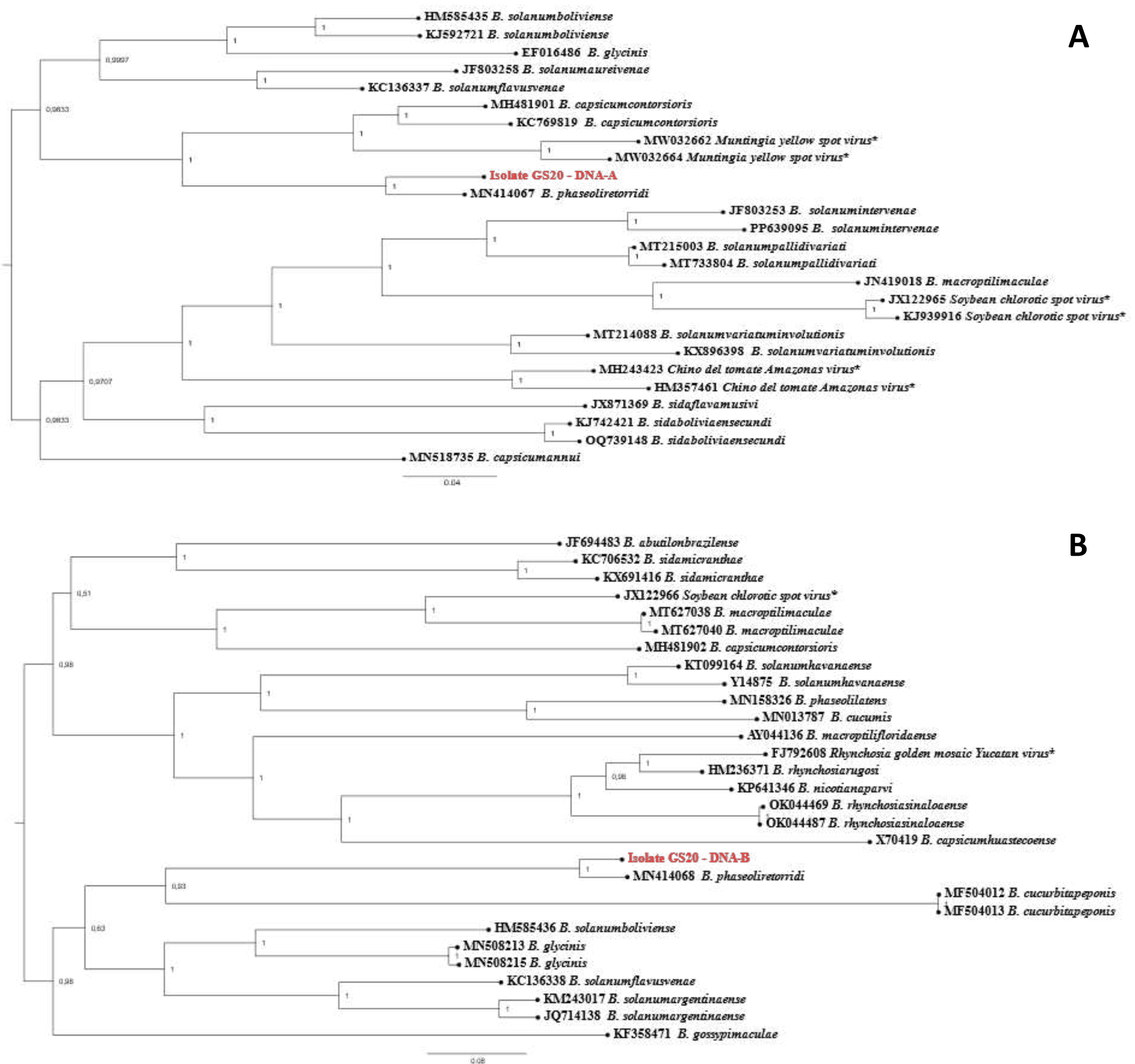


Figure 1. Phylogenetic trees based on the DNA-A (**A**) and DNA-B (**B**) of the GS-20 isolate (in red) and various begomoviruses. The trees were generated using Bayesian inference, with posterior probability values indicated at the branches. *Proposed species not yet accepted by the ICTV.

Table 1. Pairwise nucleotide and amino acid sequence identity between the genes and protein products, respectively, of bean bushy stunt virus (BBSV) isolates GS-20 (from Brazil) and General Mosconi (from Argentina).

	DNA-A					DNA-B	
	Rep	**AC4**	**TrAP**	**REn**	**CP**	**MP**	**NSP**
Nucleic acid identity	97.5	97.8	98.6	97.8	84.8	89.2	94
Amino acid identity	97.2	94.8	97.7	97.7	89.9	97.9	95.8

The results of sequence comparisons and phylogenetic analysis led us to hypothesize that the GS-20 isolate might be a recombinant. Thus, recombination analysis was performed, and five methods (RDP, MaxChi, Chimaera, SisterScan, 3Seq) consistently identified a recombination event spanning nucleotide positions 215 to 896 in the DNA-A. BBSV was inferred as the major parent and soybean chlorotic spot virus (SoCSV) as the minor parent. SoCSV was detected infecting leguminous plants in the state of Minas Gerais in 2010 and 2011 [5, 28]. The recombinant region, encompassing approximately 75% of the *CP* gene, accounts for its greater divergence compared to other genes. No recombination events were detected in the DNA-B.

To assess the pathogenic potential of this new strain in bean and soybean crops, four cultivars of each species were inoculated by particle bombardment (Table 2). Symptoms were observed only in two common bean plants from cultivars Celeiro and Pérola, where the presence of both viral components was detected by PCR (data not shown). This corresponds to an infectivity rate of 10% for each of these cultivars. The symptoms observed were consistent with those previously documented by Reyna, et al. [29], including leaf distortion, reduced leaf blade size, shortened internodes and mild mosaic. Additionally, this experiment noted the occurrence of flower abortion (Figure 2).

Table 2. Results of pathogenicity test of the bean bushy stunt virus (BBSV) isolate GS-20 in common bean and soybean cultivars.

Species	Cultivar	Infected/Inoculated
Common bean (*Phaseolus vulgaris*)	BRS Estilo	0/10
	BRS Pérola	1/8
	IPR Celeiro	1/10
	Ouro Negro	0/10
Soybean (*Glycine max*)	BRS 133	0/10
	BRS 213	0/10
	Cariri	0/10
	TMG 7063 IPRO	0/10

Figure 2. Symptoms of the bean bushy stunt virus (BBSV) isolate GS-20 in common bean (*Phaseolus vulgaris*) cvs. Pérola (**A, B**) and Celeiro (**C-F**). **A**, healthy plant; **B**, infected plant; **C**, leaf of an infected plant showing deformation and leaf blade reduction on the left, and healthy leaf on the right; **D**, flower of a healthy plant; **E**, aborted flowers in an infected plant; **F**, leaf deformation and mosaic in an infected plant.

Despite unsuccessful attempts to infect soybeans with the GS-20 isolate, previous studies have demonstrated the pathogenic potential of BBSV in this crop. It is unlikely that the recombination event detected in CP would lead to a loss of pathogenicity in this species, as recombinants that acquire the CP from a minor parent are reported to maintain the level of infectivity of the major parent [8]. Hence, the outcome observed could be due to our experimental conditions or divergent evolutionary processes in other genetic loci, resulting in the loss of pathogenicity in this host.

Despite the recent report of BBSV in Brazil [34], the hypothesis suggesting its native status in non-cultivated hosts appears more plausible than that of a recent introduction. This is consistent with its low infectivity in two cultivated leguminous plants (Table 2), potentially making its detection challenging, considering the bias in the knowledge of viral diversity towards plants of agricultural significance [2].

In conclusion, the BBSV isolate GS-20 is member of a distinct strain of the species *Begomovirus phaseoliretorridi*, isolate from a non-cultivated plant. The isolate has a recombinant coat protein, and our results indicate a low infectivity in common bean and the absence of pathogenicity in soybean. However, further inoculation experiments are necessary to eliminate potential methodological issues and validate these results.

REFERENCES

1. Aragão FJL, Barros LMG, Brasileiro ACM, Ribeiro SG, Smith FD, Sanford JC, Faria JC, Rech EL (1996) Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris* L.) co-transformed via particle bombardment. Theor Appl Genet 93:142-150
2. Bernardo P, Charles-Dominique T, Barakat M, Ortet P, Fernandez E, Filloux D, Hartnady P, Rebelo TA, Cousins SR, Mesleard F, Cohez D, Yavercovski N, Varsani A, Harkins GW, Peterschmitt M, Malmstrom CM, Martin DP, Roumagnac P (2018) Geometagenomics illuminates the impact of agriculture on the distribution and prevalence of plant viruses at the ecosystem scale. ISME J 12:173-184
3. Briddon RW, Patil BL, Bagewadi B, Nawaz-ul-Rehman MS, Fauquet CM (2010) Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. BMC Evol Biol 10:97
4. Brown JK, Zerbini FM, Navas-Castillo J, Moriones E, Ramos-Sobrinho R, Silva JC, Fiallo-Olive E, Briddon RW, Hernandez-Zepeda C, Idris A, Malathi VG, Martin DP, Rivera-Bustamante R, Ueda S, Varsani A (2015) Revision of *Begomovirus* taxonomy based on pairwise sequence comparisons. Arch Virol 160:1593-1619
5. Coco D, Calil IP, Brustolini OJB, Santos AA, Inoue-Nagata AK, Fontes EPB (2013) Soybean chlorotic spot virus, a novel begomovirus infecting soybean in Brazil. Arch Virol 158:457-462
6. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem Bull 19:11-15

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

8. Fan YY, Zhong YW, Pan LL, Wang XW, Ding M, Liu SS (2023) A shift of vector specificity acquired by a begomovirus through natural homologous recombination. Mol Plant Pathol 24:882-895

9. Fiallo-Olive E, Lett JM, Martin DP, Roumagnac P, Varsani A, Zerbini FM, Navas-Castillo J (2021) ICTV Virus Taxonomy Profile: *Geminiviridae* 2021. J Gen Virol 102:001696

10. García-Arenal F, Fraile A, Malpica JM (2001) Variability and genetic structure of plant virus populations. Annu Rev Phytopathol 39:157-186

11. Gutierrez C, Ramirez-Parra E, Castellano MM, Sanz-Burgos AP, Luque A, Missich R (2004) Geminivirus DNA replication and cell cycle interactions. Vet Microbiol 98:111-119

12. Hoffmann LV, Inoue-Nagata AK, Vaz LNA, Barroso PAV, De Faria JC (2021) Identification of sida micrantha mosaic virus as the causal agent of common mosaic in cotton in Goiás. Summa Phytopathol 47:222-224

13. Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T (2004) A simple method for cloning the complete begomovirus genome using the bacteriophage phi29 DNA polymerase. J Virol Met 116:209-211

14. Inoue-Nagata AK, Lima MF, Gilbertson RL (2016) A review of geminivirus diseases in vegetables and other crops in Brazil: current status and approaches for management. Hortic Bras 34:8-18

15. Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30:772-780

16. Macedo MA, Rego-Machado CM, Maliano ML, Rojas MR, Inoue-Nagata AK, Gilbertson RL (2020) Complete sequence of a new bipartite begomovirus infecting Sida sp. in Northeastern Brazil. Arch Virol 165:253-256

17. Mar TB, Xavier CAD, Lima ATM, Nogueira AM, Silva JCF, Ramos-Sobrinho R, Lau D, Zerbini FM (2017) Genetic variability and population structure of the New World begomovirus *Euphorbia yellow mosaic virus*. J Gen Virol 98:1537-1551

18. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, Kelz Z, Kumar V, Murrell B (2021) RDP5: A computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. Virus Evol 7:veaa087

19. Melo AM, Silva SJ, Ramos-Sobrinho R, Ferro MM, Assuncao IP, Lima GS (2016) Cnidoscolus mosaic leaf deformation virus: a novel begomovirus infecting euphorbiaceous plants in Brazil. Arch Virol 161:2605-2608

20. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE):1-8

21. Muhire BM, Varsani A, Martin DP (2014) SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLOS One 9:e108277

22. Nascimento LD, Silva SJ, Sobrinho RR, Ferro MM, Oliveira MH, Zerbini FM, Assunção IP, Lima GS (2016) Complete nucleotide sequence of a new begomovirus infecting a malvaceous weed in Brazil. Arch Virol 161:1735-1738

23. Nylander JAA (2004) MrModeltest v2. Program distributed by the author Evolutionary Biology Centre, Uppsala University

24. Passos LS, Rodrigues JS, Soares ÉCS, Silva JP, Zerbini FM, Araújo ASF, Beserra JEA (2017) Complete genome sequence of a new bipartite begomovirus infecting *Macroptilium lathyroides* in Brazil. Arch Virol 162:3551-3554

25. Passos LS, Teixeira JWM, Teixeira KJML, Xavier CAD, Zerbini FM, Araújo ASF, Beserra JEA (2017) Two new begomoviruses that infect non-cultivated malvaceae in Brazil. Arch Virol 162:1795-1797

26. Pinto VB, Silva JP, Fiallo-Olivé E, Navas-Castillo J, Zerbini FM (2016) Novel begomoviruses recovered from *Pavonia* sp. in Brazil. Arch Virol 161:735-739

27. Quadros AFF, Silva JP, Xavier CAD, Zerbini FM, Boari AJ (2019) Two new begomoviruses infecting tomato and Hibiscus sp. in the Amazon region of Brazil. Arch Virol 164:1897-1901

28. Ramos-Sobrinho R, Xavier CAD, Pereira HMB, Lima GSA, Assunção IP, Mizubuti ESG, Duffy S, Zerbini FM (2014) Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts. J Gen Virol 95:2540-2552

29. Reyna PG, Bejerman N, Laguna IG, Pardina PR (2021) Biological and molecular characterization of bean bushy stunt virus, a novel bipartite begomovirus infecting common bean in northwestern Argentina. Arch Virol 166:1409-1414

30. Rocha CS, Castillo-Urquiza GP, Lima ATM, Silva FN, Xavier CAD, Hora-Junior BT, Beserra-Junior JEA, Malta AWO, Martin DP, Varsani A, Alfenas-Zerbini P, Mizubuti ESG, Zerbini FM (2013) Brazilian begomovirus populations are highly recombinant, rapidly evolving, and segregated based on geographical location. J Virol 87:5784-5799

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

32. Rojas MR, Macedo MA, Maliano MR, Soto-Aguilar M, Souza JO, Briddon RW, Kenyon LA, Rivera-Bustamante RF, Zerbini FM, Adkins S, Legg JP, Kvarnheden A, Wintermantel WM, Sudarshana MR, Peterschmitt M, Lapidot M, Martin DP, Moriones E, Inoue-Nagata AK, Gilbertson RL (2018) World management of geminiviruses. Annu Rev Phytopathol 56:637-677

33. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539-542

34. Santos GASd (2021) Estudo da diversidade viral em maracujazeiro e outras espécies de plantas no estado do Acre. Dep de Agronomia. Universidade Federal Rural de Pernambuco, Recife, Brazil, p 98

35. Silva SJC, Castillo-Urquiza GP, Hora-Junior BT, Assunção IP, Lima GSA, Pio-Ribeiro G, Mizubuti ESG, Zerbini FM (2012) Species diversity, phylogeny and genetic variability of begomovirus populations infecting leguminous weeds in northeastern Brazil. Plant Pathol 61:457-467

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

37. Xavier CAD, Godinho MT, Mar TB, Ferro CG, Sande OFL, Silva JC, Ramos-Sobrinho R, Nascimento RN, Assuncao I, Lima GSA, Lima ATM, Zerbini FM (2021) Evolutionary dynamics of bipartite begomoviruses revealed by complete genome analysis. Mol Ecol 15:3747-3767

SUPPLEMENTARY MATERIAL

1 **Suppl. Table S1.** DNA-A sequences of begomoviruses used in this study.

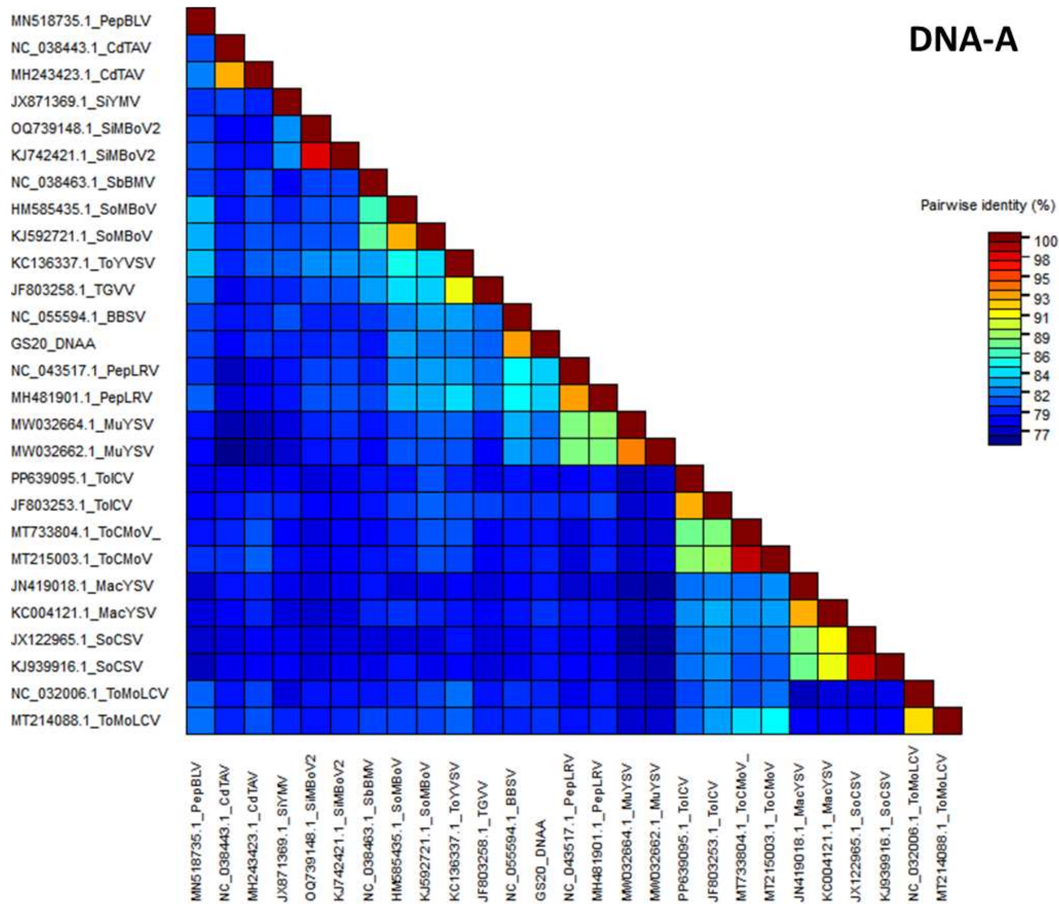
GenBank accession number	Virus	Species	Abbreviation
MN414067	bean bushy stunt virus	*Begomovirus phaseoliretorridi*	BBSV
HM357461	chino del tomate Amazonas virus	*B. solanumamazonasense*	CdTAV
MH243423	chino del tomate Amazonas virus	*B. solanumamazonasense*	CdTAV
JN419018	Macroptilium yellow spot virus	*B. macroptilimaculae*	MacYSV
KC004121	Macroptilium yellow spot virus	*B. macroptilimaculae*	MacYSV
MW032664	Muntingia yellow spot virus*		MuYSV
MW032662	Muntingia yellow spot virus*		MuYSV
MN518735	pepper blistering leaf virus	*B. capsicumannui*	PepBLV
KC769819	pepper leafroll virus	*B. capsicumcontorsioris*	PepLRV
MH481901	pepper leafroll virus	*B. capsicumcontorsioris*	PepLRV
EF016486	soybean blistering mosaic virus	*B. glycinis*	SbBMV
OQ739148	Sida mosaic Bolivia virus 2	*B. sidaboliviaensecundi*	SiMBoV2
KJ742421	Sida mosaic Bolivia virus 2	*B. sidaboliviaensecundi*	SiMBoV2
JX871369	Sida yellow mosaic virus	*B. sidaflavamusivi*	SiYMV
JX122965	soybean chlorotic spot virus*		SoCSV
KJ939916	soybean chlorotic spot virus*		SoCSV
HM585435	Solanum mosaic Bolivia virus	*B. solanumboliviense*	SoMBoV
KJ592721	Solanum mosaic Bolivia virus	*B. solanumboliviense*	SoMBoV
JF803258	tomato golden vein virus	*B. solanumaureivenae*	TGVV
MT733804	tomato chlorotic mottle virus	*B. solanumpallidivariati*	ToCMoV
MT215003	tomato chlorotic mottle virus	*B. solanumpallidivariati*	ToCMoV
PP639095	tomato interveinal chlorosis virus	*B. solanumintervenae*	ToICV
JF803253	tomato interveinal chlorosis virus	*B. solanumintervenae*	ToICV
KX896398	tomato mottle leaf curl virus	*B. solanumvariatuminvolutionis*	ToMoLCV
MT214088	tomato mottle leaf curl virus	*B. solanumvariatuminvolutionis*	ToMoLCV
KC136337	tomato yellow vein streak virus	*B. solanumflavusvenae*	ToYVSV

2 * Proposed species not yet accepted by the ICTV

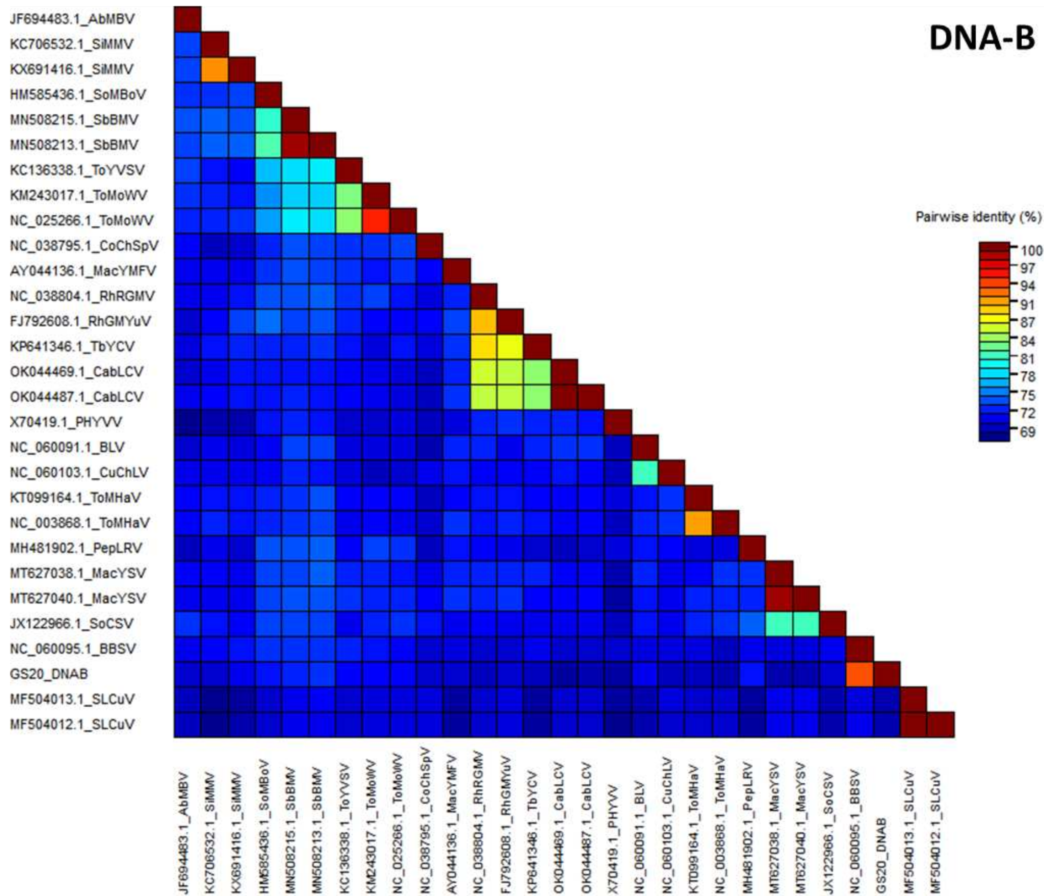
3 **Suppl. Table S2.** DNA-B sequences of begomoviruses used in this study.

GenBank accession number	Virus	Species	Abbreviation
JF694483	Abutilon mosaic Brazil virus	*Begomovirus abutilonbrazilense*	AbMBV
MN414068	bean bushy stunt virus	*B. phaseoliretorridi*	BBSV
MN158326	bean latent virus isolate	*B. phaseolilatens*	BLV
OK044469	cabbage leaf curl virus	*B. rhynchosiasinaloaense*	CabLCV
OK044487	cabbage leaf curl vírus	*B. rhynchosiasinaloaense*	CabLCV
KF358471	cotton chlorotic spot virus	*B. gossypimaculae*	CoChSpV
MN013787	cucumber chlorotic leaf virus	*B. cucumis*	CuChLV
AY044136	Macroptilium yellow mosaic Florida virus	*B. macroptilifloridaense*	MacYMFV
MT627038	Macroptilium yellow spot virus	*B. macroptilimaculae*	MacYSV
MT627040	Macroptilium yellow spot virus	*B. macroptilimaculae*	MacYSV
MH481902	pepper leafroll virus isolate	*B. capsicumcontorsioris*	PepLRV
X70419	pepper huasteco yellow vein virus	*B. capsicumhuastecoense*	PHYVV
FJ792608	Rhynchosia golden mosaic Yucatan virus*		RhGMYuV
HM236371	Rhynchosia rugose golden mosaic virus	*B. rhynchosiarugosi*	RhRGMV
MN508215	soybean blistering mosaic virus	*B. glycinis*	SbBMV
MN508213	soybean blistering mosaic virus	*B. glycinis*	SbBMV
KC706532	Sida micrantha mosaic virus	*B. sidamicranthae*	SiMMV
KX691416	Sida micrantha mosaic virus	*B. sidamicranthae*	SiMMV
MF504013	squash leaf curl virus	*B. cucurbitapeponis*	SLCuV
MF504012	squash leaf curl virus	*B. cucurbitapeponis*	SLCuV
JX122966	soybean chlorotic spot virus*		SoCSV
HM585436	Solanum mosaic Bolivia virus	*B. solanumboliviense*	SoMBoV
KP641346	tobacco yellow crinkle virus	*B. nicotianaparvi*	TbYCV
KT099164	tomato mosaic Havana virus	*B. solanumhavanaense*	ToMHaV
Y14875	tomato mosaic Havana virus	*B. solanumhavanaense*	ToMHaV
KM243017	tomato mottle wrinkle virus	*B. solanumargentinaense*	ToMoWV
JQ714138	tomato mottle wrinkle virus	*B. solanumargentinaense*	ToMoWV
KC136338	tomato yellow vein streak virus	*B. solanumflavusvenae*	ToYVSV

4 * Proposed species not yet accepted by the ICTV



Suppl. Figure S1. Pairwise comparison matrix of begomovirus DNA-A sequences, including the bean bushy stunt virus (BBSV) isolate GS-20 and the most closely related begomoviruses. Prepared with Sequence Demarcation Tool (SDT) v. 1.3.



Suppl. Figure S2. Pairwise comparison matrix of begomovirus DNA-B sequences, including the bean bushy stunt virus (BBSV) isolate GS-20 and the most closely related begomoviruses. Prepared with Sequence Demarcation Tool (SDT) v. 1.3.

CHAPTER 2

GENETIC VARIABILITY OF BLAINVILLEA YELLOW SPOT VIRUS (*Begomovirus blainvilleae*) REVEALS RECOMBINANT VARIANTS AND A DISTINCT NONANUCLEOTIDE MOTIF

GONÇALVES, M.H.O., QUADROS, A.F.F., FERRO, M. M. M., ORÍLIO, A.F., SILVA, J.P.H., LAGE, J.B., SILVA, S.J.C., ZERBINI, F.M. Genetic variability of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) reveals recombinant variants and a distinct nonanucleotide motif. Journal of General Virology, *in preparation*.

Genetic variability of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) reveals recombinant variants and a distinct nonanucleotide motif

Marcelo H.O. Gonçalves[1], Ayane F.F. Quadros[1], Mayra M. M. Ferro[2], Anelise F. Orílio[1], João Paulo H. Silva[1], Julia B. Lage[1], Sarah J.C. Silva[2], F. Murilo Zerbini[1]

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

[2]Setor de Fitossanidade, Campus de Engenharias e Ciências Agrárias, Universidade Federal de Alagoas, Rio Largo, AL, Brazil, 57100-000

ABSTRACT

Viruses from wild plants often lead to the emergence of new economically important agricultural pathogens through frequent spillover events. Blainvillea yellow spot virus (*Begomovirus blainvilleae*, BlYSV) is a bipartite geminivirus which is ubiquitous in the non-cultivated plant *Blainvillea rhomboidea*. To better understand the ecological and evolutionary aspects of this virus, we sampled *B. rhomboidea* in Minas Gerais, Rio Grande do Norte, and Alagoas states between 2022 and 2023. Following rolling-circle amplification, cloning and Sanger sequencing, a dataset was assembled with 23 complete DNA-A sequences plus an additional 30 BlYSV sequences retrieved from GenBank. Four variants (named A to D) were identified based on sequence comparisons, one of which (A) was classified as a different strain (<94% nucleotide sequence identity for the complete DNA-A segment). These isolates were structured as five independent subpopulations. It was possible to observe an alteration in the distribution pattern of variants in the municipalities of Viçosa and Coimbra. Five recombination events were identified and Four sequences in strain A exhibited a nucleotide substitution (A2667G) within the geminivirus conserved nonanucleotide motif (from TAATATTAC to TAATGTTAC). This mutation does not alter stem-loop formation. Interestingly, this subpopulation has a lower degree of genetic variability compared to the others.

Key words: Geminivirus; Virus evolution; Non-cultivated hosts; Population structure.

INTRODUCTION

The ability of viruses to evolve at a faster rate than observed in other organisms allows them to rapidly adapt to environmental changes and exploit new niches (1). The study of virus populations in non-cultivated hosts is relevant from both basic and applied perspectives: unmanaged ecosystems allow the study of virus evolution in the absence of cultivation-related bottlenecks, and introduced crops may be infected by "indigenous" viruses present in non-cultivated hosts (2).

The family *Geminiviridae* includes plant viruses characterized by twinned quasi-icosahedral particles and a single-stranded, circular DNA genome (3). These viruses are noted for their high evolutionary rates, similar to those of RNA viruses (4). Currently, 14 genera and over 500 species are recognized by the International Committee on Taxonomy of Viruses (ICTV), with new species being continuously reported (5). Most of these species belong to the genus *Begomovirus* and are naturally transmitted by whiteflies of the *Bemisia tabaci* cryptic species complex. The majority of begomoviruses found in the Americas (AM) have two genomic components, named DNA-A and DNA-B, each approximately 2,600 nucleotides (nt) long. In contrast, begomoviruses in Europe, Africa, Asia and Oceania (EAAO) may be mono- or bipartite (6).

Typical AM bipartite begomoviruses have five genes in the DNA-A (*CP*, *Rep*, *REn*, *TrAP* and *AC4*) involved in viral replication, particle formation, and suppression of RNA silencing. The DNA-B, in turn, encodes two proteins (NSP and MP) associated with nuclear transport and systemic movement in the plant. The two segments do not share sequence identity except for an intergenic region, referred to as the common region (CR), which is highly variable among different species but highly conserved in the DNA-A and DNA-B of the same species (7). This region includes a hairpin structure with a nonanucleotide motif (5'-TAATATTAC-3') in the loop region, which is the origin of rolling-circle viral replication (8). This nonanucleotide is highly conserved across the *Geminiviridae*, except in the genera *Becurtovirus*, *Eragrovirus*, and one member of the genus *Mastrevirus*, which harbor an alternative sequence (5'-TAAGATTCC-3') (9) (10). Furthermore, two begomoviruses exhibit a nonanucleotide similar to those found in members of the families *Alphasatellitidae* and *Nanoviridae* (5'-TAGTATTAC-3'), likely acquired through recombination (11) (12).

Brazil is recognized as a center of begomovirus diversity, and the genetic diversity and variability of these viruses have been extensively studied over the years in both cultivated and non-cultivated hosts (5, 13-20). These studies indicated that non-cultivated plant species of the genera *Sida* and *Macroptilium* act as "mixing vessels", showing high permissibility to mixed

infections with a diverse array of species that often recombine, conditions that are thought to favour spillover events to cultivated plants (2). Other species, such as *Cleome affinis* and *Euphorbia heterophylla*, act as "sealed containers", harboring almost exclusively a single begomovirus which, in turn, is rarely if ever found in other hosts.

Blainvillea rhomboidea exhibits a sealed container relationship with Blainvillea yellow spot virus (*Begomovirus blainvilleae*, BlYSV), currently the only begomovirus known to infect this host (18). A study that assessed the genetic variability and population structure of tomato-infecting begomoviruses in Brazil also included a small number of BlYSV isolates, and the analyses indicated that this virus has a significantly higher degree of genetic variability compared to the tomato-infecting begomoviruses (13). Interestingly, a more recent study analyzing the complete genome (DNA-A and DNA-B) of a larger number of isolates from distant geographical regions found that, unlike all other begomoviruses analyzed to date and despite their high genetic variability, BlYSV isolates shown no evidence of population subdivision or geographical segregation (15). A more detailed study of BlYSV may provide useful insights related to the evolution of viral populations in sealed container hosts.

The objective of this study was to characterize the genetic variability and evolutionary mechanisms acting upon BlYSV populations.

METHODS

Sample collection

Samples of *Blainvillea rhomboidea* and *Phyllanthus niruri* plants exhibiting symptoms of viral infection were collected from the municipalities of Viçosa, Coimbra, Muriaé (state of Minas Gerais), Touros (state of Rio Grande do Norte), Rio Largo and Maceió (state of Alagoas) between 2022 and 2023 (**Suppl. Table S1**).

Cloning and sequencing of begomovirus genomes

Total DNA was extracted from leaf samples as described by Doyle and Doyle (21). The viral genome was amplified by rolling circle amplification (RCA) (22) and was digested with *ApaI*, *HindIII*, *PstI* and *XbaI* chosen with the assistance of A Plasmid Editor (ApE) software (https://jorgensen.biology.utah.edu/wayned/ape/) to selectively cleave the DNA-A component of BlYSV at a single site. Fragments of approximately 2,600 nt were cloned into the pBluescript KSII+ vector, which had been digested with the corresponding enzyme and dephosphorylated. Recombinant plasmids were transformed into *Escherichia coli* DH5α by electroporation, and completely Sanger-sequenced at Macrogen (Seoul, Republic of Korea). Genome assembly was

performed using SeqAssem (https://science.do-mix.de/software_seqassem.php), and annotation was conducted using ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/). Species identification was preliminarily achieved via BLASTn analysis, and confirmed by pairwise comparisons using SDT v. 1.3 (23) against DNA-A sequences of BlYSV retrieved from GenBank (**Suppl. Table S2**). A threshold of 91% of sequence nucleotide identity was used to delineate individuals of the same species, and 94% identity was used for the same strain, as per the guidelines of the *Geminiviridae* and *Tolecusatellitidae* Study Group of the ICTV (6). An identity threshold of 97% was arbitrarily applied for the demarcation of BlYSV variants. The Mfold web server (24) was used to computationally predict the secondary structure in the nonanucleotide motif with lowest ΔG value.

Recombination and phylogenetic analysis

A multiple sequence alignment was obtained with the MAFFT algorithm (25) on an online server, and recombinant events were examined using RDP5 (26). Each method implemented in RDP5 was configured with default parameters, and significance was determined with a Bonferroni-corrected *p*-value of 0.05. Only recombination events identified by a minimum of four methods were considered reliable. The General Time Reversible nucleotide substitution model with gamma distribution and invariant sites (GTR+G+I) was selected using MrModelTest (27). This model was then used to construct a Bayesian phylogenetic tree using the MrBayes program (28) available on the CIPRES server (29). The tree was generated through two independent analyses of 10,000,000 generations sampled at every 1,000 generations, and was visualized and edited using Figtree (https://tree.bio.ed.ac.uk/software/figtree/). Bean golden mosaic virus served as an outgroup in a neighbor-joining analysis with 1,000 bootstrap replications using MEGA X (30) to determine the root of the Bayesian phylogenetic tree. In addition to the phylogenetic analysis of the complete DNA-A segment, independent Bayesian phylogenetic analyses were performed for the *Rep* and *CP* genes to assess the congruence between the two ORFs. The presence of a temporal signal within the population was evaluated with the TempEst program (31), by correlating the root-to-tip genetic distance with the year of collection of the isolates.

Population structure and genetic diversity

Multivariate statistical analysis was conducted using Discriminant Analysis of Principal Components (DAPC) with the adegenet package (32) implemented in R software. Subpopulations were defined using the *k*-means clustering approach, with the number of

clusters (*k*) ranging from 1 to 10 to maximize the variation between groups. The optimal number of subpopulations was determined using the Bayesian Information Criterion (BIC). The optimal principal component values were investigated to assess the proportion of successful reassignment, a-score and cross-validation. To support the results, the Nst statistic, analogous to Wright's Fixation index at the nucleotide sequence level, was estimated between subpopulations and within each one using DnaSP v. 6 (33). The main descriptors of genetic variabiltiy (haplotype number, haplotype diversity, nucleotide diversity, mutation rate, total number of polymorphic sites, total number of mutations, and average number of nucleotide differences) were evaulated using DnaSP. The statistical significance of the differences in π values between subpopulations was assessed as described by Lima et al. (34).

Selection pressures

Evolution mechanisms in individual ORFs at the population and subpopulation levels were investigated using Tajima's D (35), Fu and Li's D* and F* tests (36), conducted in DnaSP. These tests were performed under the assumption of a null hypothesis of neutrality. Only *p*-values >0.05 were considered sufficient to accept the null hypothesis. Selection pressures were assessed based on the dN/dS ratio. Sites subject to positive and negative selection pressures were assessed using the SLAC, FEL and FUBAR tests available on the Datamonkey server (37).

RESULTS

A total of 119 *B. rhomboidea* and two *P. ninuri* samples were collected, and 75 (73 *B. rhomboidea* and two *P. ninuri*) were preliminarily positive for the presence of a begomovirus based on the detection of an ~2,600-nt band after digestion of the RCA products with restriction enzymes (**Suppl. Table S1**). From these samples, 23 DNA-A clones (19 haplotypes) and 6 DNA-B clones (6 haplotypes) were obtained (**Suppl. Table S1**). BlYSV was the only species identified in this study, and the ApaI enzyme was the most efficient for cloning the DNA-A. This is the first report of this virus in Rio Grande do Norte state, and also the first time that BlYSV is found infecting *P. niruri*.

Pairwise nucleotide sequence comparisons among the 23 DNA-A clones plus 30 additional DNA-A from GenBank revealed >91% sequence identity amongst all the isolates (**Suppl. Figure S1**). Two strains, named WS (widespread) and NE (northeast), were identified in this study. They are subdivided into four variants forming a monophyletic clade in the DNA-A phylogenetic tree (**Figure 1**) with >97% identity among them, named A to D (**Suppl. Table

S3). The prevalent WS strain includes variants B, C and D, plus a group of isolates (most from AL but also including two isolates from Coimbra, MG and the single isolate from Touros, RN) that do not form a monophyletic group and therefore were not classified as a variant. The NE strain consists of isolates from variant A found in the state of Alagoas.

Variant A (*n* = 6) is the most genetically distant variant (**Figure 1**) and consists of isolates collected in three different municipalities of Alagoas state (Maceió, Junqueiro, and Limoeiro). Interestingly, despite their geographic proximity, no isolates from the municipality of Rio Largo were classified in this variant. Furthermore, all variant A isolates obtained in Maceió exhibited a nucleotide substitution (A2667G) within the conserved nonanucleotide motif (from 5'-TAATATTAC-3' to 5'-TAATGTTAC-3'). Sequencing of the PCR fragments from samples of isolates BR:Mac1:23 and BR:Mac2:23 confirmed the mutation. Although this region is highly conserved in the *Geminiviridae* family, two other nonanucleotide motifs have been reported. Viruses in the genera *Becurtovirus* and *Eragrovirus*, as well as one mastrevirus, have the sequence 5'-TAAGATTCC-3' (3), and the begomoviruses Triumfetta yellow mosaic virus (*B. triumfettae*, TrYMV) and Malvaviscus yellow mosaic virus (MvYMV) carry a nanovirus-like sequence (5'-TAGTATTAC-3') (11, 12). Nevertheless, no alteration in the stem-loop structure was observed *in silico* ($\Delta G = -16.26$) (**Suppl. Figure S2**). It is not known whether the observed mutation is also present in the cognate DNA-B components.

Variant B (*n* = 9) consists of isolates obtained in Viçosa and Coimbra (MG) collected in 2022 and 2023. All of them share a recombination event with breakpoints in the *Rep* gene and the common region (**Suppl. Table S4**). This event is supported by the incongruence of clusters of these isolates in the *Rep* and *CP* phylogenies (**Suppl. Figures S3**, **S4**). Isolates BR:Coi157:14 (var. C) and BR:Mac1:23 (var. A) were identified as the major and minor parents, respectively, indicating an origin through "intervariant" recombination. Within this variant, a second shared recombination event was identified in BR:Vic1:22, BR:Vic2:22, BR:Vic4.1:22, and BR:Vic7:22 with breakpoints in the common region, extending until the *Trap* gene.

Variant C (*n* = 19) comprises isolates sourced from Viçosa and Coimbra, spanning the years 2007 to 2022. Among the isolates sequenced in this study, the isolate BR:Vic9:22 from Viçosa was the only one classified in this variant.

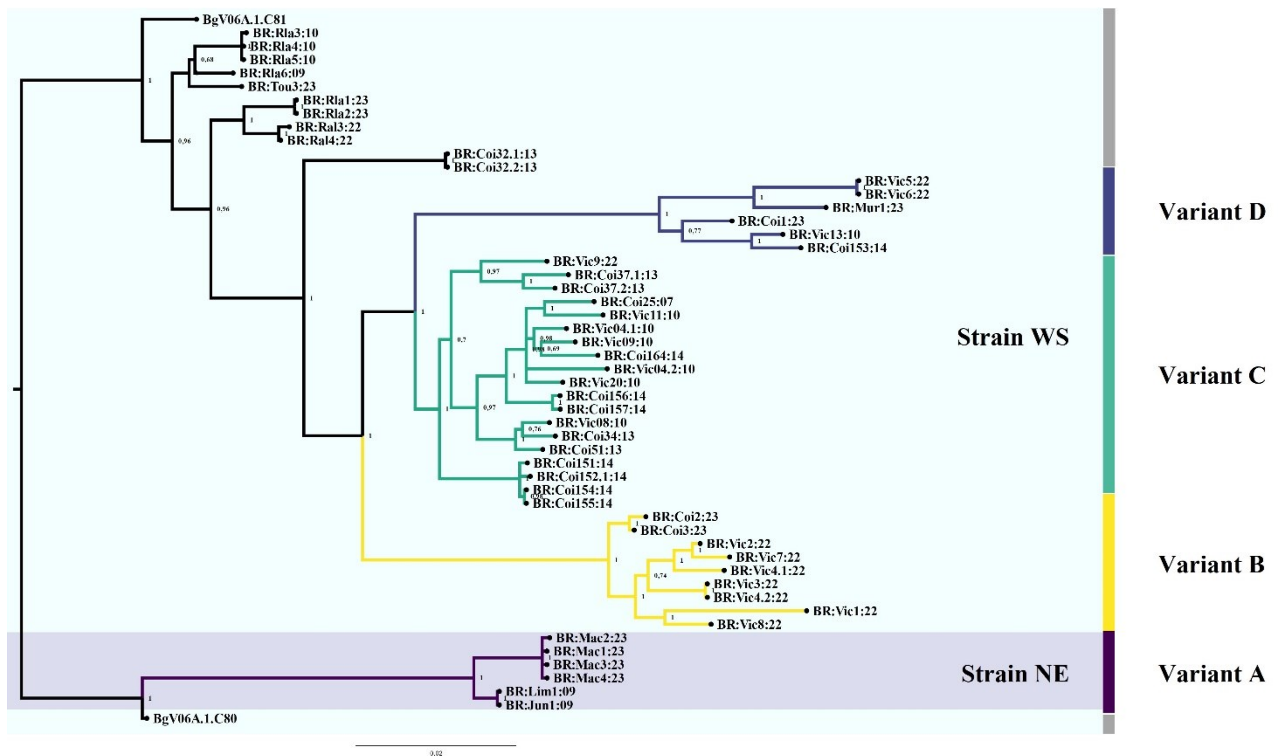


Figure 1. Bayesian phylogenetic tree based on the complete DNA-A sequences of Blainvillea yellow spot virus (*Begomovirus blainvilleae*; BlYSV) isolates collected over a 16-year period. Posterior probability values are indicated at the branches. Branch colors correspond to variant classification based on pairwise sequence comparisons. Isolate names include an abbreviation of the location where isolates were collected: Coi = Coimbra, Vic = Viçosa, Mac = Maceió, Jun = Junqueiro, Lim = Limoeiro, Rla = Rio Largo, Tou = Touros, Mur = Muriaé.

Variant D (n = 6) consists of isolates from the three municipalities in Minas Gerais (Viçosa, Coimbra, and Muriaé). Besides sharing a distinct recombination event, similar to isolates from variant B, all isolates in this variant also share a second recombination event in a similar region (**Suppl. Table S4**). This event was previously described (15) and is also supported by incongruences between the *Rep* and *CP* trees (**Suppl. Figures S3, S4**). In the *Rep* phylogeny, these isolates form a distinct branch, similar to what is observed in the full DNA-A phylogeny, whereas in the *CP* phylogeny these isolates cluster together with isolates from variants B and C.

Analysis of the DNA-A phylogenetic tree (**Figure 1**) reveals the formation of two major clades corresponding to the two strains. Clade I include the vast majority of the isolates (including all isolates from Minas Gerais) and corresponds to the WS strain. Isolates from Minas Gerais are categorized into three monophyletic subclades, corresponding to variants B,

C, and D. These variants' isolates cluster with others of the same variant, indicating a common origin. Notably, a large number of isolates deviate from this pattern, failing to form monophyletic clades and instead displaying multiple basal lineages. This suggests an ancestral nature for these isolates relative to variants B, C, and D. This hypothesis is further corroborated by the shorter genetic distance observed between the basal group of isolates and the root.

Clade II includes six isolates from Alagoas state classified as variant A, and corresponds to the NE strain. Isolates collected in Maceió in 2023 that share a common nonanucleotide mutation cluster together, forming a sister group with two other isolates. This suggests a shared origin for the mutation in their nonanucleotide. However, due to the limited sampling of this strain, more detailed inferences about the emergence of this mutation are not possible.

The distribution of variants collected over the years in Viçosa and Coimbra was analyzed together in two distinct time periods, due to the short distance between the two municipalities (less than 20 km) and the absence of any significant geographical barriers between these populations (**Figure 2**). Between 2010 and 2014, variant C was predominant in both municipalities, representing 81.8% of the sampled isolates, while variant D accounted for 9.1% of the isolates. However, a change in the distribution of the variants was observed in the same region between 2022 and 2023. During this period, variant B became predominant, accounting for 64.3% of the sampled isolates. Variant D was the second most common (28.6%) and variant C, although still present, decreased to 7.1% of the isolates. Variants' occurrence over the years in other municipalities could not be analyzed due to the limited number of sequences available.

To test the hypothesis of temporal structuring among the isolates, a correlation analysis was performed between the collection date and the root-to-tip distance of the phylogenetic tree. The analysis yielded a correlation coefficient of 0.44 (R^2 = 0.19), indicating a moderate correlation (**Table 1**). Besides the relatively low coefficient of determination, these values cannot be used to test for statistical significance because the data do not exhibit an independent distribution. Instead, they are partially correlated due to shared ancestry (31).

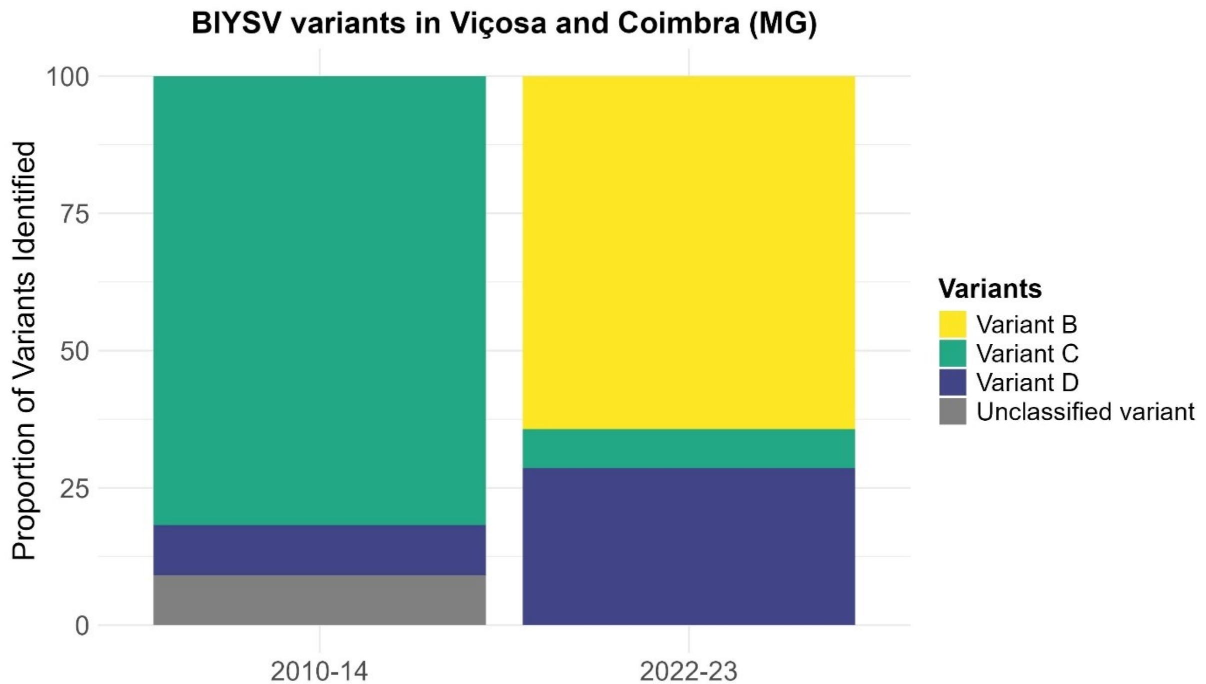


Figure 2. Occurrence of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) variants in two time periods in the municipalities of Viçosa and Coimbra, MG.

Table 1. Correlation analysis between the root-to-tip distance in the DNA-A phylogenetic tree and the year of sample collection of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) isolates, conducted using TemPest.

Dataset	n	Date range	Correlation coefficient	R^2	Slope (rate)	TMRCA*
Full dataset	53	16	0.44	0.19	2.731×10^{-3}	1994
Without variant A	47	16	0.55	0.30	2.388×10^{-3}	1995

*TMRCA = time to most recent common ancestor

Correlation analysis was then conducted excluding variant A isolates, which were identified as outliers based on the graph. These isolates could be evolving at a different evolutionary rate, but this could be masked due to undersampling. Excluding variant A isolates increased the correlation coefficient to 0.55 ($R^2 = 0.30$) (**Table 1**), indicating that the population, at least for the WS strain, is partially structured over time. The X-axis intercept indicates the time for the most recent common ancestor (TMRCA) to be around 1995, shortly before the virus was first reported (38).

Based on six principal components, DAPC confirmed the division of the BlYSV population into five subpopulations (k = 5), consistent with the variant classification (**Figure 3**). The isolates that were not classified into a variant cluster within a single subpopulation (pop3) with high membership probabilities. Reassignment percentages indicate that DAPC is effective in separating the groups, with all individuals correctly assigned to their original groups. The high differentiation between subpopulations is supported by the genetic variability observed within each group relative to the total variability, indicated by the Nst index ranging from 0.51 to 0.85 (**Table 2**).

Overall, the BlYSV population exhibited a nucleotide diversity (π) value of 0.049, although at least half of this genetic variability can be attributed to differences among variants (**Table 3**). The difference of nucleotide diversity between variant B and C isolates is not statistically significant (**Suppl. Figure S5**), while variant D and subpopulation 3 exhibited the highest nucleotide diversity (π = 0.025). Intriguingly, variant A isolates exhibit a much lower nucleotide diversity (π = 0.007) than the other subpopulations.

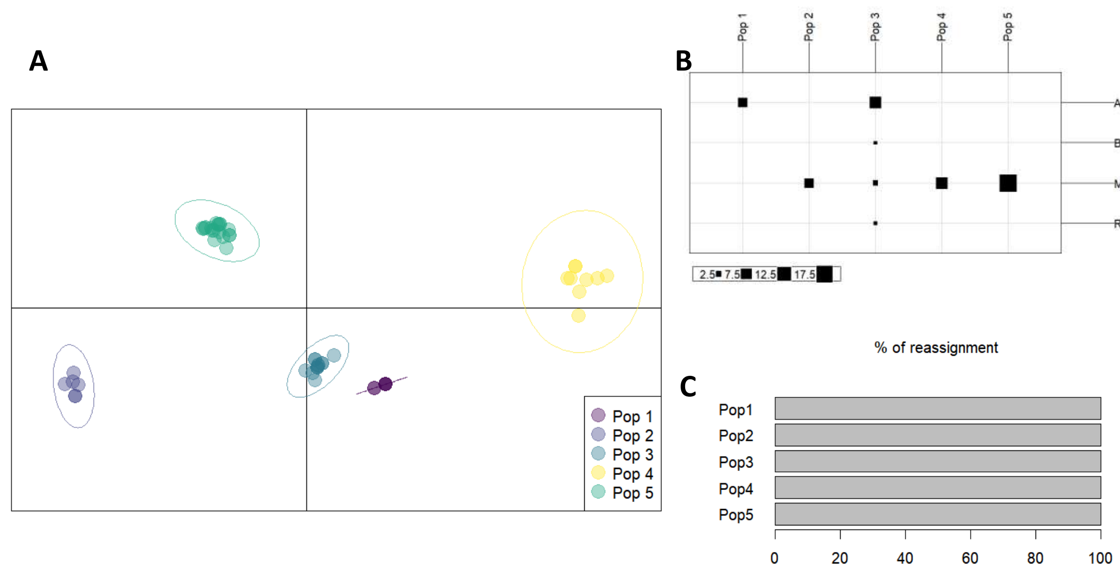


Figure 3. Multivariate statistical clustering analysis of population subdivision using Discriminant Analysis of Principal Components (DAPC) for Blainvillea yellow spot virus (*Begomovirus blainvilleae*). **(A)** DAPC scatter plots indicating population subdivision. pop1 corresponds to variant A, pop2 to variant D, pop4 to variant B, and pop5 to variant C. pop3 isolates were not classified into a variant. **(B)** Comparison between groups inferred by *k*-stat (columns) and isolate origin (rows). **(C)** Percentage of successful reassignment after randomization retaining six principal components.

Table 2. Results of subdivision test (Nst) performed among the five subpopulations* of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) as determined by Discriminant Analysis of Principal Components (DAPC).

Subpopulations		Nst#
pop1	pop2	0.81
pop1	pop3	0.79
pop1	pop4	0.86
pop1	pop5	0.84
pop2	pop3	0.59
pop2	pop4	0.72
pop2	pop5	0.59
pop3	pop4	0.67
pop3	pop5	0.51
pop4	pop5	0.66

*pop1 corresponds to variant A, pop2 to variant D, pop4 to variant B, and pop5 to variant C. pop3 isolates were not classified into a variant

#Values from 0 to 0.05 indicate little genetic differentiation between subpopulations; 0.05 to 0.15, moderate differentiation; 0.15 to 0.25, great differentiation; >0.25 high differentiation

Table 3. Genetic variability indices for Blainvillea yellow spot virus (*Begomovirus blainvilleae*) population and its five subpopulations* as determined by Discriminant Analysis of Principal Components (DAPC).

Subpopulation	n#	π	h	Hd	Θ	s	Eta	k
pop1	6	0.007	5	0.93	0.00574	35	35	17.7
pop2	6	0.025	5	0.93	0.02287	136	139	65.9
pop3	13	0.025	10	0.96	0.02832	223	234	65.8
pop4	9	0.016	8	0.97	0.01913	132	138	43.1
pop5	19	0.019	18	0.99	0.02480	218	231	49.7
TOTAL	53	0.049	46	0.99	0.05806	579	695	129.0

*pop1 corresponds to variant A, pop2 to variant D, pop4 to variant B, and pop5 to variant C. pop3 isolates were not classified into a variant

#n, number of sequences; π, nucleotide diversity; h, haplotype number; Hd, haplotype diversity; Θ, mutation rate; Eta, total number of mutations; s, total number of polymorphic sites; k, average number of nucleotide differences.

Values obtained in Tajima's D, Fu and Li's D* and F* tests did not differ statistically from zero in any protein × subpopulation analysis, indicating absence of strong deviation from neutral equilibrium across the entire BlYSV population (**Suppl. Table S5**). dN/dS values indicated that the AC4 protein of variant B isolates is under positive or diversifying selection. All other analyzed proteins showed negative selection (dN/dS < 1). The CP exhibited the

strongest negative selection pressure (dN/dS ranging from 0.0098 to 0.0847). Consistent with the negative selection pressure observed in the genes, most sites were detected as under negative selection in all three evaluated tests (**Suppl. Table S6**). A few sites under positive selection were observed in *Rep*, *REn*, *TrAP* and *AC4*. Analysis of variant B isolates did not identify sites under positive selection in AC4.

DISCUSSION

Following the initial detection of *Bemisia tabaci* MEAM1 in Brazil in the mid-1990's, numerous reports of begomoviruses causing significant economic losses in cultivated plants, particularly tomatoes, began to surface (39). A high begomovirus diversity infecting this crop was observed, and the indigenous origin of the viruses from non-cultivated plant hosts became rapidly evident. A comprehensive study carried out about ten years after the emergence of the new viruses reported that the begomovirus populations in tomatoes and associated weed hosts evolved at rapid substitution rates, were highly recombinant, and geographically structured (13). However, few studies attempted to understand the temporal evolutionary dynamics of these populations.

BlYSV was first reported in 1999 (38) and so far has only been reported in Brazil, in the states of Minas Gerais, Pernambuco, Rio Grande do Norte, Alagoas, and Bahia. Notably, it is the only begomovirus reported infecting *B. rhomboidea*. Although reports of this virus infecting other hosts are rare (13), we have found it in the non-cultivated plant *Phyllanthus niruri*. Thus, BlYSV can be considered a "sealed container" host, as defined by García-Arenal and Zerbini (2).

Despite its narrow host range, BlYSV exhibits relatively high nucleotide diversity compared to other species infecting cultivated plants such as bean golden mosaic virus (*B. costai*, BGMV), tomato common mosaic virus (*B. solanumvulgarismusivi*, ToCmMV), tomato chlorotic mottle virus (*B. solanumpallidivariati*, ToCMoV), tomato severe rugose virus (*B. solanumseverugosi*, ToSRV) and tomato yellow vein streak virus (*B. solanumflavusvenae*, ToYVSV) (13, 15). Most of this variability can be attributed to its subdivision into five subpopulations, and is comparable to that observed in subpopulations of Macroptilium yellow spot virus (*B. macroptilimaculae*, MacYSV) (15).

The WS strain is widely distributed in the northeast and southeast regions of the country and is subdivided into three variants (B - D) plus a group of unclassified isolates. Meanwhile, the NE strain has only been reported so far in the Northeast region (Alagoas and Pernambuco

states) (40). Phylogenetic analysis suggests that the northeast region is likely the center of origin for this strain, which later migrated to the southeast region.

Variant C isolates predominated in the municipalities of Viçosa and Coimbra during the period spanning 2010 to 2014. However, after recombining with a NE isolate, variant C gave rise to variant B, which quickly replaced it. This hypothesis is supported by the fact that isolates of variant B cluster with isolates of variant C in the CP phylogeny, which does not include recombination regions.

Although negative selection is the most influential evolutionary force acting on begomovirus populations (15, 41), positive selection was found to be acting on the *AC4* gene of variant B and pop3 isolates. The fact that *AC4* is often under positive selection has been reported before and is thought to be correlated with the functions of this protein as a pathogenicity factor (42, 43). Mutations that enhance the fitness of this variant and enable evasion of the plant immune system may be associated with the increased prevalence of this subpopulation in Viçosa and Coimbra. They can also be linked to the *P. niruri* spillover event observed in pop3 (44). However, based on neutrality tests, evolution through genetic drift cannot be ruled out.

BlYSV has a highly recombinant DNA-A. About one third of the isolates evaluated in this dataset showed recombination events (18 out of 53). Similar results were reported for the DNA-B (15). Begomovirus populations are often recombinant and this affects various aspects of their evolution. Despite this mechanism being crucial for increasing viral diversity, it may actually serve mainly to eliminate deleterious mutations. The intergenic region and the *Rep* gene are important recombination hotspots, thus exhibiting relatively higher genetic variability.

Although it does not encode any proteins, the intergenic region in begomoviruses plays an essential role in the viral life cycle (8). This region is highly variable among different species but exhibits high identity among isolates of the same species. It contains a CG-rich region of approximately 30 nt that is common to both segments of bi-segmented begomoviruses and forms a hairpin-shaped secondary structure with a conserved nonanucleotide motif in the loop, serving as a replication origin (45).

Isolates obtained in Maceió in the year 2023 have a mutation in the nonanucleotide motif that does not alter the secondary structure. The same mutation is not found in isolates of the same variant obtained in nearby municipalities in the year 2009. However, the same mutation is present in a phylogenetically related isolate from Pernambuco obtained in 2022 (40). The substitution does not alter the cleavage site of the Rep protein, and its biological effect is

unknown. Nevertheless, it deserves attention because it is in a region with high potential for recombination.

Two hypotheses for the emergence and spread of this mutation are proposed. The first hypothesis suggests that the effect is neutral, and its spread occurs through genetic drift. The low genetic variability observed in this subpopulation supports this hypothesis and suggests genetic bottlenecks that amplify the impacts of genetic drift. However, a second hypothesis suggesting an increase in viral fitness, justifying the spread of this mutation across different states due to natural selection, cannot be ruled out. Therefore, it is important to determine the biological effects of this alternative nonanucleotide, as well as monitor the evolutionary dynamics of this strain in the Northeast region. This can contribute to a better understanding of how viruses from the family *Geminiviridae* restructure their origin of replication.

CONCLUSIONS

- BlYSV is a begomovirus with high genetic variability, and the analyzed population can be divided into two strains (WS and NE) and four variants (A-D)
- Variant C was predominant in Viçosa and Coimbra until 2014, when it was replaced by a new recombinant variant B
- The *AC4* gene of variant B isolates is under positive selection, while all other genes are under strong negative selection
- Isolates obtained in Maceió show a substitution in the nonanucleotide motif that does not alter its secondary structure

REFERENCES

1. Jones RAC. 2021. Global plant virus disease pandemics and epidemics. Plants 10:233.
2. García-Arenal F, Zerbini FM. 2019. Life on the edge: geminiviruses at the interface between crops and wild plant hosts. Annual Review of Virology 6:411-433.
3. Fiallo-Olive E, Lett JM, Martin DP, Roumagnac P, Varsani A, Zerbini FM, Navas-Castillo J. 2021. ICTV Virus Taxonomy Profile: *Geminiviridae* 2021. Journal of General Virology 102:001696.
4. Duffy S, Holmes EC. 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus *Tomato yellow leaf curl virus*. Journal of Virology 82:957-965.
5. Ferro CG, Silva JP, Xavier CAD, Godinho MT, Lima ATM, Mar TB, Lau D, Zerbini FM. 2017. 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 170:204-218.

6. Brown JK, Zerbini FM, Navas-Castillo J, Moriones E, Ramos-Sobrinho R, Silva JC, Fiallo-Olive E, Briddon RW, Hernandez-Zepeda C, Idris A, Malathi VG, Martin DP, Rivera-Bustamante R, Ueda S, Varsani A. 2015. Revision of *Begomovirus* taxonomy based on pairwise sequence comparisons. Archives of Virology 160:1593-619.

7. Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D. 1999. Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. Critical Reviews in Plant Sciences 18:71-106.

8. Fontes EPB, Eagle PA, Sipe PS, Luckow VA, Hanley-Bowdoin L. 1994. Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. Journal of Biological Chemistry 269:8459-8465.

9. Cao M, Lan P, Li F, Abad J, Zhou C, Li R. 2017. Genome characterization of sweet potato symptomless virus 1: a mastrevirus with an unusual nonanucleotide sequence. Archives of Virology 162:2881-2884.

10. Varsani A, Navas-Castillo J, Moriones E, Hernández-Zepeda C, Idris A, Brown JK, Zerbini FM, Martin DP. 2014. Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. Archives of Virology 159:2193-2203.

11. Nascimento LD, Silva SJ, Sobrinho RR, Ferro MM, Oliveira MH, Zerbini FM, Assunção IP, Lima GS. 2016. Complete nucleotide sequence of a new begomovirus infecting a malvaceous weed in Brazil. Archives of Virology 161:1735-1738.

12. Lima ATM, Orílio AF, Almeida MMS, Rocha CS, Barros DR, Castillo-Urquiza GP, Silva FN, Xavier CAD, Bruckner FP, Alfenas-Zerbini P, Barbosa JC, Albuquerque LC, Inoue-Nagata AK, Kitajima EW, Zerbini FM. 2021. Malvaviscus yellow mosaic virus, a divergent begomovirus carrying a nanovirus-like nonanucleotide and a modified stem-loop structure. Annals of Applied Biology 179:96-107.

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

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

15. Xavier CAD, Godinho MT, Mar TB, Ferro CG, Sande OFL, Silva JC, Ramos-Sobrinho R, Nascimento RN, Assuncao I, Lima GSA, Lima ATM, Zerbini FM. 2021. Evolutionary dynamics of bipartite begomoviruses revealed by complete genome analysis. Molecular Ecology 15:3747-3767.

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

17. Pinto VB, Silva JP, Fiallo-Olivé E, Navas-Castillo J, Zerbini FM. 2016. Novel begomoviruses recovered from *Pavonia* sp. in Brazil. Archives of Virology 161:735-739.

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

19. Silva SJC, Castillo-Urquiza GP, Hora-Júnior BT, Assunção IP, Lima GSA, Pio-Ribeiro G, Mizubuti ESG, Zerbini FM. 2011. High genetic variability and recombination in a begomovirus population infecting the ubiquitous weed *Cleome affinis* in northeastern Brazil. Archives of Virology 156:2205-2213.

20. Mar TB, Xavier CAD, Lima ATM, Nogueira AM, Silva JCF, Ramos-Sobrinho R, Lau D, Zerbini FM. 2017. Genetic variability and population structure of the New World begomovirus *Euphorbia yellow mosaic virus*. Journal of General Virology 98:1537-1551.

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

22. Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T. 2004. A simple method for cloning the complete begomovirus genome using the bacteriophage phi29 DNA polymerase. Journal of Virological Methods 116:209-211.

23. Muhire BM, Varsani A, Martin DP. 2014. SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLOS One 9:e108277.

24. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31:3406-3415.

25. Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30:772-780.

26. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, Kelz Z, Kumar V, Murrell B. 2021. RDP5: A computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. Virus Evolution 7:veaa087.

27. Nylander JAA. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

28. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539-542.

29. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE):1-8.

30. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549.

31. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution 2:vew007.

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

33. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution 34:3299-3302.

34. Lima ATM, Silva JCF, Silva FN, Castillo-Urquiza GP, Silva FF, Seah YM, Mizubuti ESG, Duffy S, Zerbini FM. 2017. The diversification of begomovirus populations is predominantly driven by mutational dynamics. Virus Evolution 3:vex005.

35. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

36. Fu Y, Li WH. 1993. Statistical tests of neutrality of mutations. Genetics 133:693-709.

37. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. Molecular Biology and Evolution 35:773-777.

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

39. Ribeiro SG, Ávila AC, Bezerra IC, Fernandes JJ, Faria JC, Lima MF, Gilbertson RL, Zambolim EM, Zerbini FM. 1998. Widespread occurrence of tomato geminiviruses in Brazil, associated with the new biotype of the whitefly vector. Plant Disease 82:830.

40. Gonçalves MHO, Dias CHM, Zerbini FM, Blawid R. Expanding reports: detection and phylogeny of a blainvillea yellow spot virus isolate obtained in the state of Pernambuco. Annals of the 53th Brazilian Congress of Phytopathology, p. 682. Brasília, DF.

41. Mendes ALSF, Melo AM, Ramos-Sobrinho R, Silva SJC, Ferro CG, Ferro MMM, Murilo Zerbini F, Lima GSA, Assunção IP. 2021. High molecular diversity and divergent subpopulations of the begomovirus cnidoscolus mosaic leaf deformation virus associated with Cnidoscolus urens. Archives of Virology 166:3289-3299.

42. Deom CM, Brewer MT, Severns PM. 2021. Positive selection and intrinsic disorder are associated with multifunctional C4(AC4) proteins and geminivirus diversification. Scientific Reports 11:11150.

43. Medina-Puche L, Orilio AF, Zerbini FM, Lozano-Duran R. 2021. Small but mighty: Functional landscape of the versatile geminivirus-encoded C4 protein. PLoS Pathogens 17:e1009915.

44. Menezes LM. 2022. Primeiro relato do begomovírus Blainvillea yellow spot virus em quebra-pedra (*Phyllantus niruri*). Trabalho de Conclusão de Curso (Agronomia). Universidade Federal de Alagoas, Rio Largo, AL.

45. Fontes EPB, Luckow VA, Hanley-Bowdoin L. 1992. A geminivirus replication protein is a sequence-specific DNA binding protein. Plant Cell 4:597-608.

SUPPLEMENTARY MATERIAL

Suppl. Table S1. Blainvillea yellow spot virus (*Begomovirus blainvilleae*) isolates/clones obtained in this work.

Sample code	Host	Location[1]	Date of collection	Isolate/clone name	Component	Enzyme
BA03	*Blainvillea rhomboidea*	Viçosa, MG	May, 2022	BR:Vic1:22	DNA-A	*ApaI*
BA09	*Blainvillea rhomboidea*	Viçosa, MG	May, 2022	BR:Vic2:22	DNA-A	*ApaI*
BA11	*Blainvillea rhomboidea*	Viçosa, MG	May, 2022	BR:Vic3:22	DNA-A	*ApaI*
BA14	*Blainvillea rhomboidea*	Viçosa, MG	May, 2022	BR:Vic4.1:22	DNA-A	*ApaI*
				BR:Vic4.2:22	DNA-A	*ApaI*
BB02	*Blainvillea rhomboidea*	Viçosa, MG	May, 2022	BR:Vic5:22	DNA-A	*ApaI*
					DNA-B	*XbaI*
BB06	*Blainvillea rhomboidea*	Viçosa, MG	May, 2022	BR:Vic6:22	DNA-A	*ApaI*
BB07	*Blainvillea rhomboidea*	Viçosa, MG	May, 2022	BR:Vic7:22	DNA-A	*ApaI*
BB08	*Blainvillea rhomboidea*	Viçosa, MG	May, 2022	BR:Vic8:22	DNA-A	*ApaI*
BB10	*Blainvillea rhomboidea*	Viçosa, MG	May, 2022	BR:Vic9:22	DNA-A	*ApaI*
BC01	*Blainvillea rhomboidea*	Muriaé, MG	January, 2023	BR:Mur1:23	DNA-A	*XbaI*
BD02	*Blainvillea rhomboidea*	Coimbra, MG	May, 2023	BR:Coi1:23	DNA-A	*ApaI*
BD05	*Blainvillea rhomboidea*	Coimbra, MG	May, 2023	BR:Coi2:23	DNA-A	*ApaI*
BD06	*Blainvillea rhomboidea*	Coimbra, MG	May, 2023	BR:Coi3:23	DNA-A	*ApaI*
BI01	*Blainvillea rhomboidea*	Maceió, AL	May, 2023	BR:Mac1:23	DNA-A	*ApaI*
BI02	*Blainvillea rhomboidea*	Maceió, AL	May, 2023	BR:Mac2:23	DNA-A	*ApaI*
BI06	*Blainvillea rhomboidea*	Maceió, AL	May, 2023	BR:Mac3:23	DNA-A	*ApaI*
BI08	*Blainvillea rhomboidea*	Maceió, AL	May, 2023	BR:Mac4:23	DNA-A	*ApaI*
CECA08	*Blainvillea rhomboidea*	Rio Largo, AL	May, 2023	BR:Rla1:23	DNA-A	*PstI*
					DNA-B	*PstI*
CECA10	*Blainvillea rhomboidea*	Rio Largo, AL	May, 2023	BR:Rla2:23	DNA-A	*PstI*
					DNA-B	*PstI*
QP	*Phyllanthus niruri*	Rio Largo, AL	2022	BR:Rla3:22	DNA-A	*ApaI*
				BR:Rla4:22	DNA-A	*HindIII*
TO01	*Blainvillea rhomboidea*	Touros, RN	May, 2023	BR:Tou1:23	DNA-B	*XbaI*
TO02	*Blainvillea rhomboidea*	Touros, RN	May, 2023	BR:Tou2:23	DNA-B	*XbaI*
TO03	*Blainvillea rhomboidea*	Touros, RN	May, 2023	BR:Tou3:23	DNA-A	*ApaI*

TO04	*Blainvillea rhomboidea*	Touros, RN	May, 2023	BR:Tou4:23	DNA-B	*XbaI*

[1]State abbreviations: AL, Alagoas; MG, Minas Gerais; RN, Rio Grande do Norte

Suppl. Table S2. DNA-A sequences of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) available in NCBI database used in this work.

Isolate	**GenBank access number**	**City**	**State**	**Year of collection**	**Host**
BR:Coi164:14	MT626996	Coimbra	Minas Gerais	2014	*Blainvillea rhomboidea*
BR:Coi157:14	MT626995	Coimbra	Minas Gerais	2014	*Blainvillea rhomboidea*
BR:Coi156:14	MT626994	Coimbra	Minas Gerais	2014	*Blainvillea rhomboidea*
BR:Coi155:14	MT626993	Coimbra	Minas Gerais	2014	*Blainvillea rhomboidea*
BR:Coi154:14	MT626992	Coimbra	Minas Gerais	2014	*Blainvillea rhomboidea*
BR:Coi153:14	MT626991	Coimbra	Minas Gerais	2014	*Blainvillea rhomboidea*
BR:Coi152.1:14	MT626990	Coimbra	Minas Gerais	2014	*Blainvillea rhomboidea*
BR:Coi151:14	MT626989	Coimbra	Minas Gerais	2014	*Blainvillea rhomboidea*
BR:Coi51:13	MT626988	Coimbra	Minas Gerais	2013	*Blainvillea rhomboidea*
BR:Coi37.2:13	MT626987	Coimbra	Minas Gerais	2013	*Blainvillea rhomboidea*
BR:Coi37.1:13	MT626986	Coimbra	Minas Gerais	2013	*Blainvillea rhomboidea*
BR:Coi34:13	MT626985	Coimbra	Minas Gerais	2013	*Blainvillea rhomboidea*
BR:Coi32.2:13	MT626984	Coimbra	Minas Gerais	2013	*Blainvillea rhomboidea*
BR:Coi32.1:13	MT626983	Coimbra	Minas Gerais	2013	*Blainvillea rhomboidea*
BR:Vic20:10	KC706522	Viçosa	Minas Gerais	2010	*Physalis* sp.
BR:Vic13:10	KC706521	Viçosa	Minas Gerais	2010	*Blainvillea rhomboidea*
BR:Vic11:10	KC706520	Viçosa	Minas Gerais	2010	*Blainvillea rhomboidea*
BR:Vic09:10	KC706519	Viçosa	Minas Gerais	2010	*Blainvillea rhomboidea*
BR:Vic08:10	KC706518	Viçosa	Minas Gerais	2010	*Blainvillea rhomboidea*
BR:Vic04.2:10	KC706517	Viçosa	Minas Gerais	2010	*Blainvillea rhomboidea*
BR:Vic04.1:10	KC706516	Viçosa	Minas Gerais	2010	*Blainvillea rhomboidea*
BR:Jun1:09	JX871394	Junqueiro	Alagoas	2009	*Blainvillea rhomboidea*
BR:Lim1:09	JX871393	Limoeiro	Alagoas	2009	*Blainvillea rhomboidea*
BR:Rla6:09	JX871392	Rio Largo	Alagoas	2009	*Blainvillea rhomboidea*
BR:Rla5:10	JX871391	Rio Largo	Alagoas	2010	*Blainvillea rhomboidea*
BR:Rla4:10	JX871390	Rio Largo	Alagoas	2010	*Blainvillea rhomboidea*
BR:Rla3:10	JX871389	Rio Largo	Alagoas	2010	*Blainvillea rhomboidea*
BgV06A.1.C81	JF694476	Rio Largo	Alagoas	2009/10	*Blainvillea rhomboidea*

BgV06A.1.C80	JF694468	-	Bahia	2009/10	*Blainvillea rhomboidea*
BR:Coi25:07	EU710756	Coimbra	Minas Gerais	2007	*Blainvillea rhomboidea*

Suppl. Table S3. Classification of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) isolates according to strains, variants and subpopulations.

Isolate	**Strain**[*]	**Variant**[#]	**Subpopulation**[$]
BR:Vic1:22	WS	B	Pop4
BR:Vic2:22	WS	B	Pop4
BR:Vic3:22	WS	B	Pop4
BR:Vic4.1:22	WS	B	Pop4
BR:Vic4.2:22	WS	B	Pop4
BR:Vic5:22	WS	D	Pop2
BR:Vic6:22	WS	D	Pop2
BR:Vic7:22	WS	B	Pop4
BR:Vic8:22	WS	B	Pop4
BR:Vic9:22	WS	C	Pop5
BR:Mur1:23	WS	D	Pop2
BR:Coi1:23	WS	D	Pop2
BR:Coi2:23	WS	B	Pop4
BR:Coi3:23	WS	B	Pop4
BgV06A.1.C80	WS	-[&]	Pop3
BgV06A.1.C81	WS	-	Pop3
BR:Mac1:23	NE	A	Pop1
BR:Mac2:23	NE	A	Pop1
BR:Mac3:23	NE	A	Pop1
BR:Mac4:23	NE	A	Pop1
BR:Coi151:14	WS	C	Pop5
BR:Coi152.1:14	WS	C	Pop5
BR:Coi153:14	WS	D	Pop2
BR:Coi154:14	WS	C	Pop5
BR:Coi155:14	WS	C	Pop5
BR:Coi156:14	WS	C	Pop5
BR:Coi157:14	WS	C	Pop5
BR:Coi164:14	WS	C	Pop5

BR:Coi25:07	WS	C	Pop5
BR:Coi32.1:13	WS	-	Pop3
BR:Coi32.2:13	WS	-	Pop3
BR:Coi34:13	WS	C	Pop5
BR:Coi37.1:13	WS	C	Pop5
BR:Coi37.2:13	WS	C	Pop5
BR:Coi51:13	WS	C	Pop5
BR:Jun1:09	NE	A	Pop1
BR:Lim1:09	NE	A	Pop1
BR:Rla3:10	WS	-	Pop3
BR:Rla4:10	WS	-	Pop3
BR:Rla5:10	WS	-	Pop3
BR:Rla6:09	WS	-	Pop3
BR:Vic04.1:10	WS	C	Pop5
BR:Vic04.2:10	WS	C	Pop5
BR:Vic08:10	WS	C	Pop5
BR:Vic09:10	WS	C	Pop5
BR:Vic11:10	WS	C	Pop5
BR:Vic13:10	WS	D	Pop2
BR:Vic20:10	WS	C	Pop5
BR:Rla1:23	WS	-	Pop3
BR:Rla2:23	WS	-	Pop3
BR:Ral3:22	WS	-	Pop3
BR:Ral4:22	WS	-	Pop3
BR:Tou3:23	WS	-	Pop3

[*]According to the criterion of 94% nucleotide sequence identity for the DNA-A {Brown, 2015 #19477}

[#]According to the criteria of 96% nucleotide sequence identity for the DNA-A and forming a monophyletic cluster in the phylogenetic tree

[$]According to Discriminant Analysis of Principal Components (DAPC)

[&]Not classified into a variant

Suppl. Table S4. Recombination events detected in Blainvillea yellow spot virus (*Begomovirus blainvilleae*) DNA-A.

Event	Recombinant isolate	Recombination breakpoints*		Parents		Method#	*p*-value&
		Begin	End	Major	Minor		
1	BR:Vic1:22, BR:Vic2:22, BR:Vic7:22, BR:Vic4.2:22	50	1401	BR:Vic3:22	Unknow	**R**GBMCST	4.096×10^{-05}
2	BR:Coi153:14, BR:Vic13:10, BR:Coi1:23, BR:Mur1:23, BR:Vic6:22, BR:Vic5:22	55	2026	BR:Vic20:10	Unknow	RG**B**MCST	2.843×10^{-03}
3	BgV06A.1.C80	2650	1593	BR:Coi157:14	Unknow	R**G**BMCST	4.263×10^{-05}
4	BR:Vic3:22, BR:Coi3:23, BR:Coi2:23, BR:Vic8:22, BR:Vic7:22, BR:Vic2:22, BR:Vic4.1:22, BR:Vic4.2:22, BR:Vic1:22	1589	50	BR:Coi157:14	BR:Mac1:23	**R**MCT	3.815×10^{-02}
5	BR:Coi32.2:13, BR:Coi32.1:13	1841	2656	BR:Coi152.1:14	Unknow	MCS**T**	1.457×10^{-02}

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

#R, Rdp; G, Geneconv; B, Boostcan; M, Maxichi; C, Chimaera; S, SisterScan; T, 3Seq

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

Suppl. Table S5. Results of neutrality tests and mean ratios of non-synonymous to synonymous substitutions (dN/dS) for each gene of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) isolates classified into five subpopulations[#] by Discriminant Anaçysis of Principal Components (DAPC).

Subpopulation	Gene	Tajima's D[&]	Fu and Li's D*	Fu and Li's F*	dN/dS
pop 1	*Rep*	0,37522	0,56715	0,56279	0.169
	TrAP	1,18059	1,46717	1,48929	-
	REn	0,33839	0,51052	0,49804	0.132
	AC4	-[$]	-	-	-
	CP	1,13697	1,43971	1,49228	0.0847
pop 2	*Rep*	0,5746	0,60436	0,65624	0.181
	TrAP	0,32163	0,3419	0,36558	0.357
	REn	0,01296	0,08022	0,07133	0.142
	AC4	-0,12471	-0,1377	-0,14477	0.575
	CP	0,28268	0,54081	0,53158	0.00984
pop 3	*Rep*	-0,72617	-0,02947	-0,24858	0.153
	TrAP	-0,92501	-0,4309	-0,17465	0.626
	REn	-0,9475	-0,59439	-0,78858	0.195
	AC4	-0,76576	-0,80311	-0,90526	1.26
	CP	0,20077	0,69196	0,64165	0.0277
pop 4	*Rep*	-1,01471	-0,74684	-0,91082	0.168
	TrAP	-0,28488	0,0081	-0,07014	0.917
	REn	-0,66122	-0,57771	-0,67027	0.248
	AC4	-1,14944	-1,03151	-1,17694	1.24
	CP	-1,01516	-0,98879	-1,11705	0.0325
pop 5	*Rep*	-1,24264	-1,6965	-1,81821	0.162
	TrAP	-0,84373	-0,76933	-0,91901	0.394
	REn	-1,53543	-2,16965	-2,30463	0.320
	AC4	-1,43743	-1,94567	-2,08695	0.894
	CP	-0,90684	-0,99903	-1,13098	0.0318
TOTAL	*Rep*	-0,5724	-0,48826	-0,62346	0.168
	TrAP	-0,5994	0,35722	-0,00338	0.518
	REn	-0,98395	-0,9895	-1,18059	0.223
	AC4	-0,8372	-1,1942	-1,2682	0.874
	CP	-0,79404	-0,02372	-0,37902	0.0279

[*]pop1 corresponds to variant A, pop2 to variant D, pop4 to variant B, and pop5 to variant C. pop3 isolates were not classified into a variant

[&]No values of neutrality were significant to reject the null hypothesis of selective neutrality at $p < 0.05$

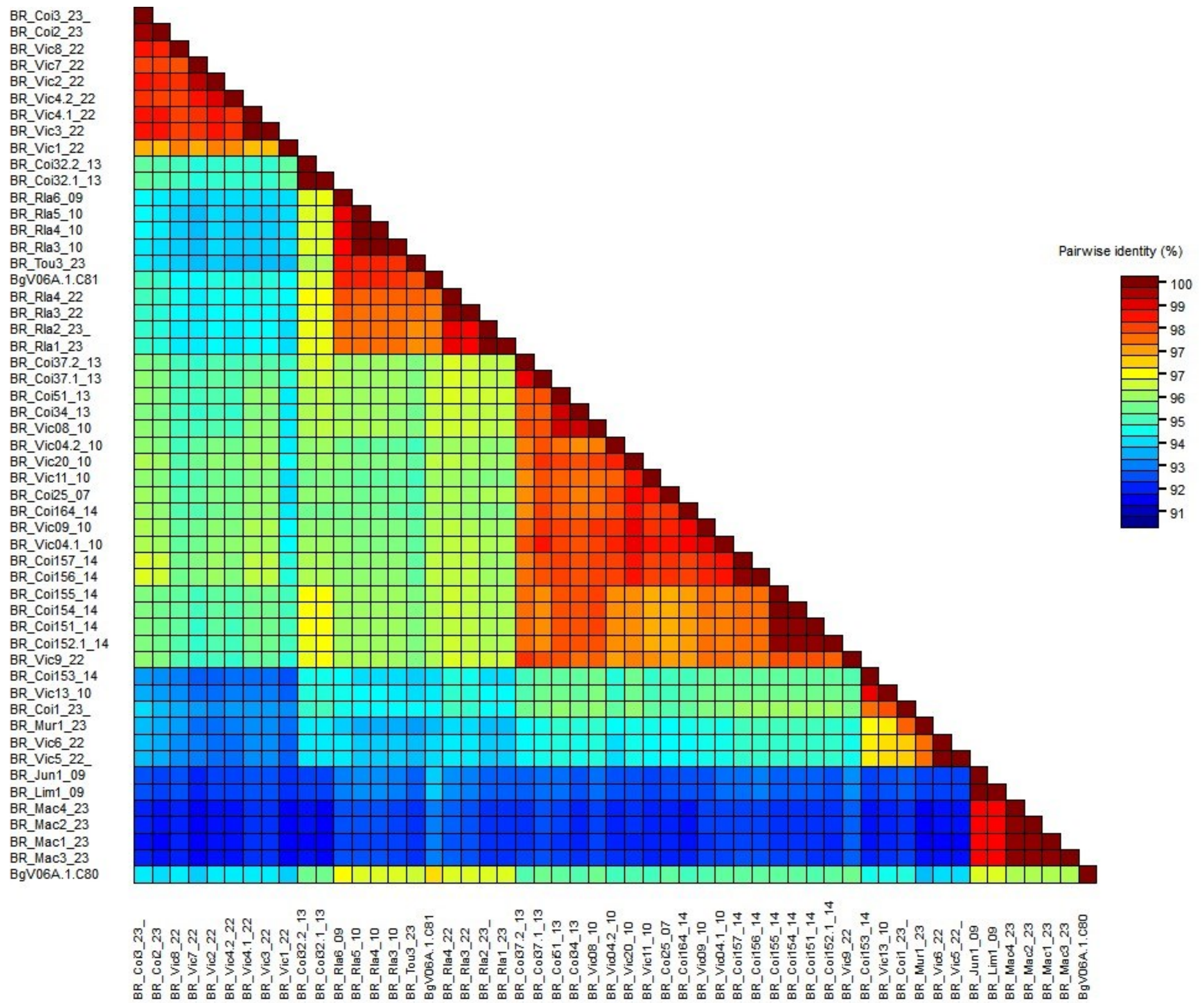
[$]Value cannot be calculated because the sequences are too similar

Suppl. Table S6. Number of amino acid sites under positive and negative selection* for each gene of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) isolates classified into five subpopulations# by Discriminant Anaçysis of Principal Components (DAPC).

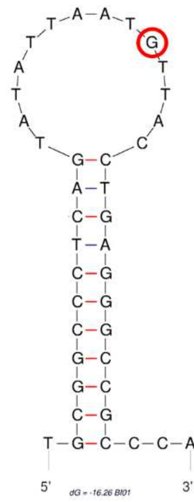
		SLAC		FEL		FUBAR	
Subpopulation	**Gene**	**Positive selection**	**Negative selection**	**Positive selection**	**Negative selection**	**Positive selection**	**Negative selection**
pop1	*Rep*	0	0	0	20	2	19
	TrAP	0	2	0	3	1	4
	REn	0	1	0	5	0	6
	AC4	0	0	0	1	1	2
	CP	0	0	0	23	0	32
pop2	*Rep*	0	4	0	27	0	24
	TrAP	0	2	0	8	2	6
	REn	0	2	0	14	0	6
	AC4	0	0	0	1	2	2
	CP	0	4	0	19	0	22
pop3	*Rep*	0	0	0	5	1	9
	TrAP	0	0	0	0	1	0
	REn	0	0	0	5	0	4
	AC4	0	0	0	0	0	1
	CP	0	1	0	23	0	20
pop4	*Rep*	0	0	0	3	0	4
	TrAP	-	-	-	-	-	-
	REn	0	0	0	0	0	1
	AC4	-	-	-	-	-	-
	CP	0	0	0	4	0	9
pop5	*Rep*	0	7	1	36	3	20
	TrAP	0	2	0	6	2	5
	REn	0	2	0	7	0	4
	AC4	0	1	0	3	1	2
	CP	0	9	0	23	0	21
TOTAL	*Rep*	1	45	2	89	2	75
	TrAP	0	8	3	16	9	12
	REn	0	11	0	24	1	18
	AC4	0	1	1	4	5	2
	CP	0	51	0	92	0	89

*Positive and negative selection at sites with a significance level of 0.1

#pop1 corresponds to variant A, pop2 to variant D, pop4 to variant B, and pop5 to variant C. pop3 isolates were not classified into a variant



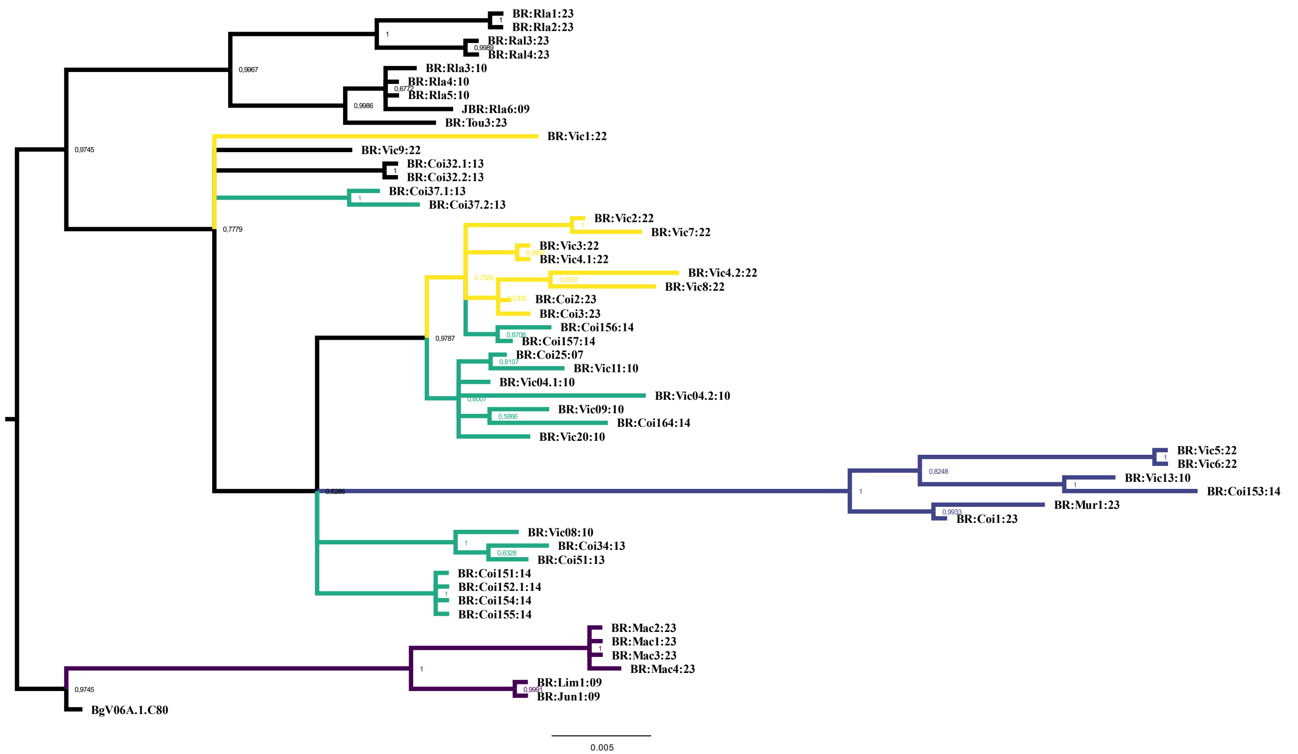
Suppl. Figure S1. Pairwise comparison matrix of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) DNA-A sequences prepared with Sequence Demarcation Tool (SDT) v. 1.3. Isolates with nucleotide identity >96% were classified as belonging to the same variant (indicated by the uppercase letter).



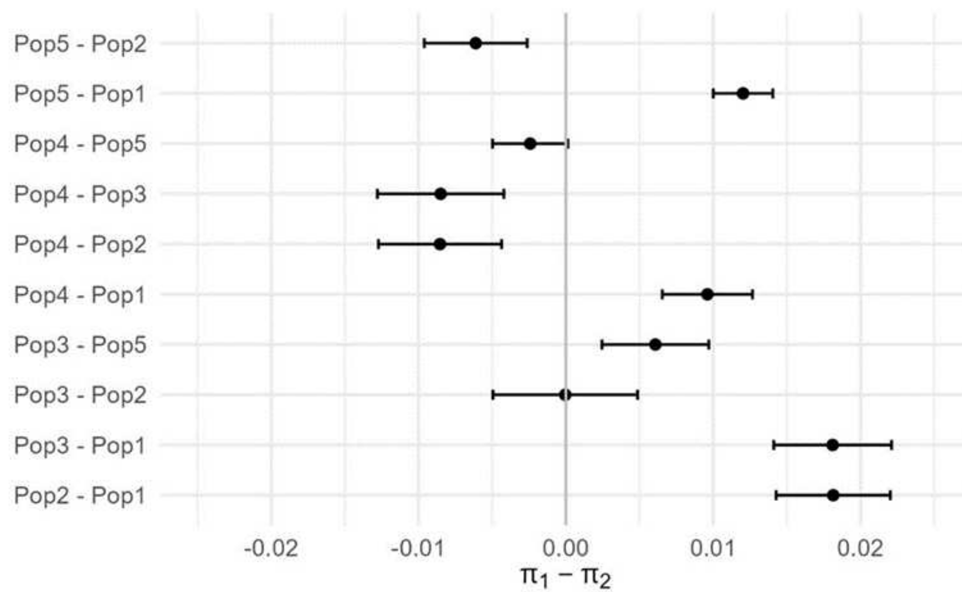
Suppl. Figure S2. Formation of the stem-loop structure of the alternative nonanucleotide motif (TAATGTTAC) of the Blainvillea yellow spot virus (*Begomovirus blainvilleae*) isolate BR:Mac1:23. The structural analysis was performed using the Mfold web server. The red circle highlights the A2667G substitution.

BgV06A.1.C81
BR:Rla3:10
BR:Rla4:10
BR:Rla5:10
BR:Rla6:09
BR:Tou3:23
BR:Rla1:23
BR:Rla2:23
BR:Ral1:22
BR:Ral2:22
BR:Coi32.1:13
BR:Coi32.2:13
BR:Vic5:22
BR:Vic6:22
BR:Mur1:23
BR:Coi1:23
BR:Vic13:10
BR:Coi153:14
BR:Vic9:22
BR:Vic08:10
BR:Coi151:14
BR:Coi152.1:14
BR:Coi154:14
BR:Coi155:2014
BR:Coi34:13
BR:Coi51:13
BR:Coi25:07
BR:Vic11:10
BR:Vic04.1:10
BR:Vic09:10
BR:Coi37.1:13
BR:Coi164:14
BR:Vic04.2:10
BR:Vic20:10
BR:Coi37.2:13
BR:Coi156:14
BR:Coi157:14
BR:Vic2:22
BB07CL06
BR:Vic3:22
BR:Vic4.1:22
BR:Vic8:22
BR:Coi2:23
BR:Coi3:230
BR:Vic1:22
BR:Vic4.2:22
BR:Mac1:23
BR:Mac2:23
BR:Mac3:23
BR:Mac4:23
BR:Lim1:09
BR:Jun1:09
BgV06A.1.C80
0.02

Suppl. Figure S3. Phylogenetic tree based on the *Rep* gene nucleotide sequence of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) isolates. The tree was constructed using Bayesian inference, with posterior probability values indicated at each branch.



Suppl. Figure S4. Phylogenetic tree based on the *CP* gene nucleotide sequence of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) isolates. The tree was constructed using Bayesian inference, with posterior probability values indicated at each branch.



Suppl. Figure S5. Assessment of the statistical significance in the differences of nucleotide diversity (π) values among the five subpopulations of Blainvillea yellow spot virus (*Begomovirus blainvilleae*) as determined by Discriminant Analysis of Principal Components (DAPC). pop1 corresponds to variant A, pop2 to variant D, pop4 to variant B, and pop5 to variant C. pop3 isolates were not classified into a variant. The differences in π where the confidence interval (bars) overlaps the value 0 are not statistically significant.