Universidade de Brasília
Instituto de Ciências Biológicas
Departamento de Fitopatologia
Programa de Pós-Graduação em Fitopatologia

Dissertação de Mestrado

Diversidade do complexo de Sida micrantha mosaic virus e caracterização de novos begomovírus em Malvaceae no Brasil

MARCOS SILVA DE QUEIROZ FERREIRA

Brasília - DF
2023

MARCOS SILVA DE QUEIROZ FERREIRA

Diversidade do complexo de Sida micrantha mosaic virus e caracterização de novos begomovírus em Malvaceae no Brasil

> Thesis presented to the University of Brasília as a partial requirement for obtaining the title of Master in Phytopathology by the Post-Graduate Program in Phytopathology.

Orientadora

Dra. Rita de Cássia Pereira Carvalho

BRASÍLIA, DF– BRASIL
2023

FICHA CATALOGRÁFICA

Diversity of the Sida micrantha mosaic virus complex and characterization of new begomoviruses in Malvaceae in Brazil

Brasília, 2023.

Pages number p.: 68

Master Thesis - Programa de Pós-Graduação em Fitopatologia, Universidade de Brasília, Brasília, DF.

I- *Virus*, *High-Throughput Sequencing*.

II- Universidade de Brasília. PPG/FIT.

**Meu pai Ológunède por guiar sempre meus caminhos e a todos os Òrìsàs
que me dão forças para seguir**

Dedico

Agradecimentos

A Meu pai Ológunède, por guiar meu Ori e me dar forças para persistir em meu odu! Oxalá por cuidar de mim, com sua calma, paciência e tranquilidade.

Aos meus pais, Denise Regina da Silva Ferreira e Marcos de Queiroz Ferreira por todo o apoio em minha trajetória de vida, dedicação e amor.

A minha, irmã Júlia Silva de Queiroz Ferreira por nossa parceria de vida.

Ao meu Bàbálórisà Arno D'Oya, por me acolher em nossa família de Àse e me ensinar todos os dias a ser uma pessoa melhor em meu caminho de vida e ensinar que Òrìsà está sempre na frente de nossas vidas.

A minha orientadora Rita de Cássia Pereira Carvalho por todos os ensinamentos, amizade e parceria durante o mestrado.

Ao professor Leonardo Silva Boiteux por todos os ensinamentos e ajuda na orientação para realização desse trabalho.

Aos colegas de laboratório Luciane, Felipe, Eduardo e Izaías por toda ajuda durante esse processo do mestrado.

A minha irmã de fé Rafaela Couto, por todo companheirismo, trocas, ensinamentos e apoio na minha vida profissional e minha jornada pessoal.

A Maíra Vitoriano, minha amiga e irmã por todos os anos de amizade e parceiria na vida.

A Vêronica Sampaio, minha amiga e minha metade, que me incentivou durante todo o mestrado, e acreditar em mim como potência de ser! Todos os dias me relembra meu potencial e de quem realmente sou, nossa cumplicidade ímpar.

Ao Programa de Pós-graduação em Fitopatologia da Universidade de Brasília (PPG-FIT) por possibilitar meu trabalho.

A Fundação de Apoio à Pesquisa do Distrito Federal (FAP-DF)

Work carried out in the Department of Plant Pathology of the Institute of Biological Sciences of the University of Brasília (UnB), under the guidance of Dr. Rita de Cássia Pereira Carvalho.

Diversity of Sida micrantha mosaic virus isolates and characterization of new begomoviruses in Malvaceae in Brazil

MARCOS SILVA DE QUEIROZ FERREIRA

Thesis approved 2 8 / 0 2 / 2023 by:

_____
Dra. Rita de Cássia Pereira Carvalho
Universidade de Brasília – UnB
(Presidente da Banca)

_____
Prof°. Dr. Cleber Furlanetto
Departamento de Fitopatologia – Universidade de Brasília – UnB
(Membro Titular da Banca)

_____
Dra. Mirtes Freitas Lima
Embrapa Hortaliças
(Membro Titular da Banca)

_____
Prof. Helson Mario Martins do Vale
Departamento de Fitopatologia – Universidade de Brasília – UnB
(Suplente)

Summary

LIST OF FIGURES

CHAPTER 1. LITERATURE REVIEW

Figure 1. Number of viruses by genera infecting Malvaceae family worldwide according based on GenBank (2023), Virus-Host DataBase (2023) and Kitajima et al (2020).

Figure 2. Worldwide distribution of begomoviruses that infect the Malvaceae family based on GenBank (2023), Virus-Host DataBase (2023) and Kitajima et al (2020).

Figure 3. Number of begomoviruses infecting genera of the Malvaceae family in the world according to GenBank database (2022), Virus-Host DataBase (2022) and Kitajima et al (2020).

Figure 4. Photos of weeds of the malvaceae family with symptoms of begomovirus.

CHAPTER 2. Complete bipartite genome characterization of three novel recombinant begomoviruses infecting Malvaceous weeds and their phylogenetic relationships with Neotropical tomato-infecting species.

Figure 1. Genomic organization of the three new bipartite *Begomovirus* species infecting Malvaceous weeds. Panel A: Diagrammatic representation of the circular genomes of new species #1, new species #2 and new species #3 and their respective open reading frames (ORFs). The ORFs AV1, AC1, AC2, AC3, AC4 and AC5 are according to the putative function of their protein products. CR = intergenic region, encompassing the hairpin.

Figure 2. A segment of the intergenic region showing iterons, TATA region, nonanucleotide, stem-loop and at the end Rep = IRD (Rep Iteron-Related Domain).

Figure 3. Pairwise identity in Sequence Demarcation Tool (SDT) analysis carried out using the information of the DNA–A sequences of selected New World *Begomovirus* species showing their phylogenetic identities/distances with three new weed plants infecting Malvaceae species: new species #1, new species #2 and new species #3. These *Begomovirus* species were identified by their accession number and by the acronym of the countries where they were described: AR = Argentina; BR = Brazil; BO = Bolivia. Viral species and GenBank accession numbers are as follow: Tomato leaf curl purple vein virus isolate – ToLCPVV [BR] KY196217; Macroptilium yellow spot virus – MaYSV [BR] KX691398; New species #1; Sidastrum golden leaf spot virus – SidGLSV [BR] HM357458; Pavonia mosaic virus – PavMV [BR] PRJNA485481; Triumfetta yellow mosaic virus – TrYMV [BR] KU131588; Sida Brazil virus – SiBrV [BR]FN436001; Sida golden mosaic Brazil virus – SiGMBRV [BR] FN436001; Leonurus mosaic virus – LeMV [BR] JX863082; Tomato yellow spot virus – ToYSV [AR] MN518733; Sida mottle virus – SiMoV [BR] AY090555; Sida yellow net virus – SiYNV [BR] MF957207; Okra mottle virus – OkMV [BR] EU914817; New species #2; Oxalis yellow vein virus – OxYVV BR] KM887907; Sida chlorotic vein virus – SiCVV

[BR] KX691405; Sida yellow golden mosaic virus – SiYGMV [BR] MF163258; Tomato mild mosaic virus – ToMIMV [BR] EU710752; Sida common mosaic virus – SiCMV [BR] JX871372; Tomato leaf distortion virus – ToLDV [BR] EU710749; Sida mosaic Alagoas virus – SiMAV [BR] JF694471; Sida yellow blotch virus – SiYBV [BR] JX871382; Sida mottle Alagoas virus – SiMoAV [BR] KX896419; Sida bright yellow mosaic virus – SiBYMV [BR] KX348184; Sida micrantha mosaic virus – SiMMV [BR] KC706535; Melochia mosaic virus - MelMV [BR] PRJNA485481; Melochia yellow mosaic virus - MelYMV [BR] PRJNA485481; Cnidoscolus blistering yellow mosaic virus – CnBYMV [BR] MT553995; Tomato interveinal chlorosis virus-2 – ToICV-2 [BR] MK087038; New species #3; Sida yellow mosaic Alagoas virus – SiYMAV [BR] JX871383; Sida mosaic Bolivia virus 1 – SiMBoV1 [BO] HM585441; Tomato rugose mosaic virus – ToRMV [BR] MT214090; Tomato severe rugose virus – ToSRV [BR] DQ207749; Tomato golden mosaic virus – TGMV [BR] PRJNA485481; Soybean blistering mosaic virus – SbBMV [BR] MN508210; Euphorbia yellow mosaic virus – EuYMV [BR] FJ619507; Tomato rugose yellow leaf curl virus – TRYLCV [AR] KC132844; Tomato common mosaic virus – ToCmMV [BR] EU710754; Sida yellow leaf curl virus – SiYLCV [BR] EU710750; Tomato chlorotic leaf curl virus – ToCLCV [BR] MK558058; Tomato bright yellow mottle virus – ToBYMoV [BR] KC791691; Tomato golden leaf spot virus – ToGLSV [BR] PRJNA485481; Tomato golden vein virus – TGVV [BR] JF803254; Tomato yellow vein streak virus – ToYVSV [AR] MN508216; Sida mosaic Bolivia virus 2 – SiMBoV2 [BO] HM585443; Sida golden yellow spot virus – SiGYSV [BR] KX348185; Tomato chlorotic mottle Guyane virus – ToCMoGV [BR] KR263181; Chino del tomate virus – CdTV [BR] AF101476; Tomato bright yellow mosaic virus – ToBYMV [BR] KC791690; Tomato interveinal chlorosis virus – ToICV [BR] JF803252; Tomato chlorotic mottle virus – ToCMOV [BR] AF490004; Tomato mottle leaf curl virus – ToMoLCV [BR] KX896398; Cnidoscolus mosaic leaf deformation virus – CnMLDV [BR] MW600763; and Maize streak virus – MSV [USA] AF32987.

CHAPTER 3. Reappraisal of the diversity of Sida micrantha mosaic virus (SiMMV) and Sida mottle virus (SiMoV) complex: Species status, host range, and geographical distribution

Figure 1. Phylogenetic tree and Sequence Demarcation Tool (SDT) of a set of DNA–A component sequences showing the phylogenetic identities/distances among Sida micrantha mosaic virus (SiMMV) and Sida mottle virus (SiMoV) isolates. These isolates are identified by their accession number GenBank accession numbers of isolates classified/named as SiMMV are following: FN436003, FN436005, KX348155, KX348157, KX348160, KX348161, KX348162, KX348163, KX348164, MT103974, MT103979, MT103980, MT103981, MT103982, MT103983, MT103984, MT103985, MT103986, HM585433, KX348156, KX348158, KX348159, AJ557451, Contig 23, Contig 44, Contig 47, FJ686693, HM585431, HM585437, HM585439, KC706535, KC706536, KC706537, KU852503, KX691401, KX691410, KY650717, KY650722, MT214092, MT733803, JX415187, JX415194, JX415195 and MT733814. These isolates are identified by their accession number GenBank accession numbers of isolates classified/named as SiMoV FN557522 and three isolates of SiMoV are following: AY090555, AJX871377 and JX871378.

LIST OF TABLES

RESUMO GERAL

Begomovirus (família *Geminiviridae*) corresponde ao maior gênero de vírus de plantas, englobando vírus que possuem uma ou duas moléculas de DNA circular de fita simples, encapsidados separadamente em partículas de 18-30 nanômetros. Isolados com genoma bipartido (apresentando os componentes DNA-A e DNA-B) predominam em tomateiro no Novo Mundo. Os begomovírus são transmitidos com grande eficiência pelo vetor *Bemisia tabaci* Middle East Asian Minor 1 – MEAM-1 (= biótipo B), que se encontra amplamente distribuído, apresentando hábito alimentar polífago e alta adaptabilidade a diferentes condições ambientais. A introdução de *B. tabaci* MEAM 1 no início da década de 1990 foi o fator desencadeador de epidemias de begomoviroses em tomateiro no Brasil, com subsequente aumento no número de relatos de novas espécies virais. Além disso, padrões distintos de diversidade e dinâmica de subpopulações virais foram observados em diferentes ambientes. Neste cenário, as plantas daninhas apresentam papel relevante por funcionarem como reservatórios naturais de begomovírus. Novas espécies virais têm sido relatadas em plantas daninhas, incluindo áreas dentro e no entorno de campos comerciais de tomateiro. Alguns begomovírus de plantas daninhas têm sido identificados e caracterizados biologicamente, entretanto acredita-se que estes estudos não tenham sido extensos o suficiente para estimar a real diversidade de novas espécies em plantas daninhas. No país, quatro begomovírus foram relatados infectando concomitantemente o tomateiro e plantas daninhas. A nível global, mais de 40 begomovírus foram descritos em espécies de Malvaceae, sendo que alguns destes foram detectados infectando originalmente espécies de Solanaceae. O genoma pequeno e bipartido somado a eficiência de transmissão e polifagia do vetor propiciam condições favoráveis para a ocorrência de infecções mistas e de eventos de recombinação e pseudo-recombinação entre begomovírus, contribuindo assim, para a frequente alteração da estrutura genética das populações virais nas nossas condições. Diferentes estratégias têm sido empregadas para analisar os processos evolutivos capazes de moldar a estrutura genético-molecular dos begomovírus. A principal delas tem sido a obtenção do genoma viral completo (DNA-A e DNA-B) para posterior análise através do uso de diferentes programas e modelos evolutivos. A metagenômica aliada ao *High Throughput Sequencing* tem permitido determinar uma grande diversidade viral presente no Brasil. Uma grande frequência de infecções mistas vem sendo detectada com muitas delas envolvendo potenciais espécies novas capazes de induzir sintomas severos em plantas. No presente trabalho, quatro *contigs* provenientes de HTS de amostras foliares de Malvaceae foram selecionados para estudo e caracterização, conforme metodologia realizada pela equipe do LVV-Fito e resumida a seguir. Inicialmente, amostras foliares de plantas daninhas exibindo sintomas típicos de begomovírus (clorose generalizada, mosaico e pontuações cloróticas) foram coletadas em áreas de produção de tomate e/ou áreas próximas ao cultivo de tomateiro, nas cinco regiões do país. As amostragens foram realizadas no período de 2001 a 2020, sendo analisado um total de 78 amostras foliares de plantas da família Malvaceae (selecionadas usando como critérios ano/local de coleta). O DNA total foi extraído via CTAB e solventes orgânicos e armazenado a -20 °C. A confirmação inicial da presença de infecção por begomovírus nas amostras foi feita por meio de ensaios de PCR (reação em cadeia da polimerase) usando *primers* degenerados 'PAR1c496' e 'PAL1v1978'. O DNA circular viral foi enriquecido nas amostras positivas via amplificação por círculo rolante (RCA). O sequenciamento de alto rendimento (HTS) foi realizado em uma plataforma *Illumina* NovaSeq *6000*. Os *contigs* virais foram anotados e as leituras foram mapeadas de volta ao genoma anotado usando a ferramenta '*Map to reference*' disponível no programa *Geneious* 11.1.5. Cerca de 7.391.728 milhões de leituras foram obtidas do *pool* de 78 amostras. Após a montagem, no programa CLC Genomics Workbench 11, foram obtidos 10.679 *contigs*. Quatro *contigs* foram selecionados. Três destes *contigs* correspondiam ao DNA-A completo e exibiram níveis de identidade inferiores a 91%, consistente com o critério taxonômico atual de definição de novas espécies dentro do gênero *Begomovirus*. O componente DNA-A da potencial espécie nova #1 apresentou 79% de identidade com Sidastrum golden leaf spot virus (HM357458), a espécie nova #2 apresentou a identidade de 81% com Oxalis yellow vein virus (KM887907), enquanto a nova #3 apresentou a identidade de 78% com Sida yellow mosaic Alagoas virus (JX871383), confirmando a ocorrência de três novas espécies de begomovírus nas plantas daninhas. Os componentes DNA-B foram recuperados e as análises realizadas, incluindo a confirmação de cognatos. O

quarto *contig* apresentou identidade de 98,24% com Sida micranta mosaic virus (SiMMV) (AJ557451). Após uso de primers específicos, já disponíveis, obteve-se um resultado de 41 amostras positivas para SiMMV A caracterização destas três espécies, bem como a ocorrência e distribuição de Sida micranta mosaic virus nas cinco regiões brasileiras, e a reavaliação da diversidade do complexo Sida micrantha mosaic virus (SiMMV) e Sida mottle virus (SiMoV) envolvendo o status da espécie, gama de hospedeiros e distribuição geográfica serão apresentados na presente dissertação.

GENERAL ABSTRACT

Begomovirus (family *Geminiviridae*) corresponds to the largest genus of plant viruses, encompassing viral species that have one or two single-stranded circular DNA molecules, separately encapsidated in 18-30 nanometer particles. Isolates with a bipartite genome (with DNA-A and DNA-B components) predominate in tomato in the New World. Begomoviruses are transmitted with great efficiency by the *Bemisia tabaci* Middle East Asian Minor 1 (MEAM-1 = biotype B) vector, which is widely distributed, with a polyphagous feeding habit and high adaptability to different environmental conditions. The introduction of *B. tabaci* MEAM 1 in the early 1990s was the triggering factor for begomovirus epidemics in tomato in Brazil, with a subsequent increase in the number of new viral species. Furthermore, distinct patterns of diversity and dynamics of viral subpopulations were observed in different environments. In this scenario, weeds play an important role as they function as natural reservoirs of begomovirus. New viral species have been reported in weeds, including areas in and around commercial tomato fields. Some weed begomoviruses have been identified and biologically characterized, however it is believed that these studies have not been extensive enough to estimate the actual diversity of new species in weeds. In the country, four begomoviruses were reported concomitantly infecting tomato and weeds. Globally, more than 40 begomoviruses have been described in Malvaceae species, and some of these have been detected originally infecting Solanaceae species. The small and bipartite genome, together with the transmission efficiency and polyphagy of the vector, provide favorable conditions for the occurrence of mixed infections and recombination and pseudo-recombination events between begomoviruses, thus contributing to the frequent alteration of the genetic structure of viral populations in our conditions. Different strategies have been employed to analyze the evolutionary processes capable of shaping the genetic-molecular structure of begomoviruses. The main strategy is to obtain the complete viral genome (DNA-A and DNA-B) for further analysis using different programs and evolutionary models. The metagenomics combined with High Throughput Sequencing (HTS) has allowed to determine a great viral diversity present in Brazil. A high frequency of mixed infections has been detected with many of them involving potential new species capable of inducing severe symptoms in plants. In the present work, four contigs obtained from HTS of Malvaceae leaf samples were selected for study and characterization, in according to the methodology carried out by the LVV-Fito team and summarized below. Initially foliar samples of weeds showing typical symptoms of begomovirus (generalized chlorosis, mosaics, and golden spots) were collected in tomato production areas and/or areas close to tomato cultivation, in the five regions of the country. Samplings were carried out from 2001 to 2020, with a total of 78 leaf samples from plants of the Malvaceae family analyzed (selected using the year/place of collection as criteria). Total DNA was extracted via CTAB and organic solvents and stored at -20°C. The initial confirmation of the presence of begomovirus infection in the samples was made through PCR (polymerase chain reaction) assays using degenerate primers 'PAR1c496' and 'PAL1v1978'. Viral circular DNA were enriched in positive samples via rolling circle amplification (RCA). High-Throughput Sequencing (HTS) was performed on an *Illumina* NovaSeq 6000 platform. The viral contigs

were annotated and the reads were mapped back to the annotated genome using the 'Map to reference' tool available in the Geneious 11.1.5 program. About 7,391,728 million readings were obtained from the pool of 78 samples. After assembly, using the CLC Genomics Workbench 11 program, 10,679 contigs were obtained. Four contigs were selected. Three of these contigs corresponded to complete DNA-A and exhibited identity levels below 91%, consistent with the current taxonomic criteria for defining new species within the genus Begomovirus. The DNA-A component of potencial new species #1 showed 79% identity with Sidastrum golden leaf spot virus (HM357458), new species #2 showed 81% identity with Oxalis yellow vein virus (KM887907), while new #3 showed 78% identity with Sida yellow mosaic Alagoas virus (JX871383), confirming the occurrence of three new species of begomovirus in weeds. Sequences of DNA-B components were obtained and analyzes performed, including confirmation of cognates species. The fourth contig showed 98,24% identity with Sida micranta mosaic virus (SiMMV) (AJ557451). After using specific primers, already available, a result of 41 positive samples for SiMMV was obtained. The characterization of these three species, as well as the occurrence and distribution of Sida micranta mosaic virus in five Brazilian regions, and the reassessment of the diversity of the Sida micrantha mosaic virus (SiMMV) and Sida mottle virus (SiMoV) complex involving species status, host range and geographic distribution regions will be presented in this dissertation.

HYPOTHESIS

The large number of begomoviruses that have been detected in tomato and other hosts (including Malvaceous weeds) indicates that there is yet a great viral diversity to be discovered in the country. In this scenario, weeds of the Malvaceae family may be contributing to the increase in the genetic diversity of begomoviruses infecting tomatoes in Brazil, since these plants are natural hosts of different begomoviruses and members of the *Bemisia tabaci* complex, have a wide geographic distribution and are commonly present in cultivation fields of this vegetable in Brazil.

MAIN GOAL

1. Prospect and molecularly characterize begomoviruses from weeds associated with tomato cultivation in the country.

SPECIFIC OBJECTIVES

1. To characterize the complete sequence and structural genomic features of three new *Begomovirus* species from weeds of the Malvaceae family associated with tomato crops in Brazil.

2. To prospect the occurrence and geographical distribution of Sida micrantha mosaic virus isolates of weeds of the Malvaceae family in five regions of Brazil.

3. To reappraise of the diversity of Sida micrantha mosaic virus (SiMMV) and Sida mottle virus (SiMoV) complex.

CHAPTER 1

Literature Review

1. Tomato crops and their Malvaceae weeds

The tomato (*Solanum lycopersicum* L.) it is a vegetable crop originated from regions of Peru and Ecuador, developing under tropical and subtropical conditions (Boiteux et al. 2016). The tomato is an important vegetable due its versatility, being used "in natura" or for processing (Vilela et al. 2012). China is the largest tomato-producer, with production estimated in 64 million tons, followed by India, Turkey, and United States. Currently, Brazil occupies the tenth position in world production (FAOSTAT 2022), with an estimated national production of 3.6 million tons (IBGE 2022). In Brazil, the major producer is the state of Goiás (GO) (972 tons) in highland central region of the country (IBGE 2022). The Southeast region also contributed with production of both fresh-market and processing tomatoes, mainly in the states of São Paulo (SP) (876 tons) and Minas Gerais (MG) (513 tons) (IBGE 2022).

Weeds have been found in continuous association with tomato crops. Weeds are aggressive plants that abre able to compete for resources with crops of economic interest, reducing the growth, development, and yield of the associated plants. Weeds are responsible for ≈ 43% of yield losses in the tomato crop (Oerke 2006). Due to their high environmental variability, weeds are widely distributed throughout the world and can adapt to different habitats. Weeds might serve as reservoir hosts for viruses, playing an important role in the persistence and spread of these pathogens (Hallan et al.1988). Several weeds species have already been shown to be natural hosts of begomoviruses, especially plants belonging to families Asteraceae, Caparaceae, Euphorbiaceae, Fabaceae and Malvaceae (Assunção et al. 2006). In fact, the number of weed species hosting begomoviruses is more likely to be underestimated (Barreto et al. 2013). Studies dealing with the relationships of weeds as alternative reservoirs of begomoviruses and the tomato as primary host plant are extremely important to clarify the evolutionary relationship that occurs under field conditions.

The Malvaceae family may play a very important role in generating diversity of tomato-infecting begomovirus since it contains many species that serve as reservoirs of a wide array of viral species. Malvaceae members are distributed widely in tropical and temperate regions. Currently, 22 genera and about 125 Malvaceae species have been reported in Worldwide. Some Malvaceae family encompass a large number of genera such as *Hibiscus*, *Sida*, *Pavonia*, *Abutilon*, *Alcea*, *Malva*, *Lavatera*, *Gossypium*, and *Althaea* (Islam 2019). Virus affecting tomato and species from Malvaceae family will be showed in this chapter.

1.2. Main diseases of viral etiology affecting the tomato crop in Brazil and role of weeds plants

The main viruses infecting tomato crop in Brazil are classified in the genera *Crinivirus*, *Orthotospovirus*, *Potyvirus* and *Begomovirus* (Inoue-Nagata et al. 2016; ICTV 2023). Begomoviruses and

orthotospoviruses are among the most yield-limiting viruses for tomato crop in Brazil (Inoue-Nagata et al. 2016; Jorge et al. 2023).

1.2.1 Genus *Orthotospovirus*

The genus *Orthotospovirus* is classified in the family *Tospoviridae*, which comprises 27 species (Pappu et al. 2009, ICTV 2023). Orthotospoviruses have a single-stranded RNA (ssRNA) with either ambisense or negative orientation, displaying a trisegmented RNA genome composed of three segments: L-RNA (large, 8.9 kb) negative sense, M-RNA (medium, 4.8 kb) ambisense and S-RNA (small, 2.9 kb). It has an ambiguous sense, the ends of the three segments are complementary, the 3'-UCUCGUUA and 5'-AGAGCAAU ends of the RNA are converted and present complementary nucleotides, favoring the non-covalent connection of the ends, forming a ribonucleocapsid (RNP) molecule (van Knippenberg et al. 2005). In Brazil, the predominant species occurring in tomatoes and other Solanaceae are groundnut ringspot orthotospovirus (GRSV), tomato spotted wilt orthotospovirus (TSWV), tomato chlorotic spot orthotospovirus (TCSV) and chrysanthemum stem necrosis orthotospovirus (CSNV), and iris yellow spot orthotospovirus (IYSV) (Pappu et al. 2009; Jorge et al. 2023).

1.3 *Geminiviridae* Family

The *Geminiviridae* is the major family of plant viruses known with 520 species (ICTV 2023) distributed in 14 genera (*Becurtovirus*, *Begomovirus*, *Capulavirus*, *Citlodavirus*, *Curtovirus*, *Eragrovirus*, *Glabovirus*, *Maldovirus*, *Mastrevirus*, *Mulcrilevirus*, *Opunvirus*, *Topilevirus*, *Topocuvirus* and *Turncurtovirus*) that are classified according to the vector, host, genome structure, and genomic organization (Zerbini et al. 2017). Geminiviruses is composed of a broad group of plants viruses causing important diseases worldwide. These viruses can induce losses in agriculture especially in tropical and subtropical regions. *Geminiviridae* viruses are characterized by small circular single-stranded ssDNA encapsidated in twinned or geminate virions (Brown, 1997). The genome of geminiviruses can be either monopartite (with a single genomic DNA–A) or bipartite composed of two (DNA–A and DNA–B) components (Rojas et al. 2005). Each one of the bipartite components are individually encapsidaded into the twinned icosahedral virions (Zerbini et al. 2017). The geminiviruses encode a replication initiator protein (Rep) and conserved intergenic region that contains a common region of the bipartite genome (Brown, 1997). The viral genome replicates via rolling circle amplification (RCA) mechanism that is initiated at stem-loop structure containing nonanucleotide sequence (TAATATT/AC) (Orozco and Hanley-Bowdoin 1996).

The genera classified in this family will be briefly described below.

1.3.1. *Becurtovirus*

Viruses classified in the genus *Becurtovirus* show monopartite genome and are transmitted by leafhoppers. Species of this genus are characterized by the nonanucleotide 5'-TAA**G**AT**C**C-3', exhibiting differences (highlighted in bold) of two nucleotide found in the almost all other geminiviruses (5'-TAA**T**ATT**A**C-3'). Three species are recognized in *Becurtovirus*: *Beet curly top Iran virus*, *Spinach curly top Arizona virus* (Varsani et al. 2014a) and *Exomis microphylla latent virus* (Claverie et al. 2018). Species in this genera have a different molecular organization, containing three ORFs (open reading frames) in viral sense, V1 (encodes for capsid protein), V2 (encodes for movement protein) and V3 (encodes for movement protein) and two ORFs in the complementary sense C1 (codifying for protein associated with replication) and C2 (codifying for transcription protein) (Varsani et al. 2014a; ICTV 2023).

1.3.2. *Capulavirus*

Capulavirus species are characterizated by a monopartite genome with 5'-TAATATTAC-3' nonanucleotide. This genus includes four species: *Alfalfa leaf curl virus* (Roumagnac et al. 2015), *Euphorbia caput-medusae latent virus* (Bernardo et al. 2013), *French bean severe leaf curl virus* and *Plantago lanceolata latent virus* (Susi et al. 2017). Members of three species are transmitted by aphids (Ryckebusch et al. 2020). The members have a distinctive feature in their genomes with a complex arrangement of ORFs encoding movement proteins in the direction of their coat protein gene (CP). Capuloviruses encodes four ORFs in viral sense: V1 encodes CP, V2 (may not be functional), V3 (encodes movement protein) and V4 (encodes movement protein) (Bernardo et al. 2013, 2016). In the complementary sense two overlapping ORFs (C1 and C2) are predicted to encode the Rep protein and C3 a large complementary sense is completely embedded in the ORF C1 (Bernardo et al. 2013; ICTV 2023).

1.3.3. *Citlodavirus*

Members of this genus have been isolated from symptomatic dicotiledenous host plants, both trees and shrubs. All citlodaviruses known are monopartite. In this genus four species are recognized: *Camellia chlorotic dwarf-associated virus* (Zhang et al. 2018), *Citrus chlorotic dwarf associated virus* (Loconsole et al. 2012), *Paper mulberry leaf curl virus 2* (Qiu et al. 2020) and *Passion fruit chlorotic mottle virus* (Fontenele et al. 2018). The genome of *Citlodavirus* contains six ORFs. The complementary strand of the genome potentially encodes geminivirus-like RepA and/or Rep proteins. The Rep and RepA proteins are expressed from an alternatively spliced complementary strand transcript, a putative stem-loop structure which includes the nonanucleotide motif 5'-TAATATTAC-3', highly conserved at the origin of virion strand replication in geminivirus genomes, the genome contains six ORFs (Fontenele et al. 2018; ICTV 2023).

1.3.4. *Curtovirus*

Members of this genus are monopartite virus they infect a wide variety of vegetable crops and are thought to be widely prevalent in non-cultivated dicot species (Varsani et al. 2014b). Curtoviruses are transmitted by leafhoppers in a persistent (circulative, non-propagative) manner (Soto and Gilbertson 2003). Three species were described in this genus: *Beet curly top virus* (Stanley et al. 1986), *Horseradish curly top virus* (Klute et al. 1996) and *Spinach severe curly top virus* (Hernandez and Brown 2010). The three ORFs encoded in the virion-sense. One of them, for the CP; involved in virus movement and insect vector transmission), movement protein, and V2 (involved in the regulation of the relative levels of ssDNA and dsDNA). In the complementary-sense four protein are codified: Rep protein, the pathogenicity factor, the replication enhancer protein, and the C4 protein (which is an important for determination of symptoms), being implicated in cell-cycle control (Hanley-Bowdoin et al. 2013; ICTV 2023).

1.3.5. *Eragrovirus*

This genus includes the single species, *Eragrostis curvula streak virus* (Varsani et al. 2014b). Eragroviruses (Varsani et al. 2014a) have monopartite genomes. All known isolates have been found infecting the monocotiledoneous plants of *Eragrostis curvula* in South Africa (Varsani et al. 2009). The genome organization in viral-sense show V1 encodes for CP) and V2 for MP. In the complementary-sense, the C1 and C2 ORFs is a positional analog of begomovirus, topocuvirus, and short-TrAP/TrAP-like genes (Varsani et al. 2014b). Eragroviruses have the nonacluotide 5'-TAA**G**ATT**C**C-3' sequence rather than the usual 5'-TAATATTAC-3' sequence found in almost all other geminiviruses (Varsani et al. 2009; ICTV 2023).

1.3.6. *Grablovirus*

Grabloviruses have a monopartite genome that includes three species: *Grapevine red blotch virus* GRBaV (Krenz et al. 2012), *Prunus latent virus* (Al Rwahnih et al. 2018) and *Wild Vitis latent virus* (Perry et al. 2018). GRBaV is transmitted by treehoppers (Bahder et al. 2016). Three porteins are encoded in the viral-sense: CP and two proteins with movement proteins functions (Sudarshana et al. 2015). Two ORFs are encoded the in the complementary sense, including the Rep protein.

1.3.7. *Maldovirus*

This genus have been described in association with apples, grapevines and the species *Juncus maritimus* (Liang et al. 2015; Claverie et al. 2018; Al Rwahnih et al. 2018). Three species were characterized thus far *viz*. *Apple geminivirus 1*, *Grapevine geminivirus A* and *Juncus maritimus geminivirus 1*. The maldovirus genomic organization contains an intergenic region that includes the stem-loop structure

including the nonanucleotide 5'-TAATATTAC-3'. The genome contains six open reading frames, putatively encoding the proteins present in monopartite begomoviruses (coat protein and movement protein in the viral sense and Rep, TrAP, Ren, and C4 proteins in the complementary sense).

1.3.8. *Mastrevirus*

Mastrevirus have monopartite genomes mainly infecting monocots, but also some dicots. The mastreviruses are transmitted by leafhoppers (Muhire et al. 2013). Currently, this genus comprises 45 species, with maize streak virus (MSV) being the type species and the most studied species due to the severe damage caused in the African continent (Shepherd et al. 2010). The genomic organization in viral-sense displays ORFs encoding for (CP) and (MP). In the complementary sense, there are two ORFS expressed from C1 and C2 genes by transcript splicing encoding the replication-associated protein (Rep) (ICTV 2023).

1.3.9. *Mulcrilevirus*

This genus is composed of monopartite genomes with origin of replication 5'-TAATATTAC-3'. Thus far, *Mulcrilevirus* genus is composed of two species: *Mulberry crinkle leaf virus* (Lu et al. 2015) and *Paper mulberry leaf curl virus 1* (Qiu et al. 2020). Mulcrileviruses have been described in association with *Morus alba* or *Broussonetia papyrifera* in China. The genome contains six ORFs. In viral sense, the ORF V1 encodes the CP and V2 encodes the putative MP. In the complementary sense, there are two ORFs (C1 and C2) encoding for the Rep protein (Lu et al. 2015; ICTV 2023).

1.3.10. *Opunvirus*

This genus has only a single member – *Opuntia virus* 1 – described infecting Cactaceae plants in the New World. Isolates of Opuntia virus 1 (OpV1) display monopartite genomes, containing the conserved geminivirus nonanucleotide (5'-TAATATTAC-3'). The genome organization of the OpV1 is similar to the genomic organization of monopartite begomoviruses from the Old World with six ORFs with the CP and MP in the viral sense and Rep, TrAP, Ren, and C4 proteins in the complementary sense (Fontenele et al. 2020; ICTV 2023).

1.3.11. *Topilevirus*

Two monopartite species are described in *Topilevirus*: *Tomato apical leaf curl virus* (Vaghi Medina et al. 2018) and *Tomato geminivirus* 1 (Fontenele et al. 2017). The genomic organization is similar to other geminiviruses with nonanucleotide 5'-TAATATTAC-3', with six ORFs. The V1 ORF is predicted to encode CP and, V2 may encode a MP. In the complementary-sense, the ORFs C1 and C2 encode for the

Rep, the V3 and C3 encode for proteins with yet unknown function (Vaghi Medina et al. 2018; ICTV 2023).

1.3.12. *Topocuvirus*

This genus is composed of a single species *Tomato pseudo-curly top virus*. The virus has a monopartite genome that encodes six genes with an organization similar to the genomes of monopartite begomoviruses. This virus is transmitted by a treehopper and it is has been restricted thus far to the southeastern United States (Briddon et al. 1996; ICTV 2023).

1.3.13. *Turncurtovirus*

This genus is composed of species with monopartite genomes. Turncurtovius are represented by *Turnip curly top* (Razavinejad et al. 2013; Varsani et al. 2014b), *Turnip leaf roll virus* (Kamali et al. 2016) and *Sesame curly top virus* (Hasanvand et al. 2018). The genomic organization of the turncurtovirus are similar to other geminivirus. In viral-sense V1 gene encodes a geminivirus-like CP that is most similar to and V2 ORFs. In the complementary-sense, the genome displays four ORFs (C1, C2, C3, and C4). The ORF C1 encodes a Rep homologue, involved in the initiation of rolling-circle replication. All characterized TCTV isolates have the same 5'-TAATATTAC-3' nonanucleotide sequence motif that is found at the virion-strand origins of replication of mastreviruses, begomoviruses, curtaviruses, and topocuviruses (Kamali et al. 2016; ICTV 2023).

1.4 Genera *Begomovirus*

The *Begomovirus* is the largest genus in the *Geminiviridae* family, with $\approx$ 445 species (ICTV, 2023). The first report of begomovirus infecting tomato in Brazil was done in São Paulo State in the late 1950s (Flores et al. 1960). The virus was subsequently characterized and named as tomato golden mosaic virus (TGMV), transmitted by *Bemisia tabaci* (Matyis et al. 1975). Until the beginning of the 1990s, diseases caused by begomovirus species were not often found under field conditions, imposing minor economic damages. Begomoviruses became relevant in tomatoes after the first outbreaks in the Federal District (Ribeiro et al. 1994). These ssDNA viruses have both monopartite and bipartite circular genomes. The bipartite genome has two components DNA-A and DNA-B ($\approx$ 2,600 nucleotides each), enconding six ORFs. Those components have a commom region with $\approx$ 200 nucleotides. The DNA–A displays (in viral sense) the AV1 gene encondig the coat protein (CP). In complementar sense, AC1 encodes for the Rep protein, AC2 encodes for the transcription-activating protein (TrAp), AC3 encodes the protein enhances viral replication (REN), AC4 encodes ac4, a factor related with symptom expression. In recent years, several reports have shown that some bipartite begomoviruses also encode an AC5/C5 protein from the complementary DNA strand. These AC5/C5 proteins play different roles in virus infections. AC5 encodes

a protein associated with pathogenicity and suppression of gene silencing (Li et al. 2015). However, Li et al. (2021) demonstrated that ageratum leaf curl Sichuan virus (ALCScV), a monopartite begomovirus, also encodes a C5 protein that is important for disease symptom expression and can affect viral replication. Wu et al. (2022) reported that the AC5 protein encoded by the bipartite squash leaf curl China virus is an RNA silencing suppressor and a virulence determinant. The ORFs in viral sense of the DNA-B encompass the BV1 (coding for the nuclear sthuttle protein – NSP). In the complementar sense, the ORF BC1 encodes for the MP. The monopartite and bipartite begomoviruses have similar genome organization, but in the viral sense there is an addional ORF (named as AV2) that encodes the protein with function similar to MP present in the DNA-B component of bipartite begomoviruses.

In most of the genera in family *Geminiviridae* there is a highly conserved sequence of nine nucleotides (TAATATTAC nucleotide motif) located in a stem-loop structure. The synthesis of a new DNA strand begins by the Rep protein cutting the strand (in the viral direction) near the 3' end of the AC nucleotide in the sequence. Replication takes with an intermediate dsDNA via the rolling circle mechanism. The synthesis of the complementary DNA strand is dependent upon host cellular machinery. In the early stages of replication, the coding regions (in the viral sense and in the complementary sense) diverge from the intergenic region. Transcription is bidirectional, with independently controlled transcriptions starting in the intergenic region (Gutierrez et al. 2004).

Argüello-Astorga and Ruiz-Medrano (2001) used iterons as an analytical tool of begomovirus diversity, seeking and evaluating the potential determinants of REP protein binding to specific motifs in viral DNA. REP subdomains found in their analysis are variable among viruses with different iterons but are conserved in viruses with identical iterons. These conserved subdomains were named "*Iteron-Related Domains*" (IRD-Rep). These analyzes allowed grouping the begomoviruses with unique iterons in Old World species and New World species.

Years late, Cantú-Iris et al. (2019) carried out a more extensive comparative study while detecting and characterizing a new species *Blechum interveinal chlorosis virus*. In this work, they confirmed the presence of the nonanucleotide **TAATATTAC** as well as three **GGGGGA** iterons in a wide array of begomoviruses. The analyzes were carried out using 130 begomovirus sequences found that the promoter regions of the CP gene had some characteristics, such as the central presence of **ACTT-N7-AAGT** and an association that did not vary with the **TATA-box**, called **TATA-Associated Composite Elements** (TACE).

Both studies (Argüello-Astorga and Ruiz-Medrano 2001 and Cantú-Iris et al. (2019) corroborate with molecular characterization of begomoviruses.

The begomoviruses can be also classified according to their global distribution. The monopartite begomoviruses are predominant in the Old World (Africa, Asia, and Europe), whereas the bipartite

begomoviruses are prevalent in the New World (Americas). However, peculiar monopartite begomoviruses have been also detected the New World (Reis et al. 2020). With the emergence of new begomoviruses, Brown et al. (2015) carried out extensive work to unify the correct taxonomic classification of these plant pathogens. According to these criteria, a new *Begomovirus* species is classified according to levels of identity of the complete DNA-A genome with previously described viruses. A new species must have identity sequence levels below 91% when compared to species sequence already described in the database. Viruses above 94% is considered a new strain of a given species (Brown et al. 2015).

This high number of new emerging species of begomovirus related could be associated to the whitefly (family *Aleyrodidae*) vector. *Bemisia tabaci* Middle East Asia Minor 1 (MEAM1) plays an important role in the transmission of viruses, due to its polyphagous habit, facilitating spread of begomoviruses across distinct hosts. The identification of cryptic species within the *B. tabaci* complex is carried out using the sequence information derived from the *mitochondrial cytochrome oxidase* 1 gene. The threshold of 3.5% divergence has been used as a criterion for demarcating a new species within the *B. tabaci* complex (Dinsdale et al. 2010; De Barro et al. 2011; Marubayashi et al. 2013). In Brazil, the MEAM-1 genetic group (former biotype B) still predominates (Fernandes et al. 2022). Non-propagative circulatory is the type of interaction among the *B. tabaci* vectors and the begomoviruses (Rosen et al 2015; De Barro et al. 2011; Watanabe et al. 2019; Fernandes et al. 2022). Recent studies on the interaction between tomato yellow leaf curl virus (ToYLCV) and *B. tabaci* have demonstrated the ability, under specific conditions, of this virus to use of the body of its vector for replication (Pakkianathan et al. 2015). Studies on virus-vector relationships are demonstrating that the coat protein plays an important role in the begomovirus transmission by members of the *B. tabaci* complex (Pan et al. 2020). Mutations in the begomovirus coat protein can alter the capacity/efficiency of virus transmission by the insect vector (Pan et al. 2020). Fernandes et al. (2022) found in field studies that the *B. tabaci* MEAM1 continues to be prevalent in monocultures such as cotton, soybeans and tomatoes. The continuous monitoring of *B. tabaci* species is crucial because changes in the agricultural landscape, climate, and in the crop management methods can alter the dominance ratio of *B. tabaci* MEAM1 as well as the distribution of these species in crop areas of Brazil (Fernandes et al. 2022).

1.5 Virus related in the family Malvaceae

Species classified in the Malvaceae family are host to several viruses. A survey was carried out using data, including Genbank, Virus-host Database and Kitajima (2020) list. The genera with the highest incidence in the Malvaceae family are ilustrade in the **Figure 1**. A greater number of species was observed for the genus begomovirus (175 species), followed by Potyvirus (14 species) and Polerovirus and Betacarmovirus both with 4 species.
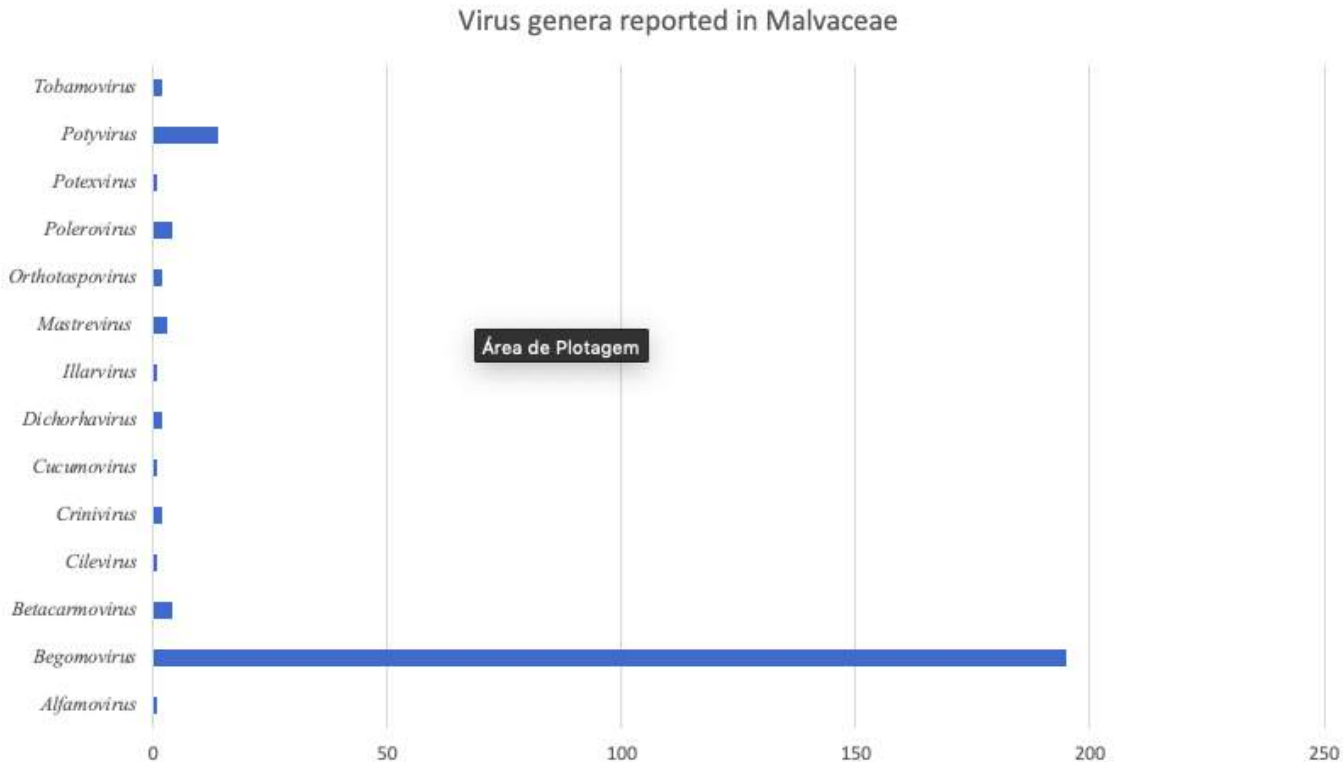
Figure 1. Number of viruses classified by genera infecting Malvaceae family worldwide according GenBank (2023), Virus-Host DataBase (2023) and Kitajima et al (2020).

1.6 Weeds as reservoir of tomato-infecting begomoviruses

Weeds are widely distributed around the world due to their superior environmental adaptability and they might serve as reservoirs of a wide array of pathogens (**Figure 2**), including begomovirus. Survival and spread of begomoviruses in weeds may occur, therefore, in the absence of their hosts (Bedford et al. 1988), making difficult the management via cultural methods. *Ageratum conyzoides* (a member of Asteraceae family) is native to Central America. It is reported as an annual invasive weed in tropical regions of the world and reported as natural host for ageratum yellow vein virus (Swanson et al. 1993). Tahir et al. (2015) reported in biological assays that Ageratum enation virus despite being a weed virus also displayed the ability to infect tomato crops. *Croton* sp. is a member of the Euphorbiaceae family, a weed native to South America but naturalized in the Indian subcontinent (Burgos et al. 2015). *Croton bonplandianum* is a natural host of croton yellow vein virus (Pramesh et al. 2013) and also a virus described in tomato, tomato leaf curl New Delhi virus (Chowda Reddy et al. 2005). In Brazil, four begomoviruses were characterized in weeds associated to tomato crops: Sida mottle virus (SiMoV) (Cotrim et al. 2007), Sida micrantha

mosaic virus (SimMV), Euphorbia yellow mosaic virus (EuYMV) (Barreto et al. 2013; Duarte et al. 2020) and Sida yellow net virus (SiYNV) (Fernandes 2015).

1.7 Malvaceae weeds as reservoir of tomato-infecting begomoviruses

Over decades of vegetative propagation of weeds from the genus *Sida* persistently harbored several begomoviruses (Prajapat et al. 2014). There is a wide variety of weed hosts of begomovirus classified in the Malvaceae family, among them many are classified in the genus *Sida*. In addition, begomoviruses have been propagated through ornamental plants also belonging to the Malvaceae family, for example plants of the genus *Abutilon* are hosts of Abutilon mosaic Brazil virus (AbMBV) (Paprotka et al. 2010). Other virus, such Sida micrantha mosaic virus (SiMMV), Sida mottle virus (SiMoV) and Sida common mosaic virus (SiCMV), have been linked to the mosaic disease in *Sida micrantha*, which was previously associated with a Brazilian strain of AbMV (Castillo-Urquiza et al. 2008; Paprotka et al. 2010). Some examples of weeds plants classified in Malvaceae family, acting as a reservoir of begomovirus for important crops, include tomato (**Figure 3**). Begomoviruses reported originally in weeds of distinct botanic families (e.g. blainvillea yellow spot virus and euphorbia yellow mosaic virus) have been also reported infecting species of the genus *Sida*.

Malvastrum coromandelianum (Malvaceae) is an annual weed native to North America and now distributed in Africa, Asia, and South America. Twenty-five begomoviruses were described in association with *Malvastrum* species, 13 of them were begomoviruses, which were reported infecting plants as tobacco, pepper, sweet-pepper, and tomato (family Solanaceae).

Some of these species are listed bellow. *Malvastrum coromandelianum* an alternative host for monopartite begomoviruses such as Malvastrum yellow mosaic virus, Malvastrum leaf curl virus, Malvastrum leaf curl Guangdong virus, and tomato yellow leaf curl China virus (Guo et al. 2007; Wu et al. 2007; Liu et al. 2009). In the New World, *M. coramandelianum* is a host reservoir of Sida golden yellow vein virus, Sida golden mosaic Florida virus, and Malvastrum yellow mosaic Jamaica virus (Graham et al. 2010; Fiallo-Olivé et al. 2013; Roshan et al. 2019). Huang and Zhou (2006) reported a begomovirus related to tobacco leaf curl Yunnan virus infecting *M. coromandelianum*. Liu et al. (2009) demonstrated that *M. coromandelianum* was serving as an alternative host for tomato yellow leaf curl China virus. In Cuba, a new species named Sida golden mosaic Florida virus (SiGMFV) was found infecting *M. coromandelianum* (Fiallo-Olivé et al. 2010). There are also reports of hollyhock leaf curl virus infecting *M. coromandelianum* in Pakistan (Zia-Ur-Rehman et al. 2017). Li et al. (2018) detected for the first time, tobacco curly shoot virus (TbCSV) infecting *M. coromandelianum* in China, a virus commonly found infecting tobacco and tomato, illustrating the role of weeds as reservoirs of begomovirus.

Four *Malva* spp were described as hosts for 16 begomoviruses. Of these, only two were originally described infecting members of the genus *Malva*. The others were described in species of the families Curcubitaceae and Solanaceae (squash, tobacco, tomato, and pepper). In addition, there are also reports of other begomoviruses within the genus of the family Malvaceae, such *Hibiscus* and *Gossipium* species. Cohen et al. (1988) carried out a survey demonstrating that plants of *M. parviflora* are natural hosts of ToYLCV. Berrie et al. (2001) demonstrated via agroinoculations that *M. parviflora* is a potential host for South African cassava mosaic virus (SACMV). Al-Musa et al. (2008) detected squash leaf curl virus (SLCV) in *M. parviflora* to naturally occur this infection in Jordan. Sattar et al. (2017) had the first report of hollyhock leaf curl virus (HoLCV) infecting *Malva parviflora* in Pakistan. Bananej et al. (2021) recently detected cotton leaf curl Gezira virus (CLCuGeV) and tomato leaf curl betasatellite (ToLCB) infecting *Malva sylvestris* plants in Iran and, emphasized the association of CLCuGeV and ToLCB in *M. sylvestris* illustrates the potential importance of this host for emergence and potential spread of this begomovirus complex to other plant hosts including important crop species. (Moradi and Mehrvar 2023) recently published a study performing a metagenomic analysis of malva vein clearing virus (MVCV) to better understand the relationship of MVCV to worldwide occurrence.

Some species of *Sida* are invasive perennial weeds found in tropical and subtropical regions. These weeds may as host reservoir of others begomoviruses associated with tomato such tomato mild mosaic virus, which were found to infect tomatoes and beans (González-Aguilera et al. 2012). *Sida micrantha* has been reported as reservoir host of Abutilon mosaic virus, Abutilon mosaic Bolivia virus, Sida mosaic Bolivia virus 1, and Sida golden mosaic buckup virus (Wyant et al. 2011; Stewart et al. 2014).

New studies aiming to elucidate the relationships and interchanges of begomoviruses among weeds and crops in Brazil have been carried out by distinct research groups. For example, three begomoviruses originally described in Malvaceae weeds were also detected infecting tomato crops in Brazil: Sida mottle virus (SiMoV) (Cotrim et al. 2007), Sida micrantha mosaic virus (SimMV), and Sida yellow net virus (SiYNV) (Fernandes 2015). In addition, Malvaceous weeds have been reported as alternative hosts of tomato-infecting begomoviruses such as tomato chlorotic mottle virus; tomato mild mosaic virus; tomato mottle leaf curl virus, and tomato yellow spot virus (Tavares et al., 2012; Ferro et al., 2017; see also **Supplementary Table 1, Chapter 2**). Passos et al. (2017) described two new begomoviruses infecting *Sida* species in the Northeast of the country. Macedo et al. (2020) reported a complete nucleotide sequence of a novel bipartite begomovirus infecting a *Sida* species in Piaui State also in the Northeastern region (Hoffmann et al. 2021), which was recently reported infecting commercial cultivars of cotton (*Gossipium hirsutum*). Lima et al. (2021) reported a new begomovirus phylogenetically related to other malvaceous-infecting begomoviruses from Brazil infecting the ornamental plant *Malvaviscus arboreus.*
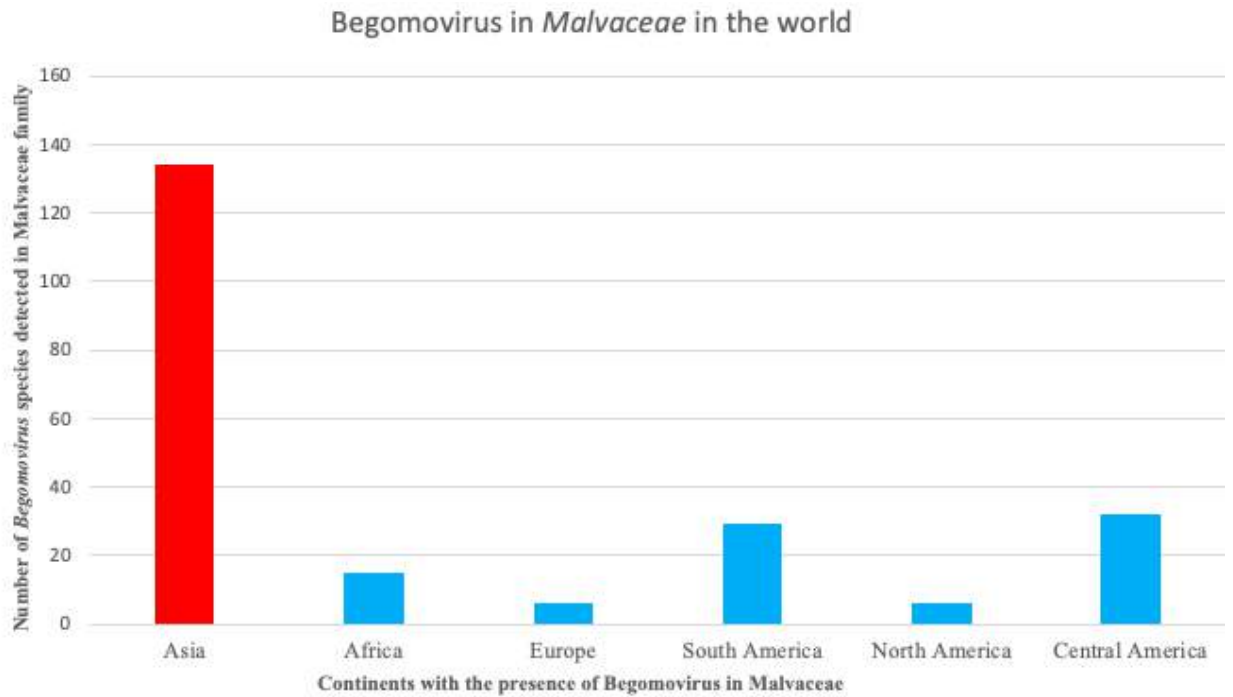
Figure 2. Worldwide distribution of begomoviruses that infect the Malvaceae family based on GenBank (2023), Virus-Host DataBase (2023) and Kitajima et al. (2020).
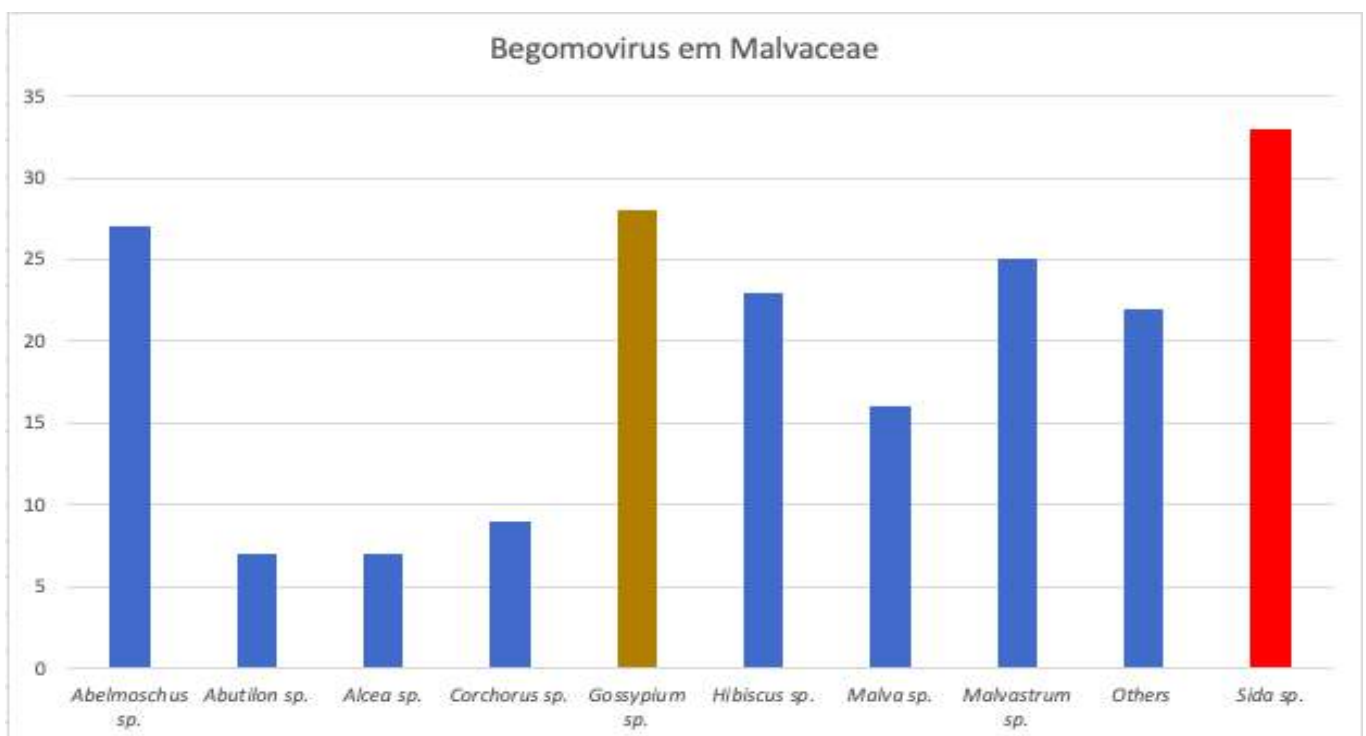
Figure 3. Number of begomoviruses infecting genera of the Malvaceae family in the world according to GenBank database (2022), Virus-Host DataBase (2022) and Kitajima et al. (2020).
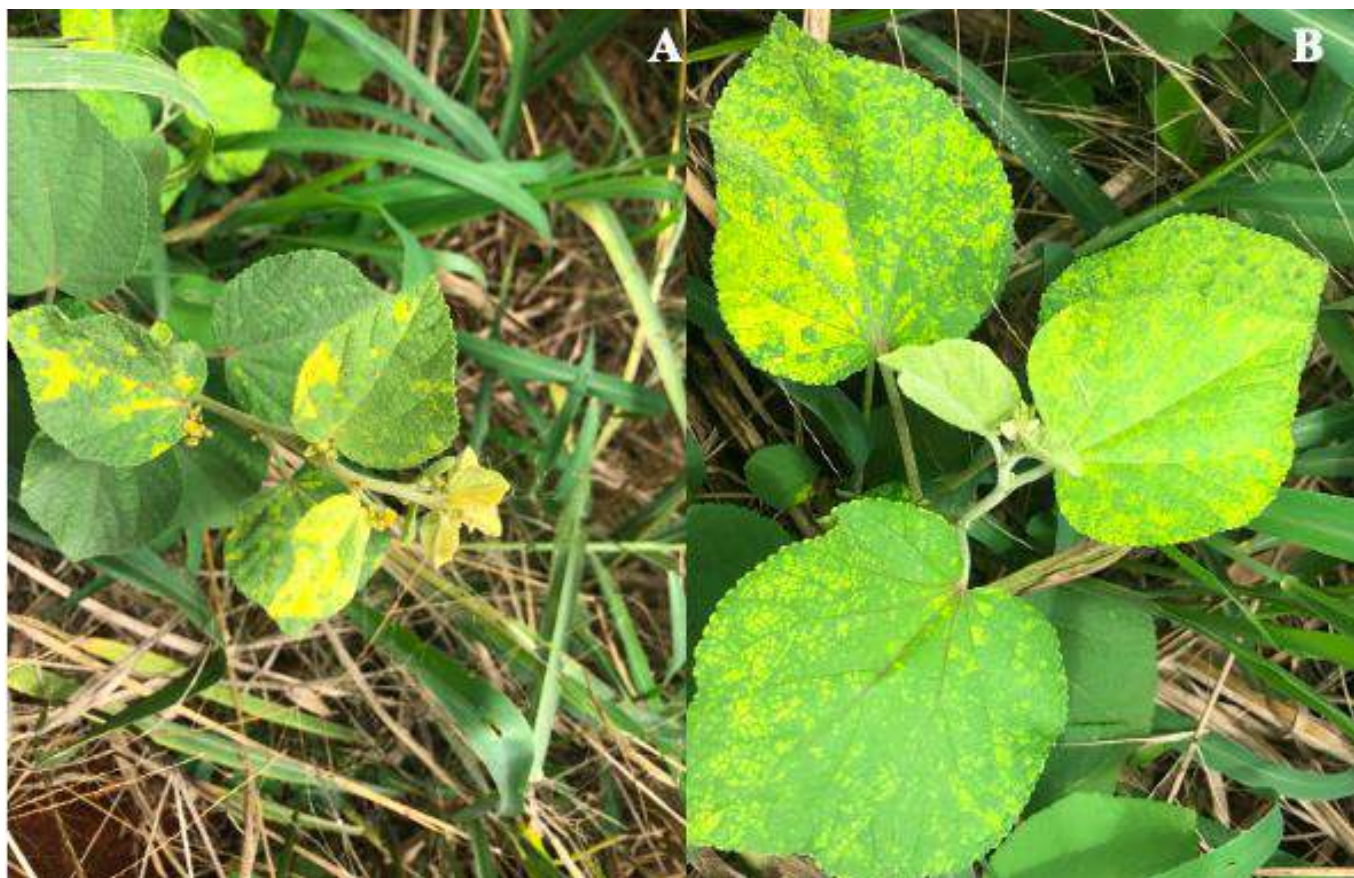
Figure 4. Photos of weeds of the malvaceae family with symptoms of begomovirus in A and B.

1.8 Genetic diversity of begomoviruses and mechanisms of variability

Begomoviruses have mechanisms that allow extreme variability. There are three mechanisms that result in a high genetic variability that can lead to the emergence of new species and strains: mutation, recombination and pseudo-recombination (Roossinck 1997; Seal et al. 2006).

A mutation genetic variation is the main mechanism of genetic variability in begomoviruses occurs due to the incorrect incorporation of nucleotides during viral replication this mechanism can generate variability before the exchange of genomic fragments associated with recombination events. Recombination is the exchanging genomic fragments between two strands of DNA during replication (Padidam et al. 1999; Lefeuvre and Moriones 2015). Pseudorecombination mechanism allows closely related isolates to share the same DNA-B component in the presence of two viruses simultaneously in the

same host (mixed infection). Pseudorecombination may allow the exchange of genomic components among distinct begomoviruses and these genetic rearrangements might allow for rapid viral evolution (Seal et al. 2006). The presence of several begomoviruses under field conditions, transmitted by the same vector, increases the frequency of mixed infections. In turn, this situation increases the probability of emerging novel recombination and pseudo-recombination among different viral genomic components, which can accelerate the generation of novel strains and species (Seal et al. 2006; Silva et al. 2014).

1.9 Use of *High-Throughput Sequencing* (HTS) in plant virology

High-Throughput Sequencing (HTS) has become a very useful tool for discovery novel species and strains in plant virology. Adams et al. (2009) were the first to development and use a metagenomic diagnostic technique for plant viruses. These novel techniques have several advantages over traditional techniques for virus detection, such as ELISA, PCR or hybridization methods. In plant virology, the employment of metagenomics in combination with HTS has enabled numerous discoveries of new viral species and subviral agents and the diversity of species in different hosts. The development of advanced DNA sequencing technologies has allowed the determination of the total nucleic acid content in biological samples. The possibility of using novel techniques for diagnosis of plant diseases, including third-generation sequencing platforms such as Oxford Nanopore is more readily available applicable to plant viruses, including begomoviruses. The *Cowpea bright yellow mosaic* was the first begomovirus sequenced using this new platform in Brazil (Naito et al. 2019). The development of new bioinformatics tools using HTS methods for precise diagnosis will help reduce losses in the field due to emerging diseases (Mehetre et al. 2021).

Blawid et al. (2017) published a practical "pipeline" map and standard alternative tools for use in HTS analysis with the major following steps for virus detection via metagenomic analyses. This pipeline starts with DNA extraction, followed by the viral enrichment in the nucleic acid preparations (for circular DNA viruses the enrichment is done by the Rolling Circle Amplification – RCA). After sample preparation and enrichment, the analyses are carried by analyzing the quality of the sequences followed by the quality control of the sequence data (providing a quick analysis of potential problems that may have occurred during the sequencing procedures). Assessment of the output quality helps to choose the correct preprocessing parameters since low scores of the baseline call quality can negatively impact assembly and mapping. The next step involves trimming of the HTS data, which stars by scanning the 5'-end of the reads and trimming when the average quality per base drops below a given threshold quality. Over-trimming often causes the loss of information, whereas retaining contaminants and low-quality base reads can interfere with downstream sequencing analyses (mapping and the *de novo* assembly of sequences). The "pipeline' map can be detailed for better understanding as follow: **(1)** sample preparation and viral DNA

enrichment; **(2)** quality analysis of advanced sequences from libraries suitable for Illumina Hiseq; **(3)** "Trimming" HTS data; **(4)** Reassembly of the contigs; **(5)** submit it to tBLASTx; **(6)** extension of contigs and *de novo* assembly; and **(7)** final assemblies. These optimized analytical steps have been recently employed by our research group (Reis et al. 2021; Nery et al. 2020).

REFERENCES

Adams IP, Glover RH, Monger WA, Mumford R, Jackeviciene E, Navalinskiene M, Samuitiene M, Boonham N (2009) Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. Molecular Plant Pathology 10:537–545

Al Rwahnih M, Alabi OJ, Westrick NM, Golino D (2018) Prunus geminivirus A: A Novel *Grablovirus* infecting *Prunus* spp. Plant Disease 102:1246–1253

Al-Musa A, Anfoka G, Misbeh S, Mohammad, Haj Ahmad F (2008) Detection and molecular characterization of Squash leaf curl virus (SLCV) in Jordan: molecular characterization of Squash leaf curl virus. Journal of Phytopathology 156:311–316

Argüello-Astorga GR, Ruiz-Medrano R (2001) 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 146:1465–1485

Asssunção IP, Listik AF, Barros MCS, Silva S, Silva I, Ramalho-Neto C, Lima GS (2006) Diversidade genética de *Begomovirus* que infectam plantas invasoras na região nordeste. Planta Daninha 24:239–244

Bahder BW, Zalom FG, Jayanth M, Sudarshana MR (2016) Phylogeny of *Geminivirus* coat protein sequences and digital PCR aid in identifying *Spissistilus festinus* as a vector of grapevine red blotch-associated virus. Phytopathology 106:1223–1230

Bananej K, Shahid MS, Shafiq M (2021) Evidence that leaf curl disease of *Malva sylvestris* in Iran is associated with *cotton leaf curl Gezira virus* and associated betasatellite. Journal of Plant Pathology 103:671–672

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

Bedford ID, Kelly A, Banks GK, Briddon RW, Cenis JL, Markham PG (1998) *Solanum nigrum*: An indigenous weed reservoir for a *tomato yellow leaf curl* geminivirus in southern Spain. European Journal of Plant Pathology 104, 221–222

Bernardo P, Golden M, Akram M, Naimuddin, Nadarajan N, Fernandez E, Granier M, Rebelo AG, Peterschmitt M, Martin DP, Roumagnac P (2013) Identification and characterisation of a highly divergent geminivirus: Evolutionary and taxonomic implications. Virus Research 177:35–45

Bernardo P, Muhire B, François S, Deshoux M, Hartnady P, Farkas K, Kraberger S, Filloux D, Fernandez E, Galzi S, Ferdinand R, Granier M, Marais A, Monge Blasco P, Candresse T, Escriu F, Varsani A, Harkins GW, Martin DP, Roumagnac P (2016) Molecular characterization and prevalence of two capulaviruses: *Alfalfa leaf curl virus* from France and *Euphorbia caput-medusae latent virus* from South Africa. Virology 493:142–153

Berrie LC, Rybicki EP, Rey MEC (2001) Complete nucleotide sequence and host range of South *African cassava mosaic virus*: further evidence for recombination amongst begomoviruses. Journal of General Virology 82:53–58

Blawid R, Silva JMF, Nagata T (2017) Discovering and sequencing new plant viral genomes by next-generation sequencing: Description of a practical pipeline: New plant virus discovery by NGS. Annals of Applied Biology 170:301–314

Briddon RW, Bedford ID, Tsai JH, Markham PG (1996) Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, *Tomato Pseudo-Curly Top Virus,* suggests a recombinant origin. Virology 219:387–394

Brown JK, Zerbini FM, Navas-Castillo J, Moriones E, Ramos-Sobrinho R, Silva JC, Fiallo-Olivé E, Briddon RW, Hernández-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 160:1593–1619

Burgos A, Barua J, Flores-Giubi ME, Bazan D, Ferro E, Alvarenga NL (2015) Antibacterial activity of the alkaloid extract and isolated compounds from *Croton bonplandianum* Baill. (Euphorbiaceae). Revista Brasileira de Plantas Medicinais 17:922–927

Calegário, RF, Ferreira SS, Andrade EC, Zerbini FM (2004) Caracterização de um isolado do Begomovirus Sida micrantha virus (SimMV) obtido de tomateiro. Fitopatologia Brasileira 29:150

Cantú-Iris M, Pastor-Palacios G, Mauricio-Castillo JA, Bañuelos-Hernández B, Avalos-Calleros JA, Juárez-Reyes A, Rivera-Bustamante R, Argüello-Astorga GR (2019) Analysis of a new begomovirus unveils a composite element conserved in the CP gene promoters of several *Geminiviridae genera*: clues to comprehend the complex regulation of late genes. PLoS One 14:e0210485

Castillo-Urquiza GP, Beserra JE Jr, Bruckner FP, Lima AT, Varsani A, Alfenas-Zerbini P, Murilo Zerbini F (2008) Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil. Archives of Virology 153:1985–1989

Chowda Reddy RV, Colvin J, Muniyappa V, Seal S (2005) Diversity and distribution of begomoviruses infecting tomato in India. Archives of Virology 150: 845–867

Claverie S, Bernardo P, Kraberger S, Hartnady P, Lefeuvre P, Lett JM, Galzi S, Filloux D, Harkins GW, Varsani A,

Martin DP, Roumagnac P (2018) From spatial metagenomics to molecular characterization of plant viruses: a geminivirus case study. In: Advances in Virus Research. Elsevier, pp 55–83

Cohen S, Kern J, Harpaz I, Ben-Joseph R (1988) Epidemiological studies of the tomato yellow leaf curl virus (TYLCV) in the Jordan Valley, Israel. Phytoparasitica 16:259

Cotrim MA de A, Krause-Sakate R, Narita N, Zerbini FM, Pavan MA (2007) Diversidade genética de begomovírus em cultivos de tomateiro no Centro-Oeste Paulista. Summa Phytopathologica 33:300–303

De Barro PJ, Liu S-S, Boykin LM, Dinsdale AB (2011) *Bemisia tabaci*: A statement of species status. Annual Review Entomology 56:1–19

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

Fernandes DS, Okuma D, Pantoja-Gomez LM, Cuenca A, Corrêa AS (2022) *Bemisia tabaci* MEAM1 still remains the dominant species in open field crops in Brazil. Brazilian Journal of Biology 84: e256949

Fiallo-Olivé E, Chirinos DT, Geraud-Pouey F, Moriones E, Navas-Castillo J (2013) Complete genome sequences of two begomoviruses infecting weeds in Venezuela. Archives of Virology 158:277–280

Fiallo-Olivé E, Martínez-Zubiaur Y, Moriones E, Navas-Castillo J (2010) Complete nucleotide sequence of *Sida golden mosaic Florida virus* and phylogenetic relationships with other begomoviruses infecting malvaceous weeds in the Caribbean. Archives of Virology 155:1535–1537.

Fontenele RS, Abreu RA, Lamas NS, Alves-Freitas DMT, Vidal AH, Poppiel RR, Melo FL, Lacorte C, Martin DP, Campos MA, Varsani A, Ribeiro SG (2018) *Passion fruit chlorotic mottle virus*: Molecular characterization of a new divergent geminivirus in Brazil. Viruses 10:169

Fontenele RS, Lamas NS, Lacorte C, Lacerda ALM, Varsani A, Ribeiro SG (2017) A novel geminivirus identified in tomato and cleome plants sampled in Brazil. Virus Research 240:175–179

Fontenele RS, Salywon AM, Majure LC, Cobb IN, Bhaskara A, Avalos-Calleros JA, Argüello-Astorga GR, Schmidlin K, Khalifeh A, Smith K, Schreck J, Lund MC, Köhler M, Wojciechowski MF, Hodgson WC, Puente - Martinez R, Van Doorslaer K, Kumari S, Vernière C, Filloux D, Roumagnac P, Lefeuvre P, Ribeiro SG, Kraberger S, Martin DP, Varsani A (2020) A novel divergent geminivirus identified in asymptomatic New World *Cactaceae* Plants. Viruses 12:398

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

Graham AP, Martin DP, Roye ME (2010) Molecular characterization and phylogeny of two begomoviruses infecting *Malvastrum americanum* in Jamaica: evidence of the contribution of inter-species recombination to the evolution of malvaceous weed-associated begomoviruses from the Northern Caribbean. Virus Genes 40:256–266

Guo X, Shi M, Zhou X (2007) Complete nucleotide sequences of *Malvastrum yellow mosaic virus* and its associated DNAβ molecule. Archives of Virology 152:641–643.

Gutierrez C, Ramirez-Parra E, Mar Castellano M, Sanz-Burgos AP, Luque A, Missich R (2004) Geminivirus DNA replication and cell cycle interactions. Veterinary Microbiology 98:111–119

Hallan V, Saxena S and Singh BP (1998) Ageratum, Croton and Malvastrum harbour geminiviruses: evidence through PCR amplication. World Journal of Microbiology and Biotechnology 14: 931-932

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

Hasanvand V, Kamali M, Heydarnejad J, Massumi H, Kvarnheden A, Varsani A (2018) Identification of a new *turncurtovirus* in the leafhopper *Circulifer haematoceps* and the host plant species *Sesamum indicum*. Virus Genes 54:840–845

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

Huang JF, Zhou XP (2006) Molecular characterization of two distinct begomoviruses from *ageratum conyzoides* and *malvastrum coromandelianum* in China. Journal of Phytopathology 154:648–653

Islam S (2019) A review study on different plants in *Malvaceae* family and their medicinal uses. AJBSR 3:94–97

Kamali M, Heydarnejad J, Massumi H, Kvarnheden A, Kraberger S, Varsani (2016) Molecular diversity of *turncurtoviruses* in Iran. Archives of Virology 161:551–561

Kitajima EW (2020) An annotated list of plant viruses and viroids described in Brazil (1926-2018). Biota Neotropica 20: e20190932

Klute KA, Nadler SA, Stenger DC (1996) *Horseradish curly top virus* is a distinct subgroup II geminivirus species with rep and C4 genes derived from a subgroup III ancestor. Journal of General Virology 77:1369–1378

Krenz B, Thompson JR, Fuchs M, Perry KL (2012) Complete genome sequence of a new circular DNA virus from grapevine. Journal of Virology 86:7715–7715

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

Li P, Jing C, Li K, Du J, Wu G, Li M, Sun X, Qing L (2018) First report of *tobacco curly shoot virus* infecting *Malvastrum coromandelianum* in China. Journal of Plant Pathology 100:339–339

Li P, Su F, Meng Q, Yu H, Wu G, Li M, Qing L (2021) The C5 protein encoded by *Ageratum leaf curl Sichuan virus* is a virulence factor and contributes to the virus infection. Molecular Plant Pathology 22:1149–1158

Liang P, Navarro B, Zhang Z, Wang H, Lu M, Xiao H, Wu Q, Zhou X, Di Serio F, Li S (2015) Identification and characterization of a novel geminivirus with a monopartite genome infecting apple trees. Journal of General Virology 96:2411–2420

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

Liu P, Xie Y, Zhou XP (2009) *Malvastrum coromandelianum* is an alternative host of *Tomato yellow leaf curl China virus*. Plant Pathology 58:403–403

Loconsole G, Saldarelli P, Doddapaneni H, Savino V, Martelli GP, Saponari M (2012) Identification of a single-stranded DNA virus associated with *citrus chlorotic dwarf disease*, a new member in the family *Geminiviridae*. Virology 432:162–172

Lu Q-Y, Wu Z-J, Xia Z-S, Xie L-H (2015) Complete genome sequence of a novel monopartite geminivirus identified in mulberry (*Morus alba* L.). Archives of Virology 160:2135–2138

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. Archives of Virology 165:253–256

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

137:113–121

Mehetre GT, Leo VV, Singh G, Sorokan A, Maksimov I, Yadav MK, Upadhyaya K, Hashem A, Alsaleh AN, Dawoud TM, Almaary KS, Singh BP (2021) Current developments and challenges in plant viral diagnostics: A systematic review. Viruses 13:412

Moradi Z, Mehrvar M Metagenomic Analysis of *Malva sylvestris* from Iran Displays a *Malva Vein Clearing Virus Genome* Journal of Agricultural Science and Technology 25:213–223

Muhire B, Martin DP, Brown JK, Navas-Castillo J, Moriones E, Zerbini FM, Rivera-Bustamante R, Malathi VG, Briddon RW, Varsani A (2013) A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus *Mastrevirus* (family *Geminiviridae*). Archives of Virology 158:1411–1424

Naito FYB, Melo FL, Fonseca MEN, Santos CAF, Chanes CR, Ribeiro BM, Gilbertson RL, Boiteux LS, Pereira-Carvalho RC (2019) Nanopore sequencing of a novel bipartite New World begomovirus infecting cowpea. Archives of Virology 164:1907–1910

Nash TE, Dallas MB, Reyes MI, Buhrman GK, Ascencio-Ibañez JT, Hanley-Bowdoin L (2011) Functional analysis of a novel motif conserved across *Geminivirus* Rep proteins. Journal of Virology 85:1182–1192

Nery FMB, Melo FL, Boiteux LS, Ribeiro SG, Resende RO, Orílio AF, Batista JG, Lima MF, Pereira-Carvalho RC (2020) Molecular characterization of Hovenia Dulcis-Associated Virus 1 (HDaV1) and 2 (HDaV2): New tentative species within the Order Picornavirales. Viruses 12:950

Oerke E-C (2006) Crop losses to pests. The Journal of Agricutural Science 144:31–43

Orozco BM, Hanley-Bowdoin L (1996) A DNA structure is required for geminivirus replication origin function. Journal of Virology 70:148–158

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265:218–225

Pakkianathan BC, Kontsedalov S, Lebedev G, et al (2015) Replication of *Tomato Yellow Leaf Curl Virus* in Its Whitefly Vector, *Bemisia tabaci*. Journal of Virology 89:9791–9803

Pappu HR, Jones RAC, Jain RK (2009) Global status of tospovirus epidemics in diverse cropping systems: Successes achieved and challenges ahead. Virus Research 141:219–236

Paprotka T, Metzler V, Jeske H (2010) The complete nucleotide sequence of a new bipartite begomovirus from Brazil infecting Abutilon. Archives of Virology 155:813–816

Perry KL, McLane H, Thompson JR, Fuchs M (2018) A novel grablovirus from non-cultivated grapevine (*Vitis* sp.) in North America. Archives of Virology 163:259–262

Prajapat R, Marwal A, Gaur RK (2014) Begomovirus associated with alternative host weeds: a critical appraisal. Archives of Phytopathology and Plant Protection 47:157–170

Pramesh D, Mandal B, Phaneendra C, Muniyappa V (2013) Host range and genetic diversity of croton yellow vein mosaic virus, a weed-infecting monopartite begomovirus causing leaf curl disease in tomato. Archives of Virology 158:531–542

Qiu Y, Zhang S, Yu H, Xuan Z, Yang L, Zhan B, Zerbini FM, Cao M (2020) Identification and characterization of two novel geminiviruses associated with Paper Mulberry (*Broussonetia papyrifera*) leaf curl disease. Plant Disease 104:3010–3018

Razavinejad S, Heydarnejad J, Kamali M, Massumi H, Kraberger S, Varsani A (2013) Genetic diversity and host range studies of *turnip curly top virus.* Virus Genes 46:345–353

Reis LNA, Boiteux LS, Fonseca MEN, Pereira–Carvalho RC (2021) *Tomato yellow vein streak virus* and *Tomato*

golden vein virus: a reappraisal of the classification status of two South American Begomovirus species based upon genome–wide pairwise identity of multiple isolates. Virus Genes 57:127–131

Reis LNA, Fonseca, MEN, Ribeiro SG, Naito, FYB, Boiteux, LS, Pereira-Carvalho, RDC. (2020). Metagenomics of neotropical single-stranded DNA viruses in tomato cultivars with and without the Ty-1 *gene*. *Viruses 12*(8), 819.

Rojas MR, Hagen C, Lucas WJ, Gilbertson RL (2005) Exploiting chinks in the plant's armor: Evolution and emergence of geminiviruses. Annual Review of Phytopathology 43:361–394.

Roossinck MJ (1997) Mechanisms of plat virus evolution. Annual Review of Phytopathology 35:191–209

Roshan P, Kulshreshtha A, Hallan V (2019) Global weed-infecting geminiviruses. In: Kumar RV (ed) Geminiviruses. Springer International Publishing, Cham, pp 103–121

Roumagnac P, Granier M, Bernardo P, Deshoux M, Ferdinand R, Galzi S, Fernandez E, Julian C, Abt I, Filloux D, Mesléard F, Varsani A, Blanc S, Martin DP, Peterschmitt M (2015) *Alfalfa Leaf Curl Virus*: an aphid-transmitted geminivirus. Journal of Virology 89:9683–9688

Ryckebusch F, Sauvion N, Granier M, Roumagnac P, Peterschmitt M (2020) *Alfalfa leaf curl virus* is transmitted by *Aphis craccivora* in a highly specific circulative manner. Virology 546:98–108

Sattar MN, Qurashi F, Iqbal Z, Saleem Haider M (2017) Molecular characterization of *Hollyhock leaf curl virus* and associated DNA-satellites infecting *Malva parviflora* in Pakistan. Canadian Journal of Plant Pathology 39:229–234

Seal SE, vandenBosch F, Jeger MJ (2006) Factors Influencing Begomovirus Evolution and Their Increasing Global Significance: Implications for Sustainable Control. Critical Reviews in Plant Sciences 25:23–46

Shepherd DN, Martin DP, Van Der Walt E, Dent K, Varsani A, Rybicki EP (2010) *Maize streak virus:* an old and complex 'emerging' pathogen. Molecular Plant Pathology 11:1–12

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

Soto MJ, Gilbertson RL (2003) Distribution and rate of movement of the curtovirus *Beet mild curly top virus* (Family *Geminiviridae*) in the beet leafhopper. Phytopathology 93:478–484

Stanley J, Markham PG, Callis RJ, Pinner MS (1986) The nucleotide sequence of an infectious clone of the geminivirus *beet curly top virus.* The EMBO Journal 5:1761–1767

Stewart C, Kon T, Rojas M, Graham A, Martin D, Gilbertson R, Roye M (2014) The molecular characterization of a Sida-infecting begomovirus from Jamaica. Archives of Virology 159:375–378

Sudarshana MR, Perry KL, Fuchs MF (2015) *Grapevine Red Blotch-Associated Virus*, an emerging threat to the grapevine industry. Phytopathology 105:1026–1032

Susi H, Laine AL, Filloux D, Kraberger S, Farkas K, Bernardo P, Frilander MJ, Martin DP, Varsani A, Roumagnac P (2017) Genome sequences of a *capulavirus* infecting *Plantago lanceolata* in the Åland archipelago of Finland. Archives of Virology 162:2041–2045

Swanson MM, Harrison BD, Wong SM (1993) A geminivirus causing vein yellowing of *Ageratum conyzoides* in Singapore. Plant Pathology 42:137–139

Tahir M, Amin I, Haider MS, Mansoor S, Briddon RW (2015) Ageratum enation virus—A Begomovirus of weeds with the potential to infect crops. Viruses 7:647–665

Vaghi Medina CG, Teppa E, Bornancini VA, et al (2018) *Tomato Apical Leaf Curl Virus*: A novel, monopartite geminivirus detected in tomatoes in Argentina. Frontiers in Microbiology 8:2665.

van Knippenberg I, Lamine M, Goldbach R, Kormelink R (2005) *Tomato spotted wilt virus* transcriptase in vitro displays a preference for cap donors with multiple base complementarity to the viral template. Virology 335:122–130

Varsani A, Martin DP, Navas-Castillo J, Moriones E, Hernández-Zepeda C, Idris A, Murilo Zerbini F, Brown JK (2014a) Revisiting the classification of *curtoviruses* based on genome-wide pairwise identity. Archives of Virology 159:1873–1882

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

Varsani A, Shepherd DN, Dent K, Monjane AL, Rybicki EP, Martin DP (2009) A highly divergent South African geminivirus species illuminates the ancient evolutionary history of this family. Virology Journal 6:36

Wu H, Liu M, Kang B, Liu L, Hong N, Peng B, Gu Q (2022) AC5 protein encoded by *squash leaf curl China virus* is an RNA silencing suppressor and a virulence determinant. Frontiers in Microbiology 13:980147

Wu J, Mugiira RB, Zhou X (2007) *Malvastrum leaf curl Guangdong virus* is a distinct monopartite begomovirus. Plant Pathology 56:771–776

Wyant PS, Gotthardt D, Schäfer B, Krenz B, Jeske H (2011) The genomes of four novel begomoviruses and a new *Sida micrantha mosaic virus* strain from Bolivian weeds. Archives of Virology 156:347–352

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

Zhang S, Shen P, Li M, Tian X, Zhou C, Cao M (2018) Discovery of a novel geminivirus associated with camellia chlorotic dwarf disease. Archives of Virology 163:1709–1712

CHAPTER 2

Complete bipartite genome characterization of three novel recombinant begomoviruses infecting Malvaceous weeds and their phylogenetic relationships with Neotropical tomato-infecting species.

Complete bipartite genome characterization of three novel recombinant begomoviruses infecting Malvaceous weeds and their phylogenetic relationships with Neotropical tomato-infecting species.

Marcos Silva de Queiroz Ferreira[1], Luciane de Nazaré Almeida dos Reis[1], Felipe Fochat Silva Melo[1], Henrique Honorato[1], Maria Esther N. Fonseca[2], Ailton Reis[2], Leonardo S. Boiteux[1,2], Rita de Cássia Pereira–Carvalho[1].

[1]Departamento de Fitopatologia, Universidade de Brasília (UnB), Brasília–DF, Brazil.

[2]National Center for Vegetable Crops Research (CNPH), Embrapa Vegetable Crops (Hortaliças), Brasília–DF, Brazil.

Abstract – Three novel Malvaceous-infecting begomoviruses were discovered via High-throughput sequencing in Brazil. These viruses were detected in foliar tissue of Malvaceae weeds growing within or near to commercial tomato fields. The new species #1 was detected in three Brazilian regions and displayed DNA-A and DNA-B genomes with 2,660 and 2,663 nts. Species #2 was detected in South Brazil and displayed DNA-A and DNA-B genomes with 2,671 and 2,640 nts. Species #3 was detected a single sample in Northeast region and displayed DNA-A and DNA-B genomes with 2,668 and 2,646 nts. All three novel species displayed typical genomic features of the New World begomoviruses Here, we corroborated the astonishing diversity of begomoviruses in Malvaceae weeds under Neotropical conditions. Moreover, the recombination analyses of these three new begomoviruses suggest that these weeds might play a role as reservoirs of recombinant viral species carrying genomic segments of viral pathogens originally detected infecting tomatoes.

To be submitted to Archives of Virology – **Annotated Sequence**

The *Geminiviridae* family (Order: *Geplafuvirales*) encompasses viruses with circular; single-stranded DNA (ssDNA) genomes, being currently composed of 14 genera. The classification at the genus level is based upon a combination of data about the vector(s), host(s) as well as genome structure and organization (Loconsole et al. 2012; Varsani et al. 2017; Rojas et al. 2018; ICTV 2023). The genus *Begomovirus* aggregates the largest number of viral species within the *Geminiviridae* family (ICTV 2023). This genus is composed by plant-infecting, whitefly-transmitted viruses with either monopartite (DNA-A) or bipartite (DNA-A and DNA-B) genomes. The components of bipartite begomoviruses (≈ 2.600 nucleotides each) are individually encapsidaded in the twinned icosahedral virions (Rojas et al. 2018).

The first reports of *Begomovirus* species infecting tomatoes (*Solanum lycopersicum* L.) in Brazil were done in the late 1950s (Flores et al. 1960). However, outbreaks of begomoviruses in tomatoes increased considerably only after the invasion of *Bemisia tabaci* Middle East-Asia Minor 1 (MEAM 1 = biotype B) in the early 1990s. Subsequent field surveys in all Brazilian regions revealed a complex of more than 20 *Begomovirus* species (composed mainly of bipartite viruses) (Reis et al. 2020; 2021). Evidence that native flora and weeds of the Malvaceae family can be reservoirs of tomato-infecting begomoviruses is accumulating and these pathogens might be transferred from the original hosts to tomatoes and vice-versa (García-Arenal and Zerbini 2019). Thus far, four begomoviruses originally characterized in weeds were detected in association with to tomato crops in Brazil: Sida mottle virus (SiMoV), Sida micrantha mosaic virus (SimMV), Euphorbia yellow mosaic virus (EuYMV) and Sida yellow net virus (SiYNV) (Duarte et al. 2021a; Duarte et al. 2021b; Reis et al. 2020). In the present study, High-Throughput Sequencing (HTS) was employed in combination with Sanger dideoxy sequencing to examine the diversity of begomoviruses associated with species of the Malvaceae family. In the present work, we provide the complete genome characterization of three putatively new bipartite *Begomovirus* species that were found naturally infecting Malvaceous weeds growing in association to commercial tomato fields.

Leaf samples of Malvaceous weeds were collected in association with tomato production areas across all five major geographical regions of Brazil. Samples were collected from plants with typical begomovirus symptoms, including overall chlorosis and typical bright yellow and golden net patterns, leaf deformation, yellow mosaic and spots. Field surveys were carried out from 2001 to 2020 and 78 samples were selected to comprise one single, major pool that was employed for analyses. Total DNA was extracted with a modified CTAB protocol with organic solvents (Boiteux et al. 1999) and stored at -20 °C for subsequent use as template in PCR (polymerase chain reaction) assays. The initial confirmation of the presence of begomovirus infection in the samples was done via PCR assays with the degenerate primers 'PAR1c496' and 'PAL1v1978' (Rojas et al. 1993). Viral circular DNA were enriched by employing rolling-circle amplification (RCA) assays (Inoue-Nagata et al. 2004). HTS was then performed at Agrega (Porto Alegre-RS, Brazil) on *Illumina* NovaSeq 6000 platform. The HTS-derived information was analyzed according to the following workflow (Nery et al. 2020; Reis et al. 2021): (*i*) elimination of low-quality reads; (*ii*) re-assembly of the sequences using the CLC Genomics Workbench 11 program; and (*iii*) validation of the contigs with the BLASTn algorithm by comparing with the ssDNA virus database (https://www.ncbi.nlm.nih.gov/). The viral

contigs were annotated and the trimmed reads were mapped back to the annotated genome using the 'Map to reference' tool available in the Geneious 11.1.5 program (Kearse et al. 2012). The complete DNA–A sequences of three putatively new viral species (detected via HTS) were verified via Sanger dideoxy sequencing employing virus-specific primer pairs (**Supplementary Table 2**). Sequencing reactions were carried out at ACTgene using the BigDye®. Sequences were initially analyzed using the BLASTn algorithm and sequence identity to the closest begomoviruses was calculated with Species Demarcation Tool v.1.2 (SDT) (Martin et al. 2015). Full–length genomes were aligned with MUSCLE. Phylogenetic trees (based on DNA-A alignments) were generated by site-specific Bayesian method with a model selected by IQtree 2 (Minh et al. 2020) with GTR+F+4 with 1,000bootstrap replications. Figures were elaborated with Adobe Illustrator CC and EvolView (He et al. 2016). To detect potential recombination events, the software RDP 5 program (Martin et al. 2021) was used. Recombination events were considered reliable only if they were detected by at least four of the seven methods implemented by the program.

As a result of the *Illumina* NovaSeq *6000* sequencing, 7,391,728 million reads were obtained from a pool of 78 weed plant leaf samples with begomovirus-like symptoms. After assembly, in the CLC Genomics Workbench 11 program, 10,679 contigs were obtained. Most of these contigs were composed of previously characterized begomoviruses. However, three full-length DNA-A contigs displayed identity levels of less than 91%, consistent with these viruses are representing new species under the current taxonomic criterion of the genus *Begomovirus* (Brown et al. 2015). The DNA-A component of the putative new species #1 (177,73 reads) had the highest identity (89.59%) with Sidastrum golden leaf spot virus, whereas the new species #2 (781,407 reads) had the highest identity (89.68 %) with Oxalis yellow vein virus and the new species #3 (1,039,723 reads) had highest identity (89.03%) with Sida yellow mosaic Alagoas virus. After PCR assays with species-specific primers (**Supplementary Table 2**), the full DNA-A genomic components of three putative new species were detected. In addition, the pBL-CRC1 primer was used in PCR assays for specific detection of the DNA-B components of these novel viruses in the individual samples.

The complete DNA-A component of **species #1** was obtained via PCR with overlapping begomovirus-specific primers and displayed a genome with 2,660 nucleotides (nts), whereas the DNA-B displayed a genome with 2,663 nts. New species #1 was detected in four samples (*viz.* DF–707, DF–708, RJ–011, and PE–146). The samples DF–707 and DF–708 were collected in 2019 in Brazlandia–DF, whereas RJ–011 was collected in 2006 in Paty do Alferes–RJ and PE–146 was collected in 2016 in Santa Maria da Boa Vista–PE. The new **species #2** was detected in a single sample (SC–035) collected in 2010 in Tijucas–SC in South Brazil, whereas the new **species #3** was detected also in a single sample (BA–109) collected in 2011 in Jequié–BA. Virus-specific primer pairs targeting the DNA-A component of each putative new species were designed using the HTS-derived genomic information. The samples DF–707 and RJ–011 displayed mixed infections with the *Sida micrantha mosaic virus*. Species #2 was detected in sample (SC–035) presenting a genome with 2,671 nts and the DNA-B displayed a genome with 2,640 nts, also was occurring in mixed infection with the Sida micrantha mosaic virus and species #3 was detected in sample (BA–109) presenting a genome with 2,668 nts and the DNA-B displayed a genome with 2,646 nucleotides

which have the typical of New World begomoviruses with the DNA-A, AV1 encoding coat protein (CP) in the virion sense strand, and four in the complementary-sense strand, AC1 enconding replication associated protein (Rep), AC2 enconding trans–acting protein (TrAP), AC3 enconding replication enhancer (REn), AC4 enconding symptom determinant and AC5 enconding the silencing suppression gene (**Figure 1**). The DNA-B component of bipartite begomoviruses comprises two ORFs, one inthe viral sense (BV1 or NSP–nuclear shuttle protein) that encodes a nuclear transport protein and the other in the complementary sense (BC1 or MP) that encodes an intercellular movement protein. All components have the conserved nonanucleotide (5'–TAATATTAC–3') located at the origin of replication. The iteron found in the sequences of new **species #1**, was GGAGA and the inverted sequence TCTCC (Rep IRD = **MPPAKRFKIQ**). DNA-B was found the iteron GGAGA and its inverted TCTCC sequence (Rep IRD = **MPPAKRFKIQ**). The iteron found in the sequences of new **species #2**, was GGGGGand the inverted sequence CCCCC (Rep IRD = **MPRKGSFCIK**). DNA-B was found the iteron GGGGG and its inverted CCCCC sequence (Rep IRD = **MPRKGSFCIK**). The iteron found in the sequences of new **species #3**, was GGGGG and the inverted sequence CCCCC (Rep IRD = **MPRKGSFCIK**). DNA-B was found the iteron GGGGG and its inverted CCCCC sequence (Rep IRD = **MPRKGSFSIK**) (Argüello-Astorga and Ruiz-Medrano, 2001; **Figure 2**). The pairwise nucleotide sequence identities of DNA-A of three new species and other begomovirus sequences from GenBank were calculated using the SDT. The analysis showed that the new **species #1** sharedin general 69–88% with other begomoviruses, the new **species #2** levels of ranged from 72% to 89% and the new **species #3** levels of ranged from 69% to 91%. Phylogenetic analyzes showed that Sidastrum golden leaf spot virus (SidGLSV; HM357458) was the virus with the closest genetic related (79% identity) to new **species #1**, whereas the new **species #2** was closely related to oxalis yellow vein virus (OxYVV; KM887907) with 81% identity. On the other hand, the new **species #3** displayed the closest genetic relationship with sida yellow mosaic Alagoas virus (SiYMAV; JX871383) and Sida mosaic Bolivia virus 1 (SiMBoV1; HM585441) (**Figure 3**).

Recombination analyzes performed with the RDP 5 program using genomic information from the new **species #1** indicated evidence of recombination events in six statistical methods GENECONV (*p-value* $=2.335 \times 10^{-18}$), BootScan (*p-value* $= 2.941 \times 10^{-24}$), MaxChi (p-value $= 4.912 \times 10^{-19}$), Chimaera (*p-value* $= 4.634 \times 10^{-3}$), SiScan (*p-value* $= 7.777 \times 10^{-15}$), 3Seq (p-value $= 3.285 \times 10^{-18}$). The analyzes showed that new **species #1** closely resemble an isolate of tomato leaf distortion virus (the major parent), whereas the minor parent was sidastrum golden leaf spot virus. The initial recombination breakpoint was detected in the nucleotide #2171 and the final breakpoint was detected in the nucleotide #461, involving sequences from CP and Rep genes. Another genomic region with evidence of recombination was detected by seven methods: RDP statistical (*p-value* $= 5.922 \times 10^{-14}$), GENECONV (*p-value* $= 9.470 \times 10^{-16}$), BootScan (*p-value* $= 1.823 \times 10^{-13}$), MaxChi (p-value $= 2.117 \times 10^{-7}$), Chimaera (*p-value* $= 1.572 \times 10^{-7}$), SiScan (*p-value* $= 1.127 \times 10^{-9}$), 3Seq (p-value $= 8.649 \times 10^{-19}$), involving one tomato leaf distortion virus isolate. The initial breakpoint is at nucleotide #478 and the final breakpoint is at nucleotide #1006 involving sequences from CP and Ren. The analyzes also demonstrated a recombination event with major parent a monopartite begomovirus from the New World tomato leaf curl purple vein virus and minor parent with Sida mottle virus four statistical methods MaxChi (p-value $= 4.340 \times 10^{-8}$), Chimaera (*p-value* $= 1.44 \times$

10^{-9}), SiScan (*p-value* = 2.511 x 10^{-13}), 3Seq (*p*-value = 3.285 x 10^{-9}). The initial breakpoint is at nucleotide #50 and the final breakpoint is at nucleotide #478 involving sequences from CP. The DNA-B of new **species #1** also exhibited evidences of recombination events according to five statistical methods of the RDP5 software (*p-value* = 2.085 x 10^{-11}), BootScan (*p-value* = 2.627 x 10^{-11}), MaxChi (*p-value* = 8.296 x 10^{-19}), Chimaera (*p-value* = 1.982x 10^{-14}), SiScan (*p-value* = 1.606 x 10^{-2}) and 3Seq (*p-value* = 4.280 x 10^{-7}). These analyzes showed that this novel species closely reseamble an isolate of tomato mild mosaic virus (major parent), and one isolate of tomato chlorotic leaf curl virus as a minor parent. The initial breakpoint was at the nucleotide #1580 and the final breakpoint was at the nucleotide #2533, involving sequences of the BC1 gene. This close genetic relationship of weed- and tomato-infecting begomoviruses from Brazil has been previously demonstrated in a wide array of studies (for review Reis et al., 2020).

The analyses of the bipartite genome of new **species #2** with the RDP 5 program showed evidence of recombination events in six statistical methods: RDP statistical (*p-value* = 6.585 x 10^{-04}), BootScan (*p-value* = 7.916 x 10^{-3}), MaxChi (p-value = 4.705 x 10^{-7}), Chimaera (*p-value* = 1.221 x 10^{-7}), SiScan (*p-value* = 2.488 x 10^{-3}) and 3Seq (p-value = 5.435 x 10^{-6}). The analyzes showed that this species closely resemble the major parent which is melonchia mosaic virus. The initial breakpoint is at nucleotide 2.037 and the final breakpoint is at nucleotide 2.648 involving sequences from Rep genes. The cognate DNA-B of the new **species #2** was found but no evidence of recombination was detected.

Analyzes with the RDP 5 program using genomic information of new **species #3** showed evidence of recombination events in seven statistical methods: RDP statistical (*p-value* = 5.906 x 10^{-8}), GENECONV (*p-value* = 3.138 x 10^{-6}), BootScan (*p-value* = 3.788 x 10^{-6}), MaxChi (p-value = 1.314 x 10^{-7}), Chimaera (*p-value* = 8.313 x 10^{-9}), SiScan (*p-value* = 1.014 x 10^{13}) and 3Seq (p-value = 1.716 x 10^{-6}). The analyzes showed that this species closely resemble the major parent, which was an isolate of sida yellow net virus. The initial breakpoint was at nucleotide #105 and the final breakpoint was at nucleotide #418 involving segments of the CP genes. Although the cognate DNA-B of the new species #3 was found, no evidence of recombination was detected.

Isolates of the new **species #1** were detected in Paty do Alferes–RJ (RJ–011), in Santa Maria da Boa Vista-PE in Pernambuco (isolate PE–146) and in Brazlandia in the Federal District (isolates DF–707 and DF–708), demonstrating that this begomovirus is widely disseminated throughout the country. The sample DF–707 had mixed infection with sida micrantha mosaic virus (SiMMV) but no evidence of interaction between the two species was obtained in our analyses. The new **species #1** also displayed significant recombination events in its DNA–B with a tomato mild mosaic virus isolate. In addition, the detection of recombination events among weed and tomato begomoviruses are endorsing the importance of carrying out studies about the relationships among these species. The new **species #2 was** detected only in a single sample from Santa Catarina State (SC–035 isolate) in Tijucas county. RPD5 analyses indicated recombination events with a distinct Malvaceous-infecting begomovirus (*Melonchia mosaic virus*). Mixed infection with SiMMV was also detected in SC–035. It is important to highlight that the number of samples collected in Malvaceous weeds in the South region of Brazil is yet negligible. The dection

of a novel species in a single sample indicates that the diversity of begomoviruses in the southern Brazilian region should be largely underestimated justifying, therefore, the intensification of field surveys in this area. Likewise, the new **species #3** was detected in a single sample (BA–109) in Jequié (Bahia State) in the warm the Northeast region. In contrast with the southern Brazilian region, a larger number of surveys for begomoviruses in Malvaceous weeds has been carried out the region. These surveys revealed that SiMMV, Sida mosaic Alagoas virus; Sida mottle Alagoas as the predominant species. Therefore, we believe that BA–109 might represent a quite unique sample composed of either a yet rare or a novel species with a very recent and endemic emergence. This new **species #3** is also recombinant with genomic contributions of sida yellow net virus, which was also previously reported infecting tomatoes.

Thus far, SiMMV is the most prevalent Malvaceous-infecting begomovirus in Neotropical areas. However, more than 20 begomoviruses have *Sida* and related Malvaceae species as their primary hosts (**Supplementary Table 1**). Begomoviruses reported originally in weeds of distinct botanic families (e.g. blainvillea yellow spot virus and euphorbia yellow mosaic virus) have been also reported infecting species of the genus *Sida*. In addition, Malvaceous weeds have been also reported as alternative hosts of tomato-infecting begomoviruses, including tomato chlorotic mottle virus; tomato mild mosaic virus; tomato mottle leaf curl virus, and tomato yellow spot virus (Tavares et al. 2012; Ferro et al. 2017; **Supplementary Table 1**). To make the scenario worst in terms of management, the recombination analyses described here involving these three novel begomoviruses are suggesting that Malvaceous weeds might play a role as reservoirs of recombinant viral species carrying genomic segments of viral pathogens previously detected infecting tomatoes. These results indicate future epidemiological challenges for the efficient management of begomovirus-induced diseases in some important agricultural crops in Brazil. The establishment of effective disease control strategies is, in fact, more complex since Malvaceous weeds are widely disseminated and many of them have life cyles with semi-perennial characteristics. Under such circsuntances, the occurrence of epidemic outbreaks as well as the emergence of new viral species are very likely biological events for tomatoes and many other host crops under Neotropical environments.

REFERENCES

Argüello-Astorga GR, Ruiz-Medrano R (2001) 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 146:1465–1485

Boiteux LS, Fonseca MEN, Simon PW (1999) Effects of plant tissue and DNA purification method on randomly amplified polymorphic DNA-based genetic fingerprinting analysis in carrot. Journal of the American Society for Horticultural Science 124:32–38

Brown JK, Zerbini FM, Navas-Castillo J, Moriones E, Ramos-Sobrinho R, Silva JC, Fiallo-Olivé E, Briddon RW, Hernández-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 160:1593–1619

Duarte MF, Pereira-Carvalho RC, Reis LNA, Rojas MR, Gilbertson RL, Costa H, Boiteux LS, Fonseca MEN (2021a). Natural infection of tomatoes (*Solanum lycopersicum*) by euphorbia yellow mosaic virus isolates across four Brazilian states. Plant Disease 105: 518.

Duarte MF, Fonseca ME, Costa H, Fernandes NAN, Reis A, Boiteux LS, Pereira-Carvalho RC (2021b) Diversity of tomato-infecting begomoviruses and spatiotemporal dynamics of an endemic viral species of the Brazilian Atlantic rain forest biome. Virus Genes 57:83–93

Ferro CG, Silva JP, Xavier CAD, Godinho MT, Lima ATM, Mar TB, & 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.

Flores E, Silberschmidt K, Kramer M (1960) Observações de "clorose infecciosa" das malváceas em tomateiros do campo. O Biológico 26:65-69

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.

He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic Acids Research. 8;44(W1): W236-41

Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T (2004) A simple method for cloning the complete begomovirus genome using the bacteriophage φ29 DNA polymerase. Journal of Virological Methods 116:209–211

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

Kitajima EW (2020) An annotated list of plant viruses and viroids described in Brazil (1926-2018). Biota Neotropica 20: e20190932.

Loconsole G, Saldarelli P, Doddapaneni H, Savino V, Martelli GP, Saponari M (2012) Identification of a single-stranded DNA virus associated with *citrus chlorotic dwarf* disease, a new member in the family *Geminiviridae*. Virology 432:162–172

Macedo MA, Albuquerque LC, Maliano MR, Souza JO, Rojas MR, Inoue-Nagata AK, Gilbertson RL (2018) Characterization of t*omato leaf curl purple vein virus*, a new monopartite New World begomovirus infecting tomato in Northeast Brazil. Archives of Virology 163:737–743

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 12:7

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37:1530–1534

Nery FMB, Melo FL, Boiteux LS, Ribeiro SG, Resende RO, Orílio AF, Batista JG, Lima MF, Pereira-Carvalho RC (2020) Molecular characterization of Hovenia Dulcis-Associated Virus 1 (HDaV1) and 2 (HDaV2): New tentative species within the Order *Picornavirales*. Viruses 12: 950

Reis LNA, Fonseca MEN, Ribeiro SG, Naito FYB, Boiteux LS, Pereira-Carvalho RC (2020) Metagenomics of Neotropical single-stranded DNA viruses in tomato cultivars with and without the *Ty-1* gene. Viruses 8:1–22.

Reis LNA, Boiteux LS, Fonseca MEN, Pereira-Carvalho RC (2021) Tomato yellow vein streak virus and tomato golden vein virus: A reappraisal of the classification status of two South American *Begomovirus* species based upon genome–wide pairwise identity of multiple isolates. Virus Genes 57:127–131

Rojas MR, Macedo MA, Maliano MR, Soto-Aguilar M, Souza JO, Briddon RW, Kenyon L, 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. Annual Review Phytopathology 56:637–677

Tavares SS, Ramos-Sobrinho R, González-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

Varsani A, Roumagnac P, Fuchs M, et al (2017) *Capulavirus* and *Grablovirus*: Two new genera in the family Geminiviridae. Archives of Virology 162:1819–1831
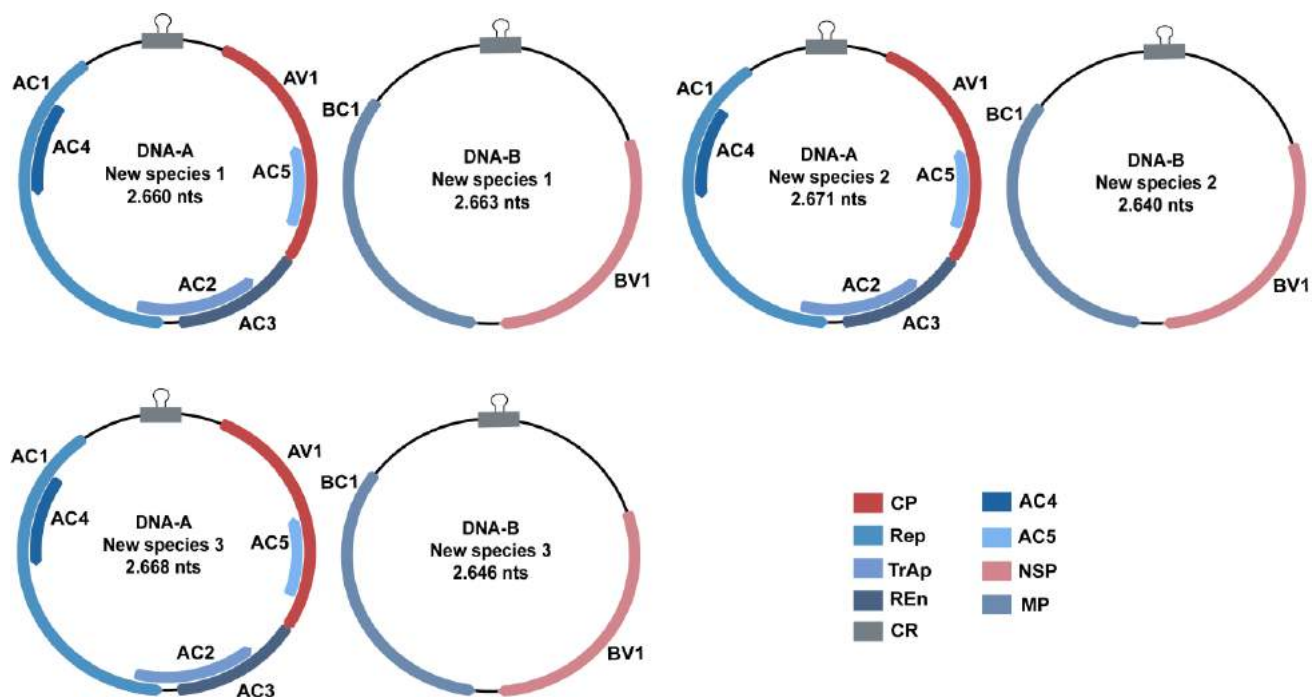
Figure 1. Genomic organization of the three new bipartite *Begomovirus* species infecting Malvaceous weeds. Diagrammatic representation of the circular genomes of **new species #1**, **new species #2** and **new species #3** and their respective open reading frames (ORFs). The ORFs AV1, AC1, AC2, AC3, AC4 and AC5 are according to the putative function of their protein products AV1, encodes the viral coat protein (CP); AV2, responsible for the movement protein (MP); AC1, viral replication-associated protein (Rep); AC2, encodes transcription activator protein (TrAp); AC3, encodes the replication potentiating protein (Ren) AC4, related to symptoms; AC5, related to suppression of gene silencing and pathogenicity CR = common region, encompassing the hairpin.

```
New species 1   A  GTCTCCAATTGAGCTCCCCTCAAACTTGCGATATGTATTGGAGACTGGAGACAATATATAGTAGAGAAGTTCTCTACGACCTCGGAACACGTGGCGG.TAATATTAC   MPPAKRFKIQ
                B  GTCTCCAATTGAGCTCCTCTCAAACTTGGCATATCAATTGGAGACAGGAGACAATATATAGTAGAGAAGTTCTCTAGGATCTC.AGAACACGTGGCGG.TAATATTAC

New species 2   A  ACCCCCAATTGCTCTCCGCTCTAAAACTCATAAGAATTGGGGGTACTGGGGGTACATTTATACTAGAAGTTCCTAAGGGCCTTAAAGGGGCACGT..G.CTAATATTAC   MPRKGSFCIK
                B  ACCCCCAGTTGCTCTCCGCTCTAAAACTCATAAGAATTGGGGGTACTGGGGGTACATTTATACTAGAAGTTCCTAAGGGCCTTAAAGGGGCACGT..G.CTAATATTAC

New species 3   A  ACCCCCAATTGCTCCGCCTTCAAAACTCTATATGAATTGGGGGAACTGGGGGAAAATATATAGTAGAAGTTCCTAGAGGGCACGTGGCGGCCATCCGTTAATATTAC   MPRKGSFSIK
                B  ACCCCCAATTGCTCCGGCTCTCAAAACTCTATACAATTGGGGGAACTGGGGGAAAATATATAGTAGAAGTTCCTAGAGGGCACGTGGCGGCCATCCGTTAATATTAC
```

Figure 2. A segment of the intergenic region showing iterons, TATA region, nonanucleotide, stem-loop and at the end Rep = IRD (Rep Iteron-Related Domain).
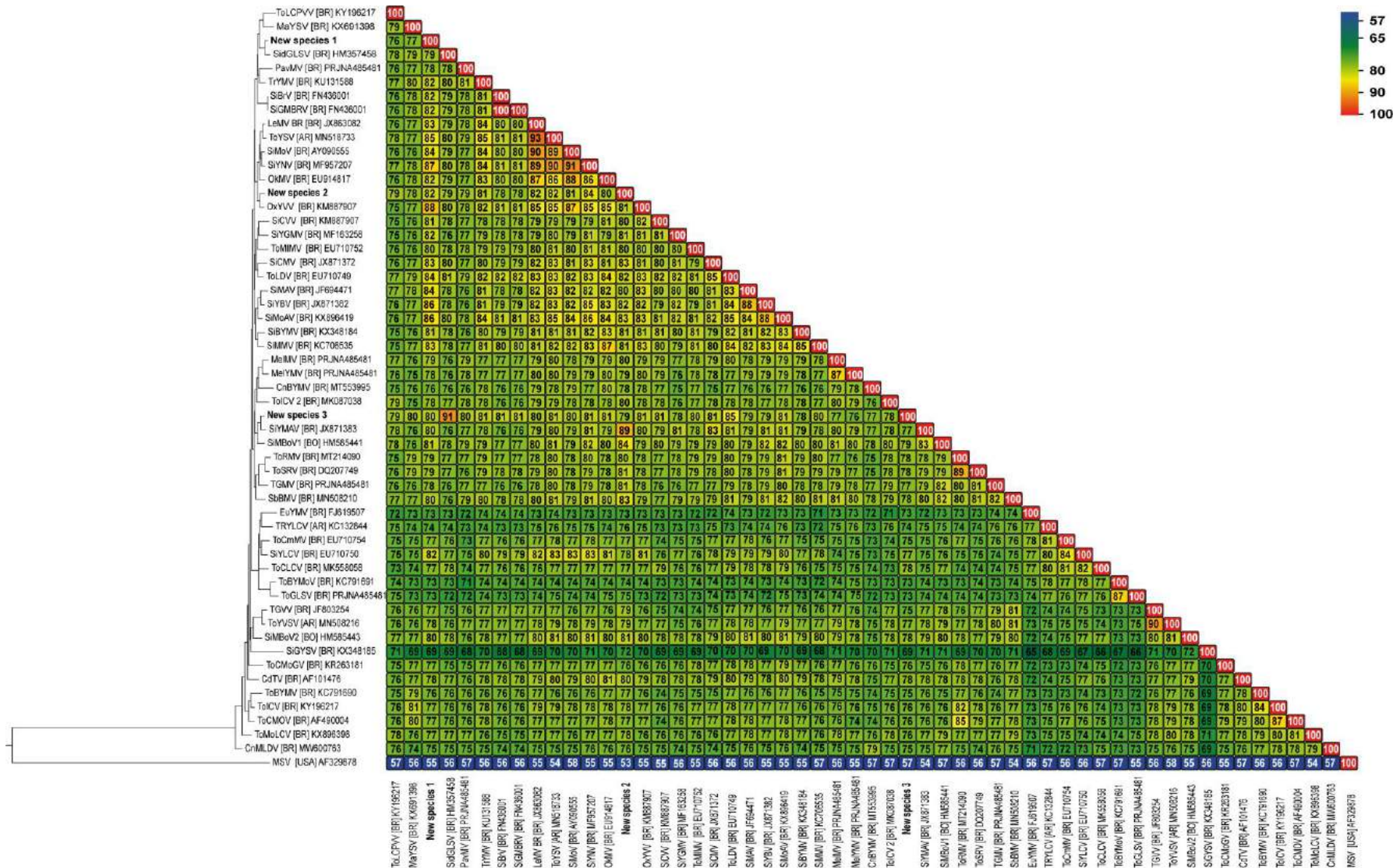
Figure 3. Pairwise identity in Sequence Demarcation Tool (SDT) analysis carried out using the information of the DNA–A sequences of selected New World *Begomovirus* species showing their phylogenetic identities/distances with three new weed plants infecting Malvaceae species: new species #1, new species #2 and new species #3. These *Begomovirus* species were identified by their accession number and by the acronym of the countries where they were described: AR = Argentina; BR = Brazil; BO = Bolivia. Viral species and GenBank accession numbers are as follow: Tomato leaf curl purple vein virus isolate – ToLCPVV [BR] KY196217; Macroptilium yellow spot virus – MaYSV [BR] KX691398; New species #1; Sidastrum golden leaf spot virus – SidGLSV [BR] HM357458; Pavonia mosaic virus – PavMV [BR] PRJNA485481; Triumfetta yellow mosaic virus – TrYMV [BR] KU131588; Sida Brazil virus – SiBrV [BR]FN436001; Sida golden mosaic Brazil virus – SiGMBRV [BR] FN436001; Leonurus mosaic virus – LeMV [BR] JX863082; Tomato yellow spot virus – ToYSV [AR] MN518733; Sida mottle virus – SiMoV [BR] AY090555; Sida yellow net virus – SiYNV [BR] MF957207; Okra mottle virus – OkMV [BR] EU914817; New species #2; Oxalis yellow vein virus – OxYVV BR] KM887907; Sida chlorotic vein virus – SiCVV [BR] KX691405; Sida yellow golden mosaic virus – SiYGMV [BR] MF163258; Tomato mild mosaic virus – ToMIMV [BR] EU710752; Sida common mosaic virus – SiCMV [BR] JX871372; Tomato leaf distortion virus – ToLDV [BR] EU710749; Sida mosaic Alagoas virus – SiMAV [BR] JF694471; Sida yellow blotch virus – SiYBV [BR] JX871382; Sida mottle Alagoas virus – SiMoAV [BR] KX896419; Sida bright yellow mosaic virus – SiBYMV [BR] KX348184; Sida micrantha mosaic virus – SiMMV [BR] KC706535; Melochia mosaic virus - MelMV [BR] PRJNA485481; Melochia yellow mosaic virus - MelYMV [BR] PRJNA485481; Cnidoscolus blistering yellow mosaic virus – CnBYMV [BR] MT553995; Tomato interveinal chlorosis virus-2 – ToICV-2 [BR] MK087038; New species #3; Sida yellow mosaic Alagoas virus – SiYMAV [BR] JX871383; Sida mosaic Bolivia virus 1 – SiMBoV1 [BO] HM585441; Tomato rugose mosaic virus – ToRMV [BR] MT214090; Tomato severe rugose virus – ToSRV [BR] DQ207749; Tomato golden mosaic virus – TGMV [BR] PRJNA485481; Soybean blistering mosaic virus – SbBMV [BR] MN508210; Euphorbia yellow mosaic virus – EuYMV [BR] FJ619507; Tomato rugose yellow leaf curl virus – TRYLCV [AR] KC132844; Tomato common mosaic virus – ToCmMV [BR] EU710754; Sida yellow leaf curl virus – SiYLCV [BR] EU710750; Tomato chlorotic leaf curl virus – ToCLCV [BR] MK558058; Tomato bright yellow mottle virus – ToBYMoV [BR] KC791691; Tomato golden leaf spot virus – ToGLSV [BR] PRJNA485481; Tomato golden vein virus – TGVV [BR] JF803254; Tomato yellow vein streak virus – ToYVSV [AR] MN508216; Sida mosaic Bolivia virus 2 – SiMBoV2 [BO] HM585443; Sida golden yellow spot virus – SiGYSV [BR] KX348185; Tomato chlorotic mottle Guyane virus – ToCMoGV [BR] KR263181; Chino del tomate virus – CdTV [BR] AF101476; Tomato bright yellow mosaic virus – ToBYMV [BR] KC791690; Tomato interveinal chlorosis virus – ToICV [BR] JF803252; Tomato chlorotic mottle virus – ToCMOV [BR] AF490004; Tomato mottle leaf curl virus – ToMoLCV [BR] KX896398; Cnidoscolus mosaic leaf deformation virus – CnMLDV [BR] MW600763; and Maize streak virus – MSV [USA] AF32987.

Supplementary Table 1. Begomoviruses described infecting Malvaceae species in Brazil according to Kitajima (2020) and with sequences deposited in the GenBank database.

Host	Begomovirus
Abelmoschus esculentus	Sida micrantha mosaic virus
	Infectious chlorosis of Malvaceae "complex"
Abutilon striatum	Abutilon mosaic Brazil virus
	Bean golden mosaic virus
Corchurus hirtus	Corchorus mottle virus
Gaya guerkeana	Gaya yellow mosaic virus
Gossypium hirsutum	Cotton chlorotic spot virus
	Infectious chlorosis of Malvaceae "complex"
	Sida micrantha mosaic virus
Herissantia crispa	Begomovirus unclassified
Hibiscus species	Hibiscus golden mosaic virus
Malva species	Sida micrantha mosaic virus
Malva parviflora	Okra mosaic Mexico virus
Malvastrum coromandelianum	Infectious chlorosis of Malvaceae "complex"
Malvaviscus arboreus	Malvaviscus yellow mosaic virus
Melochia species	Melochia mosaic virus
	Melochia yellow mosaic virus
Pavonia species	Pavonia mosaic virus
	Pavonia yellow mosaic virus
Sida species	Sida angular mosaic virus
	Sida bright yellow mosaic virus
	Sida chlorotic mottle virus
	Sida chlorotic vein virus
	Sida common mosaic virus
	Sida golden mosaic Brazil virus
	Sida micrantha mosaic virus
	Sida mosaic Alagoas virus
	Sida mottle Alagoas virus
	Sida mottle virus
	Sida yellow leaf curl virus
	Sida yellow mosaic Alagoas virus
	Sida yellow net virus
	Sida golden yellow mosaic virus
	Sida yellow spot virus
	Tomato chlorotic mottle virus
	Tomato mild mosaic virus
	Tomato mottle leaf curl virus
	Tomato yellow spot virus
Sidastrum micranthum	Sidastrum golden leaf spot virus
Triumfetta species	Triumfetta yellow mosaic virus
	Infectious chlorosis of Malvaceae "complex"
Wissadula	Infectious chlorosis of Malvaceae "complex"

Supplementary Table 2. PCR primer pairs designed based upon *High-Throughput Sequencing* viral consensus sequences for validation of the new Begomovirus species identified in the weed DNA sample pools. For = forward direction and Rev = reverse direction.

Viral species	*Primer* Name	Sequence 5'-3'	Annealing temperature °C
New *Begomovirus* species #1 DNA-A	Contig 4-For	GTTCTAAGAACATCATTACTAACTGCCTAA	
	Contig 4-Rev	ACACTTTCCCAATTATTAGCCCTAGAAACC	
New *Begomovirus* species #2 DNA-A	Contig 5-For	TTGTCGGTCCAATAATCTG	55°
	Contig 5-Rev	CTATTTCATCACGTACCCCC	
New *Begomovirus* species #3 DNA-A	Contig 11-For	CTTAAAGTGAAAGACAAACAGGAGGC	
	Contig 11-Rev	GGAGCTACACGAAGATGGGGAGC	

CHAPTER 3

Reappraisal of the diversity of Sida micrantha mosaic virus (SiMMV) and Sida mottle virus (SiMoV) complex: Species status, host range, and geographical distribution

Reappraisal of the diversity of Sida micrantha mosaic virus (SiMMV) and Sida mottle virus (SiMoV) complex: Species status, host range, and geographical distribution

Marcos Silva de Queiroz Ferreira[1], Luciane de Nazaré Almeida dos Reis[1], Maria Esther N. Fonseca[2], Henrique Honorato[1], Felipe Fochat Melo[1], Ailton Reis[2], Leonardo S. Boiteux[1,2], Rita de Cássia Pereira–Carvalho[1].

[1]Departamento de Fitopatologia, Universidade de Brasília (UnB), Brasília–DF, Brazil.

[2]National Center for Vegetable Crops Research (CNPH), Embrapa Vegetable Crops (Hortaliças), Brasília–DF, Brazil.

Abstract – Sida micrantha mosaic virus (SiMMV) and Sida mottle virus (SiMoV) were reported as distinct begomoviruses infecting Malvaceae weeds and a wide range of economically important hosts in the neotropics. However, their close phylogenetic relationship considering taxonomic criteria for demarcation *Begomovirus* species has generated uncertainties about their taxonomic status and nomenclature. Indeed, DNA-A genomic identity levels of reported isolates with an identical virus name can range from 81 to 100%. In view of the potential imprecision regarding the classification status of these viruses we performed a comprehensive set of analyzes in according to Brown et al (2015). A set of 54 isolates designated as SiMMV or SiMoV available in the GenBank with complete DNA-A sequences were analyzed. Two well-defined clusters were observed, and they are consistent with current taxonomic criteria for species determination this virus group. These distinct phylogenetic groups as well as the taxonomy position of the begomovirus of the SiMMV and SiMoV complex will be discussed and presented bellow.

To be submitted to Virus genes – Short communication

Introduction

The *Begomovirus* is the largest genus in the *Geminiviridae* family with ≈ 445 species (ICTV, 2023). These ssDNA viruses have either monopartite or bipartite circular genomes. The bipartite genome has two components DNA-A and DNA-B (≈ 2,600 nucleotides each), encoding six ORFs. This high number of new emerging of begomoviruses is related to efficient transmission by *Bemisia tabaci* Middle East Asia Minor 1–MEAM1 (= biotype B). The economic importance of begomoviruses has increasing worldwide, and the taxonomic criteria at species level were constantly reexamined over the years. Brown et al. (2015) carried out analyses combining MUSCLE alignment and Sequence Demarcation Tool (SDT) and established a new set of criteria fornew species and new strain demarcation. The complete sequence of the DNA-A component exhibiting less than 91% identity with all previously known begomoviruses is currently employed as the threshold for a novel *Begomovirus* species, which was slightly diferernt from the previous threshold of 89% of identity (Brown et al. 2015).

Malvaceous weeds are widely distributed around the world due to their outstanding environmental adaptability. These weeds may also serve as reservoirs of a wide array of begomoviruses (Garcia-Arenal and Zerbini, 2019). These weeds play relevant epidemiological roles since the survival and spread of begomoviruses may occur in the absence of their hosts (Bedford et al. 1988). Over decades of vegetative propagation of weeds, mainly from the genus *Sida* persistently harbored virus, including several begomoviruses (Prajapat et al. 2014). Beside this, there is a wide variety of weed hosts of begomovirus classified in the Malvaceae family, among them many are classified in the genus *Sida* (**Supplementary Table 1 – ST1**).

The first formal report of Sida micrantha mosaic virus (SiMMV) was carried out in a mixed infection in an member of the Abutilon mosaic virus complex with a foliar sample from Campinas, São Paulo in 1977 (Jovel et al. 2004). This initial description of SiMMV was made with complete sequences of DNA-A (2675 nts; AJ557451) and DNA-B components (2652 nts; AJ557452). After this report, a novel SiMMV isolate (MG-Bi2) was detected in association with tomato plants in São Joaquim de Bicas-MG (Calegario, 2004 Masters dissertation), but the genomic information of this tomato-infecting isolate is not yet available in GenBank. The identity after MUSCLE alignment of this tomato-infecing isolate displayed 87% identity with Sida mottle virus and 81% identity with SiMMV. Alignments with other viruses from the same region of this SiMMV isolate described by Calegario (2004) we found that it corresponds to the first description of tomato yellow spot virus with the isolate called TGV[BI-2] infecting tomato in São Joaquim de Bicas,

Minas Gerais (Calegario et al. 2004). After that, an isolate considered as the true Sida micrantha mosaic virus was published (Jeske et al. 2010) with its corresponding DNA-A (2662 nts; FN557522) and DNA-B (2632 nts; FN557523). The complete DNA-A (NC_005330) and DNA-B (NC_005331) from the sida micrantha isolate were latter established as the reference sequences for this species. In turn, the first isolate of Sida mottle virus (SiMoV), was deposited in GenBank (AY090555) in the year 2002, but without its cognate DNA-B.

Since then, SiMMV isolates have been predominantly reported in a wide range of economically important hosts in the neotropics (Alves 2012; Calegario et al. 2004; Castillo-Urquiza et al. 2008; Fernandes et al. 2009; Fernandes-Acioli 2011; Fontenele et al. 2018; Reis et al. 2020; Hoffmann et al. 2021) with no additional report of SiMoV. However, our preliminary analyses indicated a very close phylogenetic relationship of SiMMV and SiMoV, which generated uncertainties about their taxonomic status and nomenclature. We carried out preliminary analysis and we detected that the isolate described by Jeske (2010) as the real SiMMV is, in fact, closer related to the original SiMoV isolate. Indeed, DNA-A genomic identity levels of reported isolates with an identical virus name can range from 81 to 100% at the GenBank database, demonstrating a taxonomic inconsistency of SiMMV and SiMoV.

In this present work, all SiMMV and SiMoV accessions available thus far in GenBank were analyzed aiming to elucidate the actual diversity present of this complex and to verify the possible erroneous nomenclature of these Malvaceous-infecting begomoviruses. In addition, we also provided an update about the geographical distribution and host range of of the members of in SiMMV complex across all regions of Brazil via High-Throughput Sequencing (HTS) with 78 novel samples obtained from 2001 to 2020 of weeds plants classified in Malvaceae family. Therefore, our major objective is to provide an extensive reappraisal of the current diversity of SiMMV and SiMoV complex.

Material and Methods

Leaves of 78 samples of weeds plants classified in Malvaceae family, showing symptoms of begomoviruses were collected in tomato production areas and areas close to tomato crops (**Figure 1**), in all regions of the country. The samples were carried out from 2001 to 2020. Each sample was subject to individual DNA total extracted by modified protocol CTAB + solvent (Boiteux et al. 1999) (**Supplementary Table 1**). The DNA quantification of the samples was performed using nanodrop and the integrity of molecule visualized in 1% gel agarose. The total genomic DNA obtained from each sample was used as a template for Rolling Circle Amplification – RCA (Inoue-Nagata et al. 2004). The purified total DNA was subjected to

polymerase chain reaction (PCR) assays to confirm the presence of begomoviruses in these weeds leaf samples. Amplicons derived from a segment of the DNA–A component were obtained using the 'universal' primer pairs 'PAL1v1978/ PAR1c496' (Rojas et al. 1993).

Samples were combined in a *pool* and subjected to HTS to identify viral populations. The products obtained from the samples were sequenced on the *Illumina* NovaSeq 6000 platform. The sequences obtained were analyzed using bioinformatics tools. The *reads* obtained were assembled initially using CLC Genomic Workbench 7.5 program (Qiagen) to obtain *contigs*. The genomes of the individual *contigs* were assembled with *Geneious* program and the Map to reference tool (parameter 90 to 99% of *minimum overlap identity*) with mapping in the provided reads file and analyzed with the *Geneious* program (Kearse et al. 2012). The contigs were compared with sequences deposited in the GenBank database The contigs were compared with sequences deposited in the GenBank database (https://www.ncbi.nlm.nih.gov/) using BLASTn. Muscle alignments were performed using the *Geneious* program to annotate *ORFs* (*Open Reading Frames*) based on the reference genome.

Results

As a result of the *Illumina* NovaSeq 6000 sequencing, 7,391,728 million reads were obtained from a pool of 78 weed plant leaf samples with begomovirus-like symptoms. After assembly, in the CLC Genomics Workbench 11 program, 10,679 contigs were obtained. Were assembled with eight among them corresponding to SiMMV. PCR tests were conducted on the individual samples for detection of SiMMV. Of the 78 samples, 41 samples were positive for the presence of the virus. PCR assays with the previously selected virus-specific primers (**Supplementary Table 2)** were carried out with in 41 individual DNA samples (**Supplementary Table 3**). These assays were performed to validate the HTS. A Muscle alignment was performed to verify if there was a difference among the sequences and a possible new strain between the species within the *pool*. The identity among the was 97% to 99.2%, given the high similarity among the *contigs*, and three were selected for evaluation. When aligned with the sequences available in GenBank, there was a variation of 81% to 99%, demonstrating a high variation among the deposits denominated as Sida micrantha mosaic virus.

In view of the potential inaccuracy regarding the taxonomic status and nomenclature of these viruses, we carried out a comprehensive set of analyzes aiming to clarify the genetic relationships among isolates previously characterized as either SiMMV or SiMoV. The phylogenetic analyzes were performed with the

MUSCLE alignment using the software Sequence Demarcation Tool – SDT (Muhire et al. 2014) reconstructions were performed using the IQtree program by the bayesian method, model TIM3+F+R3 with 1000 bootstrap replications. All 54 available isolates (including our three new isolates, named contigs 23, 43 e 44) with complete DNA-A sequences named as either SiMMV (n=50) or SiMoV (n=4) were obtained from the GenBank characterized in our surveys as well as the ones available at the GenBank database (https://www.ncbi.nlm.nih.gov/) were employed in this set of analyses. Phylogenetic reconstructions were carried out according to Reis et al. (2021). Evaluations of iterons and conserved motifs in the common region (RC) among SiMMV and SiMoV isolates were performed according to Argüello-Astorga and Ruiz-Medrano (2001).

A subset of isolates identified as SiMMV group 1 shared up to 95 to 99%, group 2 shared up to 90% to 94% and the group 3 of SiMoV and a SiMMV isolate shared up to 81% to 83% identity. Evaluations of iterons and iteron-related domain (IRD) separated into three groups of isolates displaying either the GGGGT iteron, the GGGTA iteron or the GGTAG iteron. The group #1 comprises 18 SiMMV isolates with iteron GGGGT isolates have the same IRD (MPPPKRFK<u>I</u>S) (FN436003, FN436005, KX348155, KX348157, KX348160, KX348161, KX348162, KX348163, KX348164, MT103974, MT103979, MT103980, MT103981, MT103982, MT103983, MT103984, MT103985 and MT103986) and a subgroup #1 with four isolates of SiMMV with iteron GGGTA with IRD (MPP<u>P</u>K<u>R</u>F<u>K</u>IS) (HM585433, KX348156, KX348158 and KX348159). The group #2 included 18 isolates of SiMMV with iteron GGGGT and IRD (MPPPKRFK<u>I</u>S) (AJ557451, Contig 23, Contig 44, Contig 47, FJ686693, HM585431, HM585437, HM585439, KC706535, KC706536, KC706537, KU852503, KX691401, KX691410, KY650717, KY650722, MT214092 and MT733803) and a subgroup #2 with four isolates of SiMMV with iteron GGTAG and IRD (MPS<u>A</u>PKRFQI) (JX415187, JX415194, JX415195 and MT733814). The Group #3 comprises one isolate of SiMMV (FN557522) and three isolates of SiMoV (AY090555, AJX871377 and JX871378), after exhaustive searches for a cognate DNA-B for this group, it was found that there is still no DNA-B reported for this group #3 Sida mottle virus.

Discussion

Sida micranta mosaic virus (SiMMV) was formally described by Jovel et al. (2004). The work was entitled "Sida micrantha mosaic is associated with a complex infection of begomoviruses different from Abutilon mosaic virus". The accession code AJ557451, corresponds to the complete DNA-A sequence and was used as a SiMMV reference years later. As this accession is being used as a reference, the Reference Code is NC_005331. Subsequently, Jeske et al (2010), published the paper entitled: "In planta cloning of

geminiviral DNA: The true Sida micrantha mosaic virus". In this work, foliar tissue collected and preserved as herbarium samples in the IAC (Campinas, São Paulo-SP) were used for virus characterization. One of the, named SP77 isolate, was kindly provided by Dr. Alvaro Santos Costa. In our MUSCLE alignment and SDT analyses of the SP77 isolate (FN557522) showed an identity of 81.52% with the reference SiMMV isolate (AJ557451) initially published by Jovel et al. (2004). On the other hand, the SP77 isolate shows 91% identity with the Sida mottle virus (SiMoV) isolate, which sequence was submitted to GenBank in 2002 (by E. Fontes) and made available in 2004 (AY090555). Aftwards, Calegário (2004) carried out the molecular and biological characterization of SiMMV. However, it is important to mention that there is thus far no formal molecular characterization of the original SiMMV and SiMoV isolates.

After that first initial reports, Brown et al. (2015) considered both as reference, the deposit AY090555 (SiMoV) and the deposit AJ557450 (SiMMV). In this work by Brown et al. (2015) it is understood that the cognate DNA-B of SiMoV would be AJ557454, whose sequence corresponds to the DNA-B of SiMMV. In our analyses, the sequence of this putative SiMoV DNA-B, when subjected to common region analysis, has 200 nucleotides and an identity of 87.75%. However, our observations provided strong evidence that SiMMV isolates correspond to a complex of viral strains harboring two clusters with greater than 91% identity. Our analyzes consistently indicated identity levels above 91% among the majority of the SiMMV isolates. The SiMMV complex may have been by misnamed Sida micrantha mosaic virus, although SiMoV was first deposited in GenBank in 2002. Until the present moment, there is no molecular characterization of this virus.

In Brazil, many examples of weed and crop species were identified as reservoir of begomovirus, mainly SiMMV. Some examples can be cited. *Sida sp*. acts as reservoir of Sida micrantha mosaic virus (SiMMV) and tomato mild mosaic virus (ToMlMV), which were found infecting tomato (*Solanum lycopersicum*) and bean (*Phaseolus vulgaris*) respectively (Castillo-Urquiza et al. 2008). Fernandes et al. (2009) detected SiMMV in soybean plants collected at Santo Antonio de Goiás (Goiás State). Fernandes-Acioli et al. (2011) reported SiMMV infecting common bean in producing regions in the state of Goias. Alves et al. (2012) reported the infection of passion fruit (*Passiflora edulis*) collected from São Fidelis state of Rio de Janeiro, Paragominas (Pará), Araguari and Patos de Minas (Minas Gerais) with SiMMV. In this work, the authors related an isolate with 90% of identity SiMoV. Fontenele et al. (2018) reported SiMMV in *Oxalis* species in three tomato production areas: Gama and Rajadinha (Federal District) and Formosa, Goiás State. Hoffmann et al. (2021) reported SiMMV in cotton plants (*Gossypium*) in the state of Goiás.

It is clear that SiMMV isolates play important epidemiological role under Brazilian conditions, and your taxonomic position and nomenclature deserve attention. According to Jeske et al. (2010) the accession (FN557522) called the true SiMMV, in fact it is an isolate of SiMoV, considering the criteria of the first deposit in GB. In our analyses, there is a cognate of DNA-B but the same DNA-B has no relationship with the other isolates of SiMoV.

In summary, our hypothesis is that SiMMV-related begomoviruses comprise a complex of large-spectrum viral strains able to infect members of at least four botanical families (Solanaceae, Fabaceae, Oxalidaceae, and Passifloraceae). Reassessments of the taxonomic status of begomoviruses in Neotropical areas are necessary due to the large number of species that have been simultaneously described by distinct research groups, especially after *B. tabaci* MEAM 1 invasion in the early 1990s. Similar misidentification problem of begomoviruses reported with two South American begomoviruses (tomato golden vein virus – TGVV and tomato yellow veins streak virus –ToYVSV) in which a reappraisal of the classification status based upon genome-wide pairwise identity of multiple isolates was able to identify a large number of misnamed isolates in the GenBank database (Reis et al. 2021).
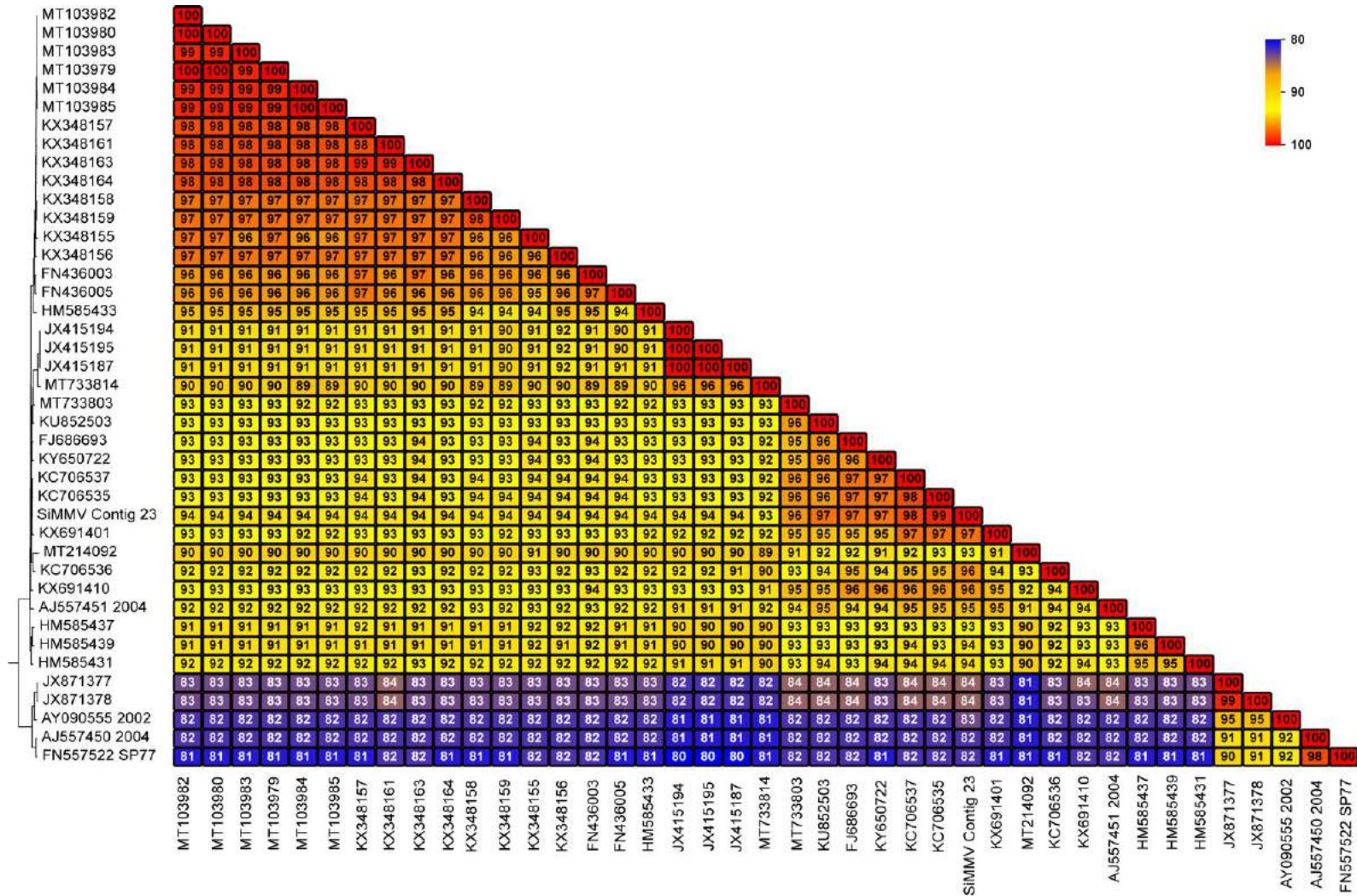
Figure 1. Phylogenetic tree and Sequence Demarcation Tool (SDT) of a set of DNA-A component sequences showing the phylogenetic identities/distances among Sida micrantha mosaic virus (SiMMV) and Sida Mottle virus (SiMoV) isolates. These isolates are identified by their accession number GenBank accession numbers of isolates classified/named

as SiMMV are following: FN436003, FN436005, KX348155, KX348157, KX348160, KX348161, KX348162, KX348163, KX348164, MT103974, MT103979, MT103980, MT103981, MT103982, MT103983, MT103984, MT103985, MT103986, HM585433, KX348156, KX348158, KX348159, AJ557451, Contig 23, Contig 44, Contig 47, FJ686693, HM585431, HM585437, HM585439, KC706535, KC706536, KC706537, KU852503, KX691401, KX691410, KY650717, KY650722, MT214092, MT733803, JX415187, JX415194, JX415195 and MT733814. These isolates are identified by their accession number GenBank accession numbers of isolates classified/named as SiMoV FN557522 and three isolates of SiMoV are following: AY090555, AJX871377 and JX871378.

References

Alves ACCN (2012) Identificação de isolados do Sida mottle virus e Sida micrantha mosaic virus não transmissíveis por *Bemisia tabaci* biótipo B que infectam maracujazeiros (*Passiflora edulis* f. *flavicarpa*). Doutorado em Fitopatologia, Universidade de São Paulo

Argüello-Astorga G, Ruiz-Medrano R (2001) 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. 146:1465–1485.

Bedford ID, Briddon RW, Brown JK, Rosell RC and Markham PG (1998) Geminivirus transmission and biological characterization of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. Annals of Applied Biology. 125: 311-325

Boiteux LS, Fonseca MEN, Simon PW (1999) Effects of plant tissue and DNA purification method on randomly amplified polymorphic DNA-based genetic fingerprinting analysis in carrot. Journal American Society Horticulture Science.124:32–38.

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 Evolutionary Biology. 10:1–17.

Calegario RF (2004) Characterization of an isolate of the begomovirus Sida micrantha mosaic virus (SimMV), Master thesis

Calegario R, Ferreira S, Andrade C, Zerbini F (2004) Caracterização de um isolado do begomovírus sida micrantha mosaic virus (SiMMV) obtido de tomateiro. Fitopatologia Brasileira. 29:150.

Castillo-Urquiza GP, Beserra JE Jr, Bruckner FP, Lima AT, Varsani A, Alfenas-Zerbini P, Murilo Zerbini F (2008) Six novel begomoviruses infecting tomato and associated weeds in Southeastern Brazil. Archives of Virology 153:1985–1989

Fernandes DS, Okuma D, Pantoja-Gomez LM, Cuenca A, Corrêa AS (2022) *Bemisia tabaci* MEAM1 still remains the dominant species in open field crops in Brazil. Brazilian Journal of Biology 84: e256949

Fernandes FR, Cruz AR, Faria JC, Zerbini FM, Aragão FJ (2009) Three distinct begomoviruses associated with soybean in central Brazil. Archives of Virology 154:1567–1570.

Fernandes-Acioli NAN, Pereira-Carvalho RC, Fontenele RS, Lacorte C, Ribeiro SG, Fonseca MEN, Boiteux LS (2011) First Report of Sida micrantha mosaic virus in *Phaseolus vulgaris* in Brazil. Plant Disease 95:1196.

Fontenele RS, Ribeiro GC, Lamas NS, Ribeiro SG, Costa AF, Boiteux LS, Fonseca MEN (2018) First report of Sida micrantha mosaic virus infecting *Oxalis* species in Brazil. Plant Disease 102:1862.

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.

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

Inoue-Nagata AK, Albuquerque LC, Rocha WB, Nagata T (2004) A simple method for cloning the complete begomovirus genome using the bacteriophage φ29 DNA polymerase. Journal of Virological Methods 116:209–211.

Jeske H, Gotthardt D, Kober S (2010) In planta cloning of geminiviral DNA: The true Sida micrantha mosaic virus. Journal of Virological Methods 163:301–308.

Jovel J, Reski G, Rothenstein D, Ringel M, Frischmuth T, Jeske H (2004) Sida micrantha mosaic is associated with a complex infection of begomoviruses different from Abutilon mosaic virus. Archives of Virology 149:829–841

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

Prajapat R, Marwal A, Gaur RK (2014) Begomovirus associated with alternative host weeds: A critical appraisal. Archives of Phytopathology and Plant Protection 47:157–170.

Reis LNA, Fonseca, MEN, Ribeiro SG, Naito, FYB, Boiteux, LS., & Pereira-Carvalho, R. D. C. (2020). Metagenomics of neotropical single-stranded DNA viruses in tomato cultivars with and without the *Ty-1* gene *Viruses 12*(8), 819

Reis LNA, Boiteux LS, Fonseca MEN, Pereira–Carvalho RC (2021) Tomato yellow vein streak virus and Tomato golden vein virus: a reappraisal of the classification status of two South American Begomovirus species based upon genome–wide pairwise identity of multiple isolates. Virus Genes 57:127–131.

Supplementary table 1. Identification of 78 samples exhibiting begomovirus–like symptoms that were obtained from weed plants in Brazil. Information is provided about the region where the isolate was collected, year of collection, and the respective isolate code.

Geographic region	Year of collection	Isolate code
Amazonas - AM	2016	AM-28, AM-30 and AM- 44.
Bahia - BA	2007	BA-004 and BA018.
	2011	BA-082, BA-088, BA-098, BA-103, BA-107, BA-109 and BA-138.
	2012	BA-169 e BA-170,
	2016	BA-180.
	2019	BA-189, BA-190, BA-192 and BA-193.
Ceará - CE	2012	CE-060 and CE-061.
	2017	CE-075.
Distrito Federal - DF	2003	DF0-36 and DF-069.
	2009	DF-275.
	2010	DF-358.
	2011	DF-332 and DF-394.
	2019	DF-707 and DF-708.
Espírito Santo - ES	2001	ES-002.
	2011	ES-047.
	2012	ES-072, ES-076 and ES-077.
Goiás - GO	2003	GO-235 e G0-243.
	2007	GO-366.
	2008	GO-413.
	2009	GO-431 and G0-440
	2010	GO-460, GO-462 and GO-472
	2013	GO-548.

	2019	GO-623.
	2020	GO-268 and GO-628
Minas Gerais - MG	2002	MG-032
	2008	MG-059
	2007	PE-006 and PE-009
Pernambuco - PE	2016	PE-146
Rio de Janeiro - RJ	2006	RJ-010, RJ-011, RJ-012, RJ-013 and RJ-014.
	2016	RJ-055 and RJ-056
Santa Catarina - SC	2010	SC035.
	2005	TO-016 e TO-025,
	2008	TO-100, TO-105, TO-124, TO-134, TO-175 and TO-229.
Tocantins - TO	2009	TO-242, TO-250, TO-259, TO-270, TO-273, TO-275 and TO-285.
	2010	TO-304, TO-307.
	2013	TO-324

Supplementary table 2. Specifc *primers* for *Sida micrantha mosaic virus* to detect the presence of SiMMV in weed plants.

Viral Species	Primer Name	Sequence 5'-3'	Annealing temperature°C
Sida micrantha mosaic virus DNA-A	SiMMV-For	GATCTCGCTCCCCCCTCT	58
	SiMMV-Rev	AGATCGCACGACAACCAG	

Supplementary table 3. Positive samples for Sida micrantha mosaic virus (SiMMV).

Virus	Positive samples by regions				
	North	Northeast	South	Southeast	Midwest
SiMMV	TO-016, TO-100, TO-105, TO-124, TO-175, TO-229, TO-250, TO-259, TO-275, TO-285, TO-304 and TO-324	BA-082, BA-088, BA-098, BA-169, BA-190, BA-193 and CE-061	SC-035	ES-002, ES-076, MG-032, RJ-010, RJ-011, RJ-012, RJ-013 and RJ-056	DF-069, DF-332, DF-394, DF-707, GO-235, GO-243, GO-440, GO-462, GO-472, GO-548, GO-623 and GO-628

Supplementary table 4. Compilation of information about Sida mottle virus

Access Number	Submission	Available	Author team	Partial/complete sequence	DNA-A	Identity with SiMMV	Range of identity with contig 23
AY090555	19-MAR-2002	18-JUN-2004	Fontes,E.P.B.	2668	DNA-A	81,28%	81,59%
AY436328	10-OCT-2003	26-JUL-2016	Faria,J.C.	Partial- 667	CP	23,34%	95,36%
AY444554	20-OCT-2003	26-JUL-2016	Faria,J.C.	Partial- 692	Rep	97,83%	94,80%
JX871377	26-SEP-2012	04-FEB-2013	Tavares,S.S.	2662	DNA-A	82,96%	82,85%